# Classification and Regression Tree (CART) analysis to predict successful delivery of a YMCA childhood obesity intervention program

Meredith Sommars, B.S.

Northwestern University

meredithsommars2018@u.northwestern.edu

630-913-8292

Submitted in fulfillment of the MPH degree.

Sources of Support: YMCA of the USA, Chicago, IL.

*Disclaimer: The views expressed in this article are her own and not an official position of Northwestern University or Y-USA.*

## Abstract

***Background:*** Obesity and its co-morbidities account for a large proportion of chronic disease in the U.S. In an effort to reduce rates of obesity among children, YMCA of the USA (Y-USA) has developed a program called Healthy Weight and Your Child (HWYC), a multi-session intervention program for children carrying excess weight aged 7-13 years old delivered through local YMCAs. To date, 19 local YMCAs have successfully delivered pilot versions of the program. As Y-USA moves from the validation stage of HWYC to translating and scaling the program, Y-USA hopes to improve the model by identifying factors in program delivery that best lead to successful delivery of the program. This analysis used classification and regression tree (CART) modeling to identify such explanatory factors that predict successful delivery of HWYC in local YMCAs, as measured by 'Change in child BMI' and '% Attendance.'

***Methods:*** Participants in HWYC pilot programs included children with obesity

(BMI $\geq$ 95$^{th}$ percentile for age and gender) (n = 782). Data collection included baseline and session 20 weight, height, and waist circumference measures, as well as demographic information. Attendance was measured at every program session. Bivariate and CART analysis was performed in R using previously collected and de-identified data.

***Results:*** CART analysis found that a participant's attendance is most predicted by the YMCA site delivering HWYC, as well as demographic factors such as race and eligibility for free/reduced lunch. The baseline BMI of the accompanying adult also influenced attendance. CART also found that class size was a major predictor of change in child's BMI, as were program model (yearlong vs. condensed) and sex.

***Conclusions:*** The CART analysis successfully identified several factors that predict both attendance and child's BMI change among participants in the HWYC program. Results suggest that Y-USA should scale HWYC to more YMCAs using the condensed 4-month program model to achieve greater successful outcomes. Additionally, individual YMCAs may need to consider the demographics of their unique communities in order to improve both attendance and BMI reduction. As the sample size of the condensed program was small (n = 120), future studies are required to better understand unique explanatory factors that predict successful delivery of HWYC under this program model.

***Keywords:*** CART, regression, childhood obesity, YMCA, HWYC

## Introduction

### *Public Health Relevance*

According to the Center for Disease Control and Prevention (CDC), chronic diseases account for 7 of 10 deaths in the U.S. and 86% of U.S. health care costs, thus contributing great disease and economic burden in this country (1). Chronic diseases include heart disease, stroke, cancer, type 2 diabetes, and obesity. Despite the severe morbidity and mortality associated with chronic disease, many of these diseases could be prevented, delayed, or alleviated through simple lifestyle changes (2).

Obesity and its co-morbidities account for a large proportion of chronic disease in the U.S. According to the National Institute of Diabetes and Digestive and Kidney Diseases, in 2010 roughly 66.7% of U.S adults were considered overweight or obese (3), and 33.3% of U.S. children between the ages of 6 and 19 were considered overweight or obese. Furthermore, approximately half of all Americans are projected to suffer from obesity in 2030 (4). Obesity is associated with several co-morbidities, such as type 2 diabetes, cardiovascular disease, and stroke, all of which are leading causes of death in the U.S. (5). For example, the CDC estimates that 29.1 million or 9.1% of the population currently suffers from type 2 diabetes and that an additional 37% of adults over age 20 are pre-diabetic (6).

With such a large burden of morbidity and mortality associated with obesity, it is of urgent public health importance to identify effective treatment and prevention strategies to curb the incidence of this chronic disease. Many efforts to reduce morbidity and mortality focus on reducing incidence and prevalence of obesity in adults. Others instead focus on childhood obesity to prevent eventual adult incidence of both obesity and its co-morbidities.

***YMCA's Healthy Weight and Your Child Program***

YMCA of the USA (Y-USA) is one such organization that is focusing on reducing childhood obesity to prevent adult obesity. Y-USA is the national resource office for local Ys in the U.S. Their mission is to help Ys achieve their goals by providing them with services, support, and programming. Y-USA is dedicated to strengthening community through youth development, healthy leaving, and social responsibility (7). One objective of Y-USA is to promote community integrated health by strengthening linkages between traditional health care and community-based prevention strategies in order to help individuals prevent, delay, or live better with chronic diseases (8).

Y-USA is currently contributing to the efforts of combating childhood obesity through its program called Healthy Weight and Your Child (HWYC), a childhood obesity intervention program modeled after the evidence-based Mind, Exercise, Nutrition, Do it! program more commonly known as MEND. MEND is a 20-session program that meets for 2 hours roughly twice a week and is attended by the child and at least one parent or caregiver (9-11). The first hour consists of an interactive family session on nutrition and behavior topics, while the second hour focuses on fun exercise for the children while the parents meet for support and discussion.

Y-USA has a license to adapt MEND for delivery in YMCAs and seeks to implement its own version (HWYC) in YMCAs across the country (12). HWYC is currently in the validation and translation stages of program development. In adapting MEND for delivery in YMCAs, Y-USA has made several adjustments to the original model. First, HWYC specifically targets children aged 7 to 13 years old with a BMI above the $95^{th}$ percentile. Second, Y-USA adjusted the original MEND schedule with a step down approach so that sessions occur weekly, then every other week, and then eventually monthly (9). Thus, overall the program lasts nearly an entire year rather than 20 weeks in MEND.

To date, nineteen individual YMCAs have piloted the program since 2015. Though the program appears successful in lowering BMI and waist circumference in participants, the improvement in BMI and waist circumference outcomes have not been as great as in MEND. Additionally, attendance rates have been lower than expected. Thus, one major adjustment that Y-USA has recently implemented is reverting the program schedule to mirror the original program cadence and length of MEND. As such, five YMCAs have piloted a condensed version of the program that consists of 25 sessions delivered over 4 months. Preliminary analysis seems to suggest that this adjustment improves outcomes to a greater degree than the yearlong model. Y-USA plans to move forward with this shorter version of the program.

As Y-USA moves from the validation stage of HWYC to translating and scaling the program, Y-USA hopes to confirm that a shorter condensed program model indeed improves outcomes to a greater degree than a yearlong model. Y-USA also seeks to identify and address other factors relating to program implementation that may affect HWYC outcomes. The literature surrounding the success of MEND may inform potential factors that influence success of HWYC. For instance, one study found that higher attendance was correlated with better outcomes (13). Another study found that sex, race, socio economic status, baseline BMI, class size, attendance, and instructor experience all predicted changes in BMI and self-esteem after completion of the program (14). Yet another study found that certain costs (both monetary and time) related to the program may be barriers to participation, and thus success, of the program (15). Since these factors have already been shown to affect outcomes of MEND, they may potentially also influence outcomes of HWYC. However, additional analysis is required to determine if these factors or others influence delivery of this curriculum in the YMCA setting. By identifying these explanatory factors, Y-USA can then focus adjustments of HWYC to further improve program outcomes as HWYC scales to more YMCAs across the country.

Preliminary linear logistic regression analyses of pilot data conducted by Y-USA have identified several factors that are associated with successful delivery of the pilot program. For instance, Y-USA has identified race and eligibility for free/reduced lunch, as well as whether or not a participant's family has a YMCA membership, as potential predictors of attendance (personal communication). They also have found that whether or not a healthcare professional recommended the participant to the program may predict change in BMI. Since these original analyses, additional data has been collected on YMCA HWYC pilot programs. Here, we describe implementation of a classification and regression tree analysis (CART) to confirm these findings from the preliminary logistic regression in an expanded data set. In addition to identifying explanatory factors in predicting HWYC outcomes, this method also ranks explanatory factors in order of decreasing influence on the outcome. Thus, this method has the additional benefit of prioritizing factors that may be adjusted to improve delivery of HWYC in local YMCAs.

## Methods and Resources

### *Participants*

Participant data from HWYC programs included in this study was previously collected and de-identified by employees in the Research, Evaluation, and Data Sciences (REDS) department at Y-USA, located in Chicago, IL. Participants included children aged 7-13 years old with a BMI at or over the $95^{th}$ percentile for age and gender. Participants and an accompanying adult (mother, father, or grandparent) were recruited to participate in HWYC by each individual YMCA. In total, 19 local YMCAs located across the United States recruited and collected data on a total of 782 individuals.

The HWYC programs included in the analysis were delivered in two formats: first, as a yearlong program with a step-down approach such that sessions occurred weekly, then every other week, and then eventually monthly; and second, as a condensed program delivered within 4 months. While a majority of data collected are from participants within the yearlong model, 5 YMCAs studied here delivered both models. Thus, analysis is done in two ways: first, utilizing the full data set (including participants in both models (n = 782)); and second, directly comparing both models of the program by only including individual data from the 5 YMCAs that delivered both the yearlong model (n = 351) and the condensed program model (n = 120).

### *Data Collection*

Data collection included weighing both child participants and their accompanying adults at session 1, as well as measuring waist circumference and height of the child at session 1. These measures were also taken at the 20$^{th}$ session of the HWYC program to assess change. The measurements were recorded and stored in REDCap (Research Electronic Data Capture) (16) by trained administrators of the HWYC curriculum at each local YMCA and analyzed by members of the REDS department at Y-USA. Local Y program leaders also recorded attendance at every session, defined as the number of sessions attended divided by the number of sessions offered ('% Attendance'). As part of the curriculum, families were given an at-home activity that corresponded to the content of the session. Families were given credit for goal tracking if they demonstrated that they completed some of this activity. Thus, 'Goal Tracker Completion %' is a measure of the number of take-home activities completed divided by the number of take-home activities assigned.

Upon enrollment, program administrators also gathered demographic information from the accompanying adults, including child age, sex and race. As a proxy for socioeconomic status, it was also determined whether the children were eligible for free or reduced lunch at school. Other information was collected about participants such as whether the family of the child had a membership at a local YMCA; the relationship of the accompanying adult to the child; and whether the family and child were referred to the program by a health care professional, including doctors and school nurses or another source.

Additional variables include 'class size,' defined as the number of participants in a group who attended more than 5 sessions. For other analyses done by Y-USA, they define regular participation as those who attend more than 5 sessions (also used by MEND). By defining class size in this way, one can understand how the number of regular participants in a group influences overall attendance rate and success of the program, as measured by a drop in BMI. The variable 'session 1 size' refers to the number of participants who attended the very first session of the class. This variable was used to determine how the number of participants at the start of the HWYC program may influence the likelihood of participants returning to future sessions and in the ultimate success of the program.

***Bivariate Analysis***

All data analysis was conducted using R (17). First, descriptive statistics were calculated on all variables, both in the full data set (Table 1) and in the yearlong vs. condensed program data set (Table 2). The variables of class size and goal tracking completion % include all enrollees regardless of whether they attended >5 sessions. The Shapiro-Wilk normality test and Q-Q plots were used to assess normality of all continuous variables, both in the full data set as well as in the yearlong vs. condensed program data set. Descriptive statistics were calculated and reported using the ‘tableone’ package in R (18).

Next, bivariate analysis was used to assess association between predictors and two outcomes of interest: ‘% Attendance’ and ‘Change in child BMI’. Since the goal is to identify factors that may predict either ‘% Attendance’ or ‘Change in child BMI’, only baseline factors and other variables representing change (Session 1 – Session 20), were evaluated in the bivariate analysis and subsequently included in the CART analysis. Those variables included as explanatory factors of interest are shown in Table 3 for the full data set and in Table 4 for the yearlong vs. condensed program data set. If a participant was missing data on an outcome, this individual was not included in the analysis of that particular outcome. Similarly, if an individual was missing data on a predictor, that individual was not included in bivariate analysis of that predictor with the outcome.

Since both outcomes are non-normal continuous variables, the Wilcoxon rank sum test was used to assess statistical significance between the individual outcomes and each binary predictor. The Kruskal-Wallis rank sum test was used to assess significant association between the outcomes and categorical predictors with greater than 2 groups. Lastly, the Spearman rank correlation analysis was used to assess significant correlation between the outcomes and each continuous predictor. The results of these tests of association are summarized in Table 3 for the full data set and in Table 4 for the yearlong vs. condensed program data set. Predictors with association p-values less than or equal to 0.15 were used as inputs for the CART models to identify explanatory variables. This more lenient statistical cut off was used in order to allow identification of higher order interactions between variables in CART, as detailed below, that may otherwise be missed in a bivariate analysis alone.

***CART Analysis***

A Classification and Regression Tree (CART) analysis was performed to identify explanatory factors that predicted successful delivery of HWYC, as measured by two outcomes: '% Attendance' and 'Change in child BMI'. This method creates a tree that arranges explanatory variables in a hierarchy of decreasing influence on the response variables (19). The CART algorithm analyzes all input explanatory variables (continuous or categorical) and determines which binary division of a single explanatory variable best reduces variance in the response variable (continuous or categorical) (20). For continuous variables, the split is chosen that maximizes the between-groups sum-of-squares in an analysis of variances, which is a measure of the impurity, or diversity, of a node (21). The aim of each split is thus to improve the purity of the child nodes and reduce residuals.

Another way of thinking about this algorithm is that it identifies the variable that splits the data into two subgroups that are each more homogenous with respect to the outcome. It continues to do this by splitting each subgroup into more and more homogenous subgroups by choosing the predictor that best splits the subgroup. After partitioning the data along that variable, the method creates cut off values for continuous outcomes and group values for categorical outcomes (9, 22). The algorithm determines the next most explanatory variable for each partition and continues splitting the data until further splits no longer explain additional variance in the data set.

To assess the two continuous outcomes of ‘Change in child BMI’ and ‘% Attendance’, regression trees were generated using the ‘rpart’ package in R which generates decision trees using the original CART algorithm (23). Rpart uses ANOVA to minimize variance at each split along an explanatory variable and continues recursively splitting the data until no improvement can be made. At each node, a summary value is given to describe that node, which represents the mean value of the outcome in that subset of the data. The program first builds an over fit model using inputs that include all explanatory variables identified from the bivariate analysis independently associated with the outcome (21). Then, this initial model is manually “pruned” by selecting the fewest number of splits that contribute to the overall fit of the model, as determined by root mean squared error (RMSE) and the complexity parameter (cp), a measure of the ‘cost’ of each additional split (19, 21, 24).

With regards to missing data, ‘rpart’ takes an ambitious approach towards handling missing values. Any observation that contains values for the dependent variable (outcome) and at least one independent variable (predictor) will participate in the modeling (21). At each split, the purity of the node is calculated only over the observations that are not missing the relevant predictor. However, observations that are missing values for the relevant predictor must still be partitioned into either group along a split. In deciding which way an observation goes along the split, rpart uses surrogate variables. Surrogate variables are determined by calculating to what extent alternative splits resemble the best split in terms of how many observations they send the same way (21, 24). Observations with missing data for the best split are thus grouped based upon the next best surrogate split. For the yearlong data set, 2 individuals had missing data on '% Attendance' and 453 individuals had missing data on 'Change in child BMI.' Thus, these trees were constructed using 780 and 329 individuals, respectively. For the yearlong vs. condensed data set, 2 individuals had missing data on '% Attendance' and 241 individuals had missing data on 'Change in child BMI.' Thus, these trees were constructed using 469 and 230 individuals, respectively.

In order to maximize predictive certainty of the CART models generated by 'rpart', the full data set was randomly portioned into a training set (90%) and a test set (10%). The training set was used to both build the initial model and the final model after pruning based upon RMSE and cp. The test set was then fed into the model to compare RMSE and confirm validity of the model. Training and test sets were not used for the yearlong vs. condensed program data set because the sample size was much smaller; thus, the initial model was constructed using all the data and the final model was chosen based on the lowest value of the cp and RMSE.

In addition to the 'rpart' package, the 'rattle' (version 4.1.0) (25), 'rpart.plot' (26), and 'RColorBrewer' (27) R packages were used to generate the decision trees.

## Results

Distribution of demographic variables for all individuals (n = 782) included in this study can be seen in Table 1. All variables used as inputs for the bivariate analysis are shown. Distribution of demographic variables for individuals who participated in the HWYC program at the 5 YMCAs that delivered both models (yearlong vs. condensed program) can be seen in Table 2 (n = 471). These 5 locations include participant data from sites A, B, D, K, and N. The data has been stratified by program model to directly compare success of the two program models. As shown, 351 individuals are included in the yearlong program and 120 individuals are included in the condensed 4-month program.

***Bivariate analysis of full data set***

Bivariate analysis of each predictor and its association with the two outcomes of interest, '% Attendance' and 'Change in child BMI', are shown for the full data set (n = 782) in Table 3. As shown, '% Attendance' demonstrated a significant association ($p \leq 0.15$) with several demographic characteristics, including race; child age; eligibility for free/reduced lunch; whether the accompanying parent was the mother; and whether the family had a Y membership. Attendance also was significantly associated with the YMCA site. All variations of baseline measures of child BMI (child BMI, BMI z-score, BMI % over the 95$^{th}$ percentile, and BMI percentile) were also significantly associated with '% Attendance', though baseline waist circumference was not significantly associated with '% Attendance'. Class size was also associated with '% Attendance'. Additionally, change in waist circumference, change in BMI z-score, and change in adult BMI were also all associated with '% Attendance'. Lastly, goal tracker completion % was also significantly associated with '% Attendance', though these two variables are highly correlated.

The other outcome of interest, 'Change in child BMI', was associated with a few demographic variables including sex and child age. It was also significantly associated with the program type. 'Change in child BMI' was also significantly associated with the class size. Change in adult BMI was also significantly associated with the 'Change in child BMI'. Lastly, change in BMI z-score, change in BMI % over $95^{th}$ percentile, and change in waist circumference were all associated with 'Change in child BMI', but these are all highly correlated and essentially capture the same measure.

***Bivariate analysis of yearlong vs. condensed program data set***

Bivariate analysis of each predictor and its association with the two outcomes of interest, 'Change in child BMI' and '% Attendance', are shown for the yearlong vs. condensed program data set (n = 481) in Table 4. Within this data set, '% Attendance' was significantly associated with several demographic factors, such as race and eligibility for free/reduced lunch. It was also associated with several baseline measures, including: adult BMI; BMI % over $95^{th}$ percentile; child BMI; child BMI percentile; and child BMI z-score. '% Attendance' was also associated with change in adult BMI, as well as YMCA site. Both class size and session 1 size were also significantly associated with '% Attendance'. Again, goal tracker % completion was associated with '% Attendance', but these variables are highly correlated.

Bivariate analysis indicates that several demographic characteristics are significantly associated with 'Change in child BMI' as well, such as sex, child age, and whether or not a health practitioner referred the child to the program. The program model was also significantly associated with 'Change in child BMI'. Both class size and session 1 size were significantly associated with 'Change in child BMI', as well as baseline child BMI. There change variables (change in BMI % over $95^{th}$ percentile, change in waist circumference, and change in BMI z-score) were all significantly associated with ‘Change in child BMI’, but these variables are all highly correlated and capture the same measure.

***CART analysis of ‘% Attendance’ within full data set***

Within the full data set, the regression tree of the outcome ‘% Attendance’ was constructed using a training set (90%) consisting of 702 observations, shown in Figure 1. As shown at the top of the tree, the average attendance within the entire group was 55%. The most explanatory variable in predicting attendance was YMCA site. Higher average attendance (63%) was observed in individuals who attended a HWYC program at sites B, C, D, J, M, N, O, P, and Q, which represents 358 individuals. The lower attendance group (46% on average) consisted of individuals (n = 344) who attended programming at sites A, E, F, G, H, I, K, L, R, and S.

The ‘% Attendance’ of the higher attendance group (n = 358) was then predicted by an additional break down of YMCA site, where the average attendance among participants at sites J and O was 87% (n = 19) and the average attendance among the remaining participants was 61% (n = 339). In this program, ‘% Attendance’ was further predicted based upon eligibility for free/reduced lunch. Among those who were not eligible for free/reduced lunch, participants attended 67% of the sessions (n = 179), whereas those participants who were eligible for free/reduced lunch attended only 55% of the sessions (n = 160).

Moving back to the top of the tree after the initial split based upon YMCA site, the lower attendance group (46%) consisting of 344 individuals was further broken down based upon race to predict ‘% Attendance’. Within this group, participants who identified as white attended 53% of the sessions on average (n = 152) compared to non-white identifying individuals who attended 41% of the sessions (n = 192). Among white identifying individuals, '% Attendance' was further predicted again based on YMCA site. Participants who attended HWYC programs at sites F or S attended 75% of the sessions (n = 26) while the remaining participants attended 48% of the sessions on average (n = 126).

***CART analysis of 'Change in child BMI' in the full data set***

Within the full data set, the regression tree of the outcome 'Change in child BMI' was constructed using a training set (90%) consisting of 293 observations, shown in Figure 2. Many observations were excluded from the analysis due to missing data on the outcome. Within the entire group (n = 293), the average 'Change in child BMI' was an increase of 0.22. The most explanatory factor in predicting 'Change in child BMI' was class size. Participants in classes greater than or equal to 19 children were likely to gain 1.8 BMI units on average (n = 16), as compared with participants in classes with fewer than 19 participants who gained 0.13 units in BMI on average (n = 277). Among these participants, 'Change in child BMI' was further predicted by the program model. Children who participated in the shorter condensed program (n = 52) lost 0.45 BMI units on average, while those who participated in the yearlong model (n = 225) gained 0.26 BMI units on average. Among these participants, class size again predicted 'Change in child BMI'. Participants in classes with 9 or fewer children (n = 103) gained 0.55 BMI units, while participants in classes with 10 or greater children (n = 122) gained only 0.011 BMI units. This group is further broken down by class size, as those with 12 or greater children (n = 64) gained 0.45 BMI units and those with fewer than 12 participants lost 0.48 BMI units (n = 58).

***CART analysis of '% Attendance' in yearlong vs. condensed program data set***

Within the yearlong vs. condensed program data set, the regression tree of the outcome '% Attendance' was constructed using 469 observations where the mean attendance was 57%, shown in Figure 3. Similar to the full data set, the YMCA site was the most explanatory factor in predicting attendance. Sites B, D, and N demonstrated a higher average attendance rate at 62% (n = 323) compared to sites A and K where the average attendance rate was 45% (n = 146).

In the higher attendance group (n = 323), attendance was further predicted by eligibility for free/ reduced lunch. Those who were not eligible for free lunch attended 68% of sessions on average (n = 178), while those who were eligible attended 55% of the sessions on average (n = 145). Moving back to the top of the tree, among the lower attendance group (n = 146), the baseline adult BMI was the next most explanatory factor in predicting attendance. If the accompanying adult had a starting BMI of less than 25, the mean attendance was only 16% (n = 10). However, if the accompanying adult had a starting BMI of 25 or greater, the average attendance was 47% (n = 136). Among these individuals, the adult baseline BMI further predicted attendance, as those with a baseline BMI greater of 32 or greater attended 42% of the sessions (n = 103) and those with a baseline BMI less than 32 attended 63% of the sessions (n = 33).

***CART analysis of 'Change in child BMI' in yearlong vs. condensed program data set***

Within the yearlong vs. condensed program data set, the regression tree of the outcome 'Change in child BMI' was constructed using 230 observations, shown in Figure 4. Overall, participants gained 0.28 BMI units by the end of the program. Similar to the full data set, class size was the most explanatory factor in predicting 'Change in child BMI'. Participants in class with 19 or more children gained 1.8 BMI units on average (n = 16) compared with participants in classes with fewer than 19 children who gained 0.17 BMI units on average (n = 214). Among these participants, the program model was the next most explanatory factor in predicting 'Change in child BMI'. Participants in the shorter 4-month program lost 0.39 BMI units (n = 56) compared to participants in the yearlong program who gained 0.37 BMI units (n = 158). Among the yearlong participants, sex was the next most explanatory factor in predicting 'Change in child BMI'. On average, girls gained 0.58 BMI units (n = 88) while boys only gained 0.11 BMI units (n = 70). Among boys in the yearlong program, class size was the next most explanatory factor in predicting 'Change in child BMI'. Boys in a class with 12 or more participants gained 0.44 BMI units (n = 33) compared to participants in classes with fewer than 12 children who lost 0.19 BMI units on average (n = 37). In this group, class size again predicted 'Change in child BMI'. Participants in classes with 9 or fewer children gained 0.3 BMI units (n = 28) while participants in classes with 10 or greater children lost 1.7 BMI units on average (n = 9).

### *Additional analysis*

Within the yearlong vs. condensed program data set, both class size and session 1 size were associated with '% Attendance' and 'Change in child BMI'. However, since these variables are so correlated with each other, they were assessed in separate CART models.

When tested as an input for '% Attendance', session 1 size did not appear as an explanatory model so that resulting tree did not change. However, when tested as an input for 'Change in child BMI', session 1 size slightly altered the resulting tree, as shown in Figure 5. As before, the overall 'Change in child BMI' was 0.28 (n = 230). Here, the most explanatory factor was session 1 size. Participants in a class with 20 or more children at the first session gained 1.3 units in BMI on average (n = 30) as compared to participants in a class with fewer than 20 children at the first session who gained only 0.13 BMI units on average (n = 200). Among these participants, program model was the next most explanatory factor, as participants in the shorter condensed program lost 0.39 BMI units on average (n = 56) and participants in the yearlong program gained 0.33 BMI units on average (n = 144).

## Discussion

The described CART analysis identified several key findings regarding the pilot implementation of the HWYC program in local YMCAs. First, the analysis found that the YMCA location explains the greatest variance in attendance. This observation was clear in both data sets. In going forward, Y-USA may need to ensure a certain level of standardization and fidelity in program delivery across YMCAs currently offering the program. Additionally, it may be useful for Y-USA to delve deeper into understanding why some YMCAs struggle to maintain high attendance, while others do not. Perhaps YMCAs with high attendance rates are using incentives to keep participants coming back or have learned other strategies that may be useful to other locations that struggle to maintain higher attendance rates. There may also be barriers, such as poor HWYC instruction; participant proximity to the YMCA; the work schedule of the accompanying adult; or other community-specific factors that influence the lower attendance rate at some YMCAs. In improving attendance rates, each YMCA may need to learn how to best serve their community as each community is different with unique challenges and strengths.

Second, the CART analysis found that attendance is also influenced by demographic factors such as race and socioeconomic status, consistent with the linear regression models previously conducted on HWYC. After accounting for YMCA location, the analysis found within the full data set that participants that identity as white have a higher attendance rate than those identifying as non-white at certain locations (Figure 1). Thus, the YMCAs that are specifically found to have lower attendance among non-white participants may need to develop specific strategies for improving attendance in these demographic groups. For instance, they might consider matching the race of the instructor with the race of the majority of the participants to increase familiarity and comfort, in an effort to achieve racial/cultural concordance. This same analysis finds that among other YMCA locations, socioeconomic status influences attendance rate (Figure 1). Specifically, participants who were eligible for free/reduced lunch had lower attendance rates as compared to participants who were ineligible (also seen in Figure 3). In this instance, there may be other barriers related to socioeconomics that influence the ability of participants to consistently attend HWYC sessions. For instance, the accompanying adult may need to work longer or inconsistent hours, thus inhibiting the ability of the child to consistently attend class. Alternatively, some families may be intimidated by the promotion of healthy eating, thinking it may be too expensive. This could then influence the likelihood of that family attending consistent sessions. Ultimately, the difference in attendance rate observed at different YMCA sites are likely a reflection of these underlying demographic influences on attendance.

Though the CART findings were fairly consistent between the full and yearlong vs. condensed data sets, the latter analysis also identified adult baseline BMI as an explanatory factor in predicting attendance. Specifically, at certain YMCAs, adults with a baseline BMI less than 25 were the least likely to attend (16%). Adults with a baseline BMI between 25 and 32 were the most likely to attend (63%), while adults with a baseline BMI greater than 32 were less likely to attend (42%). These findings suggest that the baseline BMI of the accompanying adult may influence how often the participant attends the sessions, with those adults with a low or high BMI being less likely to consistently attend. This is particularly relevant as children aged 7-13 rely on adults to even attend the sessions.

The next finding from the CART analysis of HWYC relates to explanatory factors regarding ‘Change in child BMI’. Across both data sets, class size consistently arose as the most explanatory factor in predicting ‘Change in child BMI’ (Figure 2, Figure 4, and Figure 5). Specifically, participants in classes with 19 or more students gained more weight than participants in smaller classes. However, the CART trees also suggest that there may be an ideal mid-range in class size. Shown in Figure 2, among participants in the yearlong model, participants in classes with 10-11 children exhibited weight loss, while classes of other sizes tended to gain weight. This same pattern is observed in the yearlong vs. condensed data set (Figure 4). This finding is consistent with a study on

MEND which found that participants in smaller classes had greater reductions in BMI by the end of the program (14). The beneficial effects of smaller class size are seen in many settings, particularly among educational settings (28-31). Even fitness-specific classes for children have been shown more effective when class size is reduced (32). As such, YMCA's may want to consider limiting class size when registering future participants in an effort to maximize weight maintenance and loss.

A fourth finding from the CART analysis of 'Change in child BMI' indicates that the shorter condensed program is more effective than the yearlong program in reducing BMI. This finding is consistent with the original linear regression models previously done on HWYC. This is observed in both the full data set (Figure 2) and the direct yearlong vs. condensed data set (Figure 4 and Figure 5). In both data sets, participants in the condensed program lost weight while participants in the yearlong program gained weight on average. This shorter program is a return to the original MEND model and should be used in future HWYC programs in order to maximize program effectiveness.

There were a few other similarities between the CART analysis of HWYC and other studies assessing the effectiveness of the original MEND curriculum. For instance, CART identifies sex as an explanatory factor in predicting 'Change in child BMI', at least in the yearlong vs. condensed data set. The model finds that among yearlong participants, females were more likely to gain weight than males by the end of the program, consistent with one study that found greater BMI reductions in boys (14). Though age did not emerge as explanatory in predicting change in BMI, it was significantly associated with 'Change in child BMI' in the bivariate analysis, again consistent with studies evaluating MEND (14).

Despite these, there were also some differences as well. A major difference is that the literature on MEND identifies a dose-responsive relationship between attendance and reduction in BMI, that participants who attend more sessions also see greater reductions in BMI by the end of the program (11, 13-15). However, neither the CART analysis nor the initial bivariate analysis finds this association between ‘Change in child BMI’ and attendance in HWYC. The reason for this is not understood, but perhaps future studies with a greater number of participants will uncover a similar dose-response. Other factors, including socioeconomic status, baseline BMI, and race influenced the success of MEND in terms of reduction in BMI, but were not identified as explanatory factors in this analysis of HWYC. Lastly, the original linear regression models performed by Y-USA found that participants who were recommended to HWYC by health care providers had better outcomes; however, the CART analysis described here did not find this association.

In addition to assembly of these findings in this thesis format, the results of the CART analysis were shared with the Y-USA during a presentation at the HWYC Issues and Actions Committee. This committee evaluates progress of HWYC pilot programs and discusses necessary changes and improvements. The CART findings were also summarized in a PowerPoint presentation and given to the REDS department.

***Limitations and strengths***

On major limitation of this analysis relates to the CART methodology itself. CART does not provide a p-value to assess statistical significance of the associations it identifies. However, use of a bivariate analysis before the CART analysis was used to counter this limitation. Related, the CART method also tends to over fit the data. This is related to the fact that the method attempts to account for missing data by use of surrogate variables. Over fitting also can result from the fact that the sample pool is made smaller with each split. Thus, sometimes the splits can result in overly complex trees. The act of “pruning” the tree based upon cp and RMSE attempts to counter any over fitting. Other disadvantages of this study include a limited sample size of participants in the condensed program. However, future analyses will include more participants in this program as Y-USA plans to implement this model going forward.

Despite the limitations, an advantage of CART is that it can identify higher interactions among variables (33). Its ability to handle missing data is also an advantage as long as pruning is used to prevent over fitting. Another advantage of this study is the geographic diversity of YMCA’s used in the pilot study.

### *Conclusions*

In conclusion, CART analysis has identified YMCA site, race, and eligibility for free/reduced lunch as major determinants of attendance among participants in the HWYC program. Baseline BMI of the accompanying adult may also influence attendance. Additionally, CART analysis has identified class size and program model as major determinants of change in child’s BMI. Sex may also influence change in child’s BMI. However, further analyses with additional participant data for the condensed program may help inform additional explanatory factors, as this will likely be the HWYC program format going forward.

## References

1. Chronic Disease Prevention and Health Promotion 2016 [Available from: https://www.cdc.gov/chronicdisease/index.htm.

2. The Growing Crisis of Chronic Disease in the United States. In: Disease PtFC, editor.

3. Overweight and Obesity Statistics. In: NIDDK, editor. 2010.

4. Michael J. O'Grady JCC. Assessing the Economics of Obesity and Obesity Interventions. Campaign to End Obesity. 2012:1-40.

5. Leading Causes of Death Centers for Disease Control and Prevention: National Center for Health Statistics; 2014 [Available from: https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm.

6. 2014 National Diabetes Statistics Report. Atlanta: Centers for Disease Control and Prevention; 2015. Report No.: .

7. Delivering Our Cause. Report. Chicago, IL: YMCA of the USA; 2014.

8. Community Integrated Health [Website]. The YMCA; 2017 [Available from: http://www.ymca.net/community-integrated-health.

9. Implementing the Healthy Weight and Your Child Program. Chicago, IL: Altarum Institute; 2017.

10. MEND stands for Mind, Exercise, Nutrition, Do It! [MEND Foundation Website]. Available from: http://www.mendfoundation.org/.

11. Paul M. Sacher MK, Paul M. Chadwick, Tim J. Cole, Margaret S. Lawson, Alan Lucas, Atul Singhal. Randomized Controlled Trial of the MEND Program: A Family-

based Community Intervention for Childhood Obesity. Obesity Journal. 2010;18:S62-S8.

12. YMCA of the USA and Healthy Weight Partnership will expand childhood obesity program, providing children and families critical support for reaching a healthy weight [press release]. San Diego, CA, June 30, 2015 2015.

13. Louise L. Hardy SM, Joanne Gale, Binh Nguyen, Louise A. Baur, Blythe J. O'Hara. Translational research: are community-based child obesity treatment programs scalable? . BMC Public Health. 2015;15(652):1-8.

14. J Fagg PC, TJ Cole, S Cummins, H Goldstein, H Lewis, S Morris, D Radley, P Sacher, C Law. From trial to population: a study of a family-based community intervention for childhood overweight implemented at scale. International Journal of Obesity. 2014;38:1343-9.

15. Patricia J Lucas KC-T, Lisa Arai, Sally Stapley, Jamie Fagg, Helen Roberts. What works in practice: user and provider perspectives on the acceptability, affordability, implementation, and impact of a family-based intervention for child overweight and obesity delivered at scale. BMC Public Health. 2014;2014(14):1-12.

16. Paul A. Harris RT, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, Jose G. Conde. Research electronic data capture (RedCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-81.

17. Team RC. R: A language and environment for statistical computing. . Vienna, Austria: R Foundation for Statistical Computing; 2014.

18. Kazuki Yoshida JB. Create “Table 1” to Describe Baseline Characteristics. 2018.

19. Leo Breiman JF, Charles J. Stone, R.A. Olshen. Classification and Regression Trees: Chapman and Hall/CRC; 1984.

20. Rick L. Lawrence AW. Rule-Based Classification Systems Using Classification and Regression (CART) Analysis. Photogrammetric Engineering & Remote Sensing. 2001;67(10):1137-42.

21. Terry M. Therneau EJA. An Introduction to Recursive Partitioning Using the RPART Routines. Mayo Foundation2018.

22. Institute D. CART Algorithm for Decision Tree: Dni Institute; 2017 [Available from: http://dni-institute.in/blogs/cart-algorithm-for-decision-tree/.

23. Ripley B. Recursive Partitioning for Classification. 2017.

24. Feelders AJ. Handling missing data in trees: surrogate splits or statistical imputation? PKDD. 1999:329-34.

25. Williams GJ. Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery. Use R!: Springer; 2011.

26. Milborrow S. Plot ‘rpart’ Models: An Enhanced Version of ‘plot.rpart’. 2017.

27. Neuwirth E. ColorBrewer Palettes. 2014.

28. Class size and student achievement: Research review Alexandria, VA: Center for Public Education;  [Available from: http://www.centerforpubliceducation.org/Main-Menu/Organizing-a-school/Class-size-and-student-achievement-At-a-glance/Class-size-and-student-achievement-Research-review.html.

29. Mosteller F. The Tennessee Study of Class Size in the Early School Grades. The Future of Children. 1995;5(2):113-27.

30. Matthew M. Chingos GJW. Class Size: What Research Says and What it Means for State Policy. Brown Center on Education Policy at Brookings. 2011.

31. Gene V. Glass MLS. Meta-Analysis fo Research on Class Size and Achievement. Educational Evaluation and Policy Analysis. 1979;1(1):2-16.

32. Kirkham-King M BT, Hannon JC, Castelli DM, Hilton K, Burns RD. Elementary physical education: A focus on fitness activities and smaller class sizes are associated with higher levels of physical activity. Prev Med Rep. 2017;8:135-9.

33. Richard K. Zimmerman ea. Classification and Regression Tree (CART) analysis to predict influenza in primary care patients. BMC Infectious Disease. 2016;16(503).

**Table 1.** Sample characteristics in the full data set (n = 782).

| **Demographic Variables** | **Level** | **Measure (n=782)** | | |
|---|---|---|---|---|
| | | **Median (IQR)** | | |
| Child Age | | 10.0 (9.0-11.0) | | |
| Session 1 Size | | 12.0 (10.0-16.0) | | |
| Class Size | | 10.0 (7.0-12.0) | | |
| Attendance (%) | | 65.0 (25.0-85.0) | | |
| Goal Tracking Completion (%) | | 40.0 (10.5-68.4) | | |
| | | **n (%)** | | |
| Sex | Male | 357 (45.7) | | |
| | Female | 425 (54.3) | | |
| Race - White | No | 391 (50.0) | | |
| | Yes | 391 (50.0) | | |
| Accompanying Parent - Mother | No | 168 (21.6) | | |
| | Yes | 611 (78.4) | | |
| Program Referral - Doctor | No | 385 (49.5) | | |
| | Yes | 392 (50.5) | | |
| Family Y Membership | No | 506 (65.6) | | |
| | Yes | 265 (34.4) | | |
| Eligibility for Free/Reduced Lunch | No | 318 (40.7) | | |
| | Yes | 463 (59.3) | | |
| Program Type | Condensed | 120 (15.3) | | |
| | Yearlong | 662 (84.7) | | |
| YMCA Site | Site A | 77 (9.8) | | |
| | Site B | 84 (10.7) | | |
| | Site C | 30 (3.8) | | |
| | Site D | 123 (15.7) | | |
| | Site E | 17 (2.2) | | |
| | Site F | 28 (3.6) | | |
| | Site G | 57 (7.3) | | |
| | Site H | 14 (1.8) | | |
| | Site I | 48 (6.1) | | |
| | Site J | 18 (2.3) | | |
| | Site K | 71 (9.1) | | |
| | Site L | 8 (1.0) | | |
| | Site M | 8 (1.0) | | |
| | Site N | 116 (14.8) | | |
| | Site O | 4 (0.5) | | |
| | Site P | 7 (0.9) | | |
| | Site Q | 13 (1.7) | | |
| | Site R | 14 (1.8) | | |
| | Site S | 45 (5.8) | | |
| **Anthropomorphic Variables** | | **Median (IQR)* or *Mean (SD)*** | | |
| | | **Baseline** | **Final** | **Change** |
| Child BMI (kg/m$^2$) | | 29.5 (26.0-33.6)* | 29.2 (25.8-33.5)* | 0.2 (-0.8-1.0)* |
| Child BMI z-score | | *2.3 (0.3)* | 2.3 (2.0-2.5)* | *-0.05 (0.1)* |
| BMI % over 95th Percentile | | *21.8 (12.4)* | *19.6 (12.4)* | *-1.3 (4.6)* |
| Child Waist Circumference (in) | | *35.1 (5.4)* | 34.6 (31.5-38.5)* | *0.2 (2.0)* |
| Adult BMI | | *36.0 (8.2)* | 34.9 (29.5-40.8)* | *-0.16 (2.4)* |

**Table 2.** Sample characteristics in the yearlong (n = 351) vs. condensed (n = 120) data set (only those YMCA sites that have offered both program models).

| Variable | Level | Yearlong (n=351) | | | Condensed (n=120) | | |
|---|---|---|---|---|---|---|---|
| | | **Median (IQR)** | | | **Median (IQR)** | | |
| Child Age | | 10.0 (9.0-11.0) | | | 10.0 (8.0-10.0) | | |
| Session 1 Size | | 14.0 (10.0-18.0) | | | 11.0 (9.0-11.0) | | |
| Class Size | | 11.0 (7.0-15.0) | | | 8.0 (6.8-10.0) | | |
| Attendance (%) | | 65.0 (27.5-80.0) | | | 70.0 (25.0-80.0) | | |
| Goal Tracking Completion (%) | | 42.1 (15.8-63.2) | | | 40.0 (10.0-70.0) | | |
| | | **n (%)** | | | **n (%)** | | |
| Sex | Male | 154 (43.9) | | | 56 (46.7) | | |
| | Female | 197 (56.1) | | | 64 (53.3) | | |
| Race – White | No | 179 (51.0) | | | 62 (51.7) | | |
| | Yes | 172 (49.0) | | | 58 (48.3) | | |
| Accompanying Parent – Mother | No | 88 (25.1) | | | 31 (26.3) | | |
| | Yes | 263 (74.9) | | | 87 (73.7) | | |
| Program Referral – Doctor | No | 181 (51.6) | | | 49 (41.2) | | |
| | Yes | 170 (48.4) | | | 70 (58.8) | | |
| Family Y Membership | No | 213 (60.7) | | | 76 (65.0) | | |
| | Yes | 138 (39.3) | | | 41 (35.0) | | |
| Eligibility for Free/Reduced Lunch | No | 175 (49.9) | | | 60 (50.0) | | |
| | Yes | 176 (50.1) | | | 60 (50.0) | | |
| Program Type | Condensed | 0 (0.0) | | | 120 (100.0) | | |
| | Yearlong | 351 (100.0) | | | 0 (0.0) | | |
| YMCA Site | Site A | 61 (17.4) | | | 16 (13.3) | | |
| | Site B | 71 (20.2) | | | 13 (10.8) | | |
| | Site D | 92 (26.2) | | | 31 (25.8) | | |
| | Site K | 32 (9.1) | | | 39 (32.5) | | |
| | Site N | 95 (27.1) | | | 21 (17.5) | | |
| **Anthropomorphic Variables** | | **Median (IQR)* or *Mean (SD)*** | | | **Median (IQR)* or *Mean (SD)*** | | |
| | | **Baseline** | **Final** | **Change** | **Baseline** | **Final** | **Change** |
| Child BMI (kg/m$^2$) | | 29.1 (25.9-32.9)* | 28.6 (25.7-33.5)* | 0.4 (-0.6-1.5)* | 30.0 (26.4-30.0)* | 29.9 (26.0-30.0)* | -0.3 (-0.9-0.3)* |
| Child BMI z-score | | 2.3 (2.0-2.5)* | 2.3 (2.0-2.5)* | -0.02 (-0.10-0.04)* | 2.4 (2.1-3.0)* | 2.4 (2.1-3.0)* | -0.04 (-0.09--0.09)* |
| BMI % over 95th Percentile | | *20.7 (12.2)* | *19.0 (12.2)* | *-0.7 (4.7)* | *23.2 (12.8)* | *20.7 (12.7)* | *-1.7 (2.5)* |
| Child Waist Circumference (in) | | 35.6 (32.0-39.0)* | 35.8 (31.9-39.0)* | *0.4 (2.2)* | 34.0 (31.1-40.0)* | 33.1 (31.1-40.0)* | *-0.3 (1.3)* |
| Adult BMI | | *35.8 (7.8)* | *35.2 (7.8)* | *0.1 (2.8)* | *36.2 (8.5)* | *36.1 (7.0)* | *-0.3 (0.7)* |

**Table 3.** Bivariate analysis of outcomes and predictors within the full data set (n = 782). Comparisons with p-value $\leq$ 0.15 are bolded and used as inputs for CART analysis.

| Predictor | Outcome – p-value | |
|---|---|---|
| | % Attendance | Change in Child BMI |
| ***Normal Variables*** | *a* | *a* |
| BMI z-score, baseline | **0.0192** | 0.8270 |
| BMI % over 95th Percentile, baseline | **0.0185** | 0.8915 |
| Waist Circumference (in), baseline | 0.4873 | 0.6601 |
| Adult BMI, baseline | 0.1698 | 0.2531 |
| Change in BMI z-score | **0.0977** | **<0.0001** |
| Change in BMI % over 95th Percentile | 0.1668 | **<0.0001** |
| Change in Waist Circumference (in) | **0.0947** | **<0.0001** |
| Change in Adult BMI | **0.1526** | **0.0226** |
| ***Non-Normal Variables*** | *a* | *a* |
| % Attendance | - | 0.2466 |
| Goal Tracker Completed % | <**0.0001** | 0.7534 |
| Child Age | **0.0940** | **0.0823** |
| Child BMI, baseline | **0.0130** | 0.7090 |
| Child BMI Percentile, baseline | **0.0154** | 0.8116 |
| Class Size | **0.0502** | **0.1319** |
| Session 1 Size | 0.8655 | 0.2538 |
| Change in Child BMI | 0.2466 | - |
| ***Binary Variables*** | *b* | *b* |
| Sex | 0.3105 | **0.0294** |
| Race - White | **0.0001** | 0.6009 |
| Eligibility for Free/Reduced Lunch | **0.0001** | 0.5604 |
| Accompanying Parent - Mother | **0.0805** | 0.5390 |
| Program Referral - Doctor | 0.3653 | 0.6091 |
| Family Y Membership | **0.0753** | 0.2273 |
| Dose | 0.4987 | **0.0002** |
| ***Categorical Variables*** | *c* | *c* |
| YMCA Site | <**0.0001** | 0.5397 |

*a = Spearman's rank correlation; rho (p-value)*

*b = Wilcoxon rank sum test; Wilcoxon Rank (p-value)*

*c = Kruskal – Wallis Rank Sum Test; chi-squared value (p-value)*

**Table 4.** Bivariate analysis of outcomes and predictors among Ys with both yearlong and condensed programs (n = 471). Comparisons with p-value ≤ 0.15 are bolded and used as inputs in CART analysis.

| Predictor | Outcome - p-value | |
|---|---|---|
| | % Attendance | Change in Child BMI |
| ***Normal Variables*** | *a* | *a* |
| BMI % Over 95th Percentile, baseline | **0.0910** | 0.3315 |
| Adult BMI, baseline | **0.0759** | 0.3373 |
| Change in BMI % Over 95th Percentile | 0.5534 | **0.0001** |
| Change in Waist Circumference (in) | 0.4087 | **0.0001** |
| Change in Adult BMI | **0.1514** | 0.3437 |
| ***Non-Normal Variables*** | *a* | *a* |
| % Attendance | - | 0.5828 |
| Goal Tracker % Completed | **0.0001** | 0.9129 |
| Child Age | 0.1616 | **0.0584** |
| Child BMI, baseline | **0.0418** | **0.1459** |
| Child BMI Percentile, baseline | **0.0952** | 0.3775 |
| Child BMI z-score, baseline | **0.0984** | 0.3756 |
| Waist Circumference (in), baseline | 0.5607 | 0.9514 |
| Session 1 Size | **0.0381** | **0.0262** |
| Class Size | **0.0774** | **0.0244** |
| Change in Child BMI | 0.5828 | - |
| Change in BMI z-score | 0.5299 | **0.0001** |
| ***Binary Variables*** | *b* | *b* |
| Sex | 0.4795 | **0.0814** |
| Race - White | **0.0002** | 0.4101 |
| Eligibility for Free/Reduced Lunch | **0.0001** | 0.2509 |
| Accompanying Parent - Mother | 0.5359 | 0.6639 |
| Program Referral - Doctor | 0.7684 | **0.1088** |
| Family Y Membership | 0.2567 | 0.4789 |
| Dose | 0.9373 | **<0.0001** |
| ***Categorical Variables*** | *c* | *c* |
| YMCA Site | **<0.0001** | 0.3537 |

*a = Spearman's rank correlation; rho (p-value)*

*b = Wilcoxon rank sum test; Wilcoxon Rank (p-value)*

*c = Kruskal – Wallis Rank Sum Test; chi-squared value (p-value)*

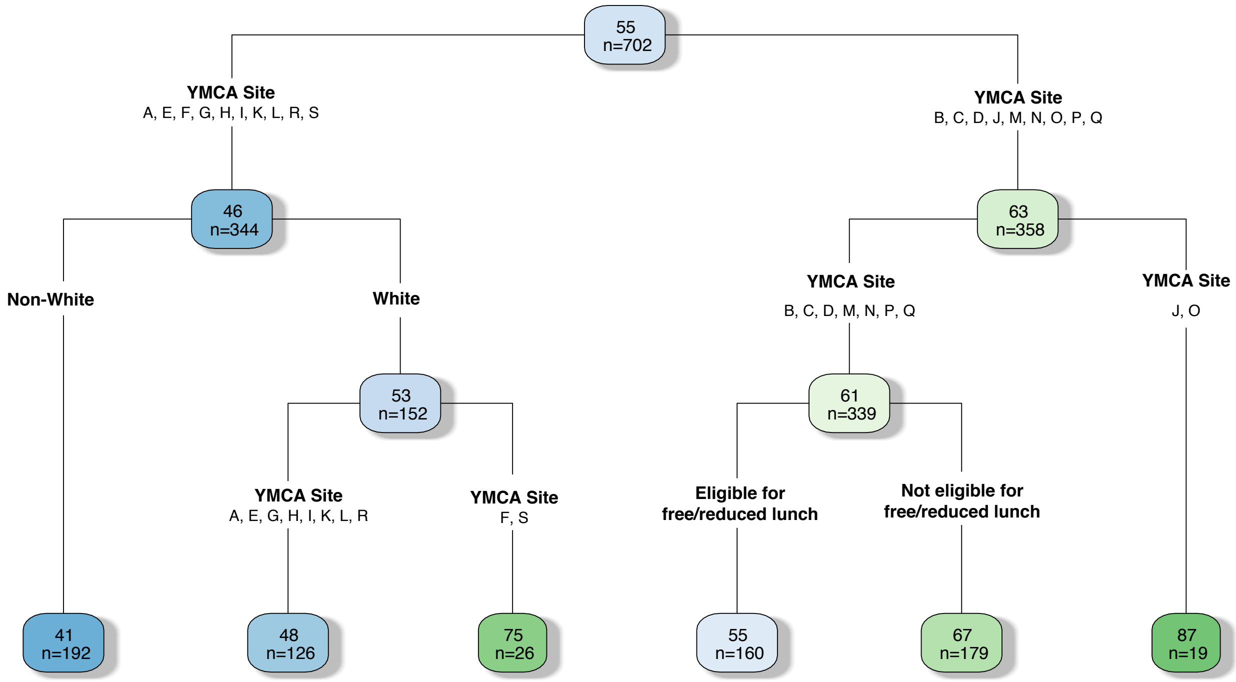


**Figure 1.** Regression tree of the outcome, "% Attendance," within the **Full** dataset. Boxed nodes represent splitting of the data along a single variable, where the top number represents the mean value for the outcome and the lower number represents the number of individuals within that group. The shades of the nodes visually represent the mean value of the outcome from low (blue) to high (green).

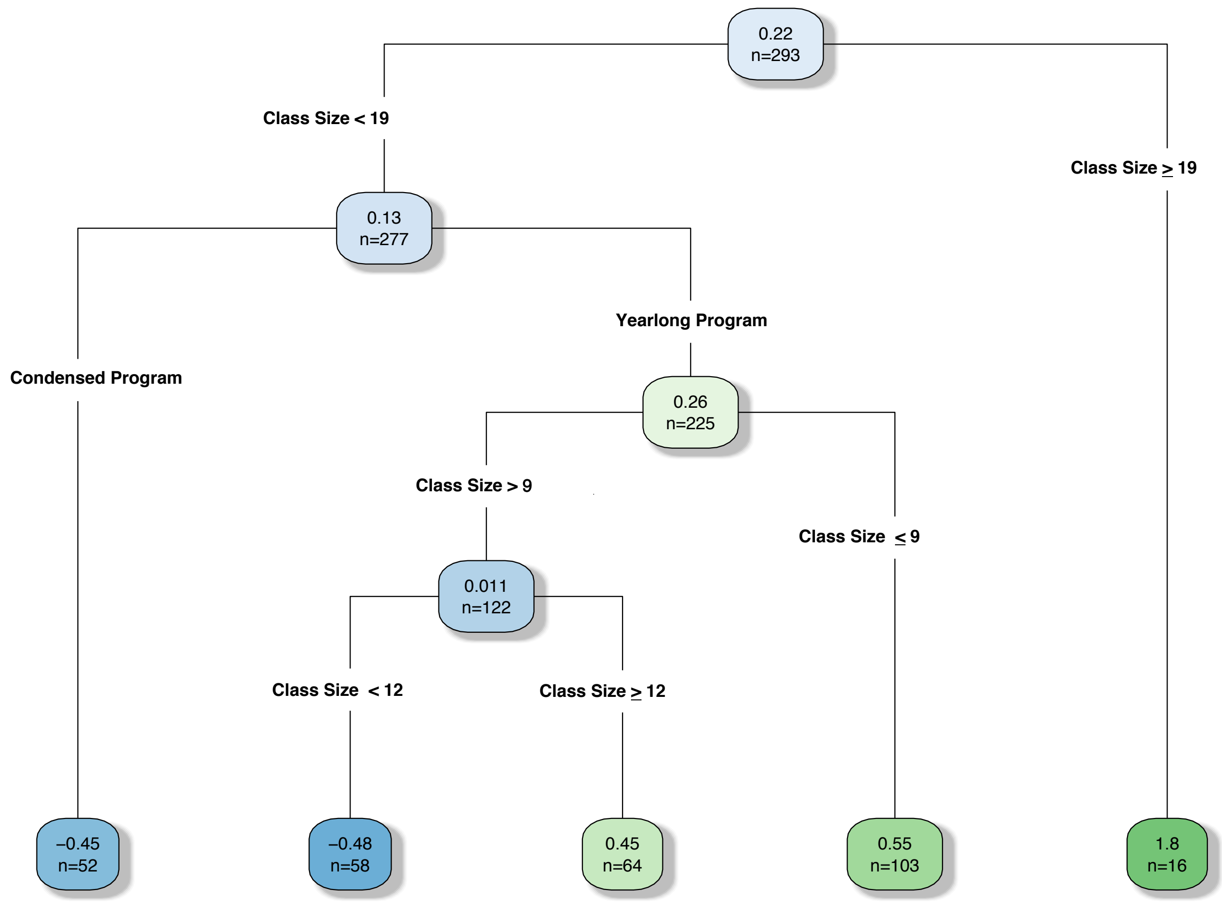


**Figure 2.** Regression tree of the outcome, "Change in child BMI," within the **Full** dataset. Boxed nodes represent splitting of the data along a single variable, where the top number represents the mean value for the outcome and the lower number represents the number of individuals within that group. The shades of the nodes visually represent the mean value of the outcome from low (blue) to high (green).

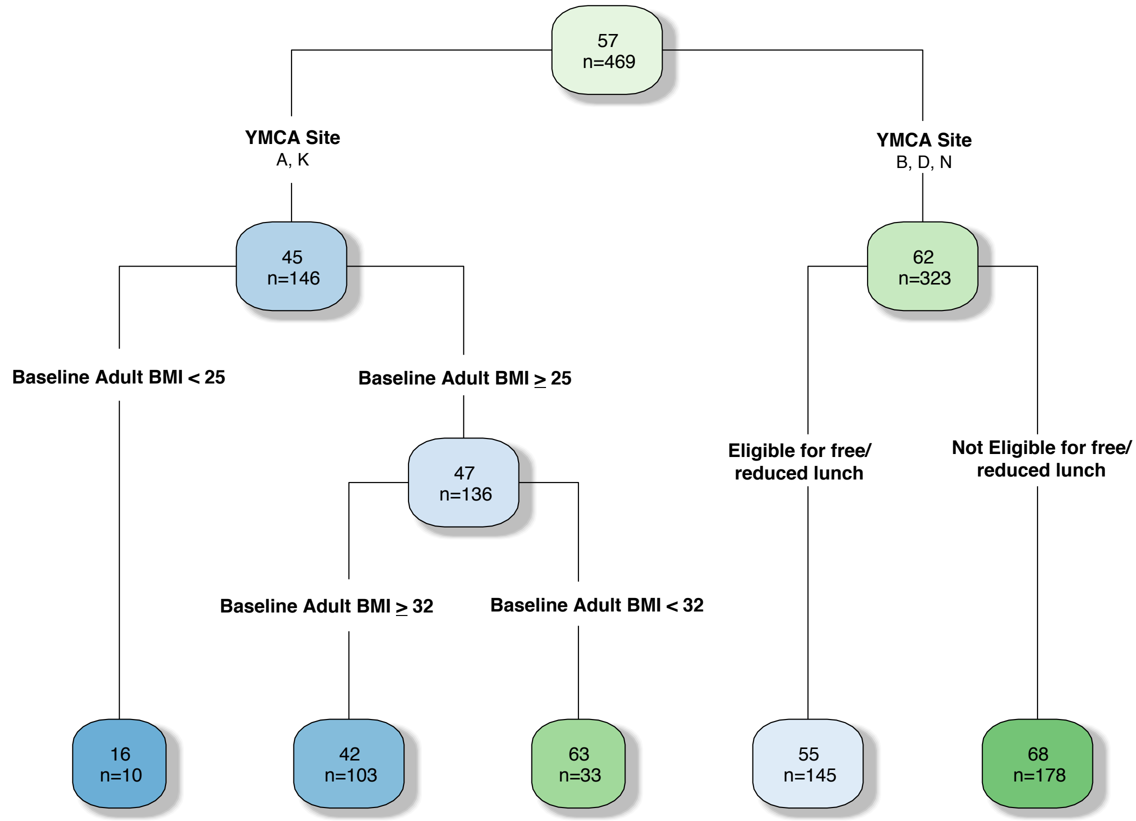


**Figure 3.** Regression tree of the outcome, "% Attendance," within the **Yearlong vs. Condensed Program** dataset. Boxed nodes represent splitting of the data along a single variable, where the top number represents the mean value for the outcome and the lower number represents the number of individuals within that group. The shades of the nodes visually represent the mean value of the outcome from low (blue) to high (green).

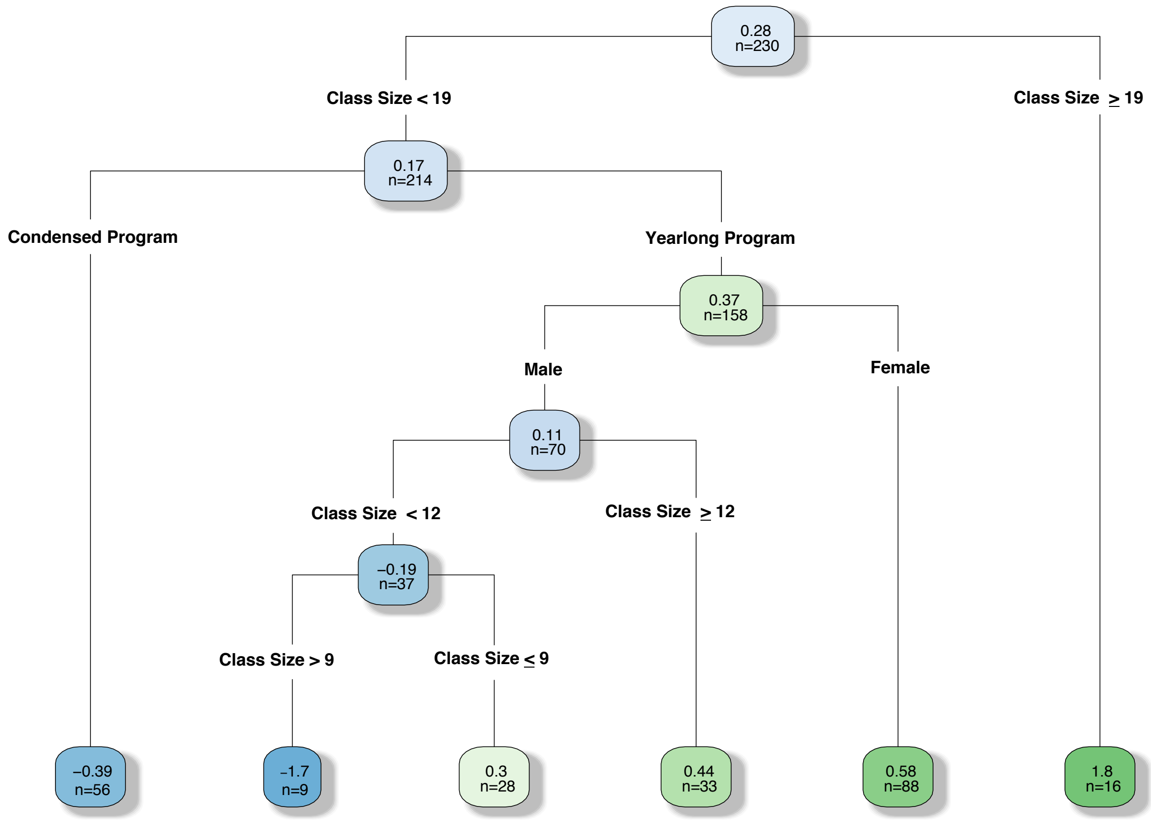


**Figure 4.** Regression tree of the outcome, "Change in child BMI," within the **Yearlong vs. Condensed Program** dataset. Boxed nodes represent splitting of the data along a single variable, where the top number represents the mean value for the outcome and the lower number represents the number of individuals within that group. The shades of the nodes visually represent the mean value of the outcome from low (blue) to high (green).

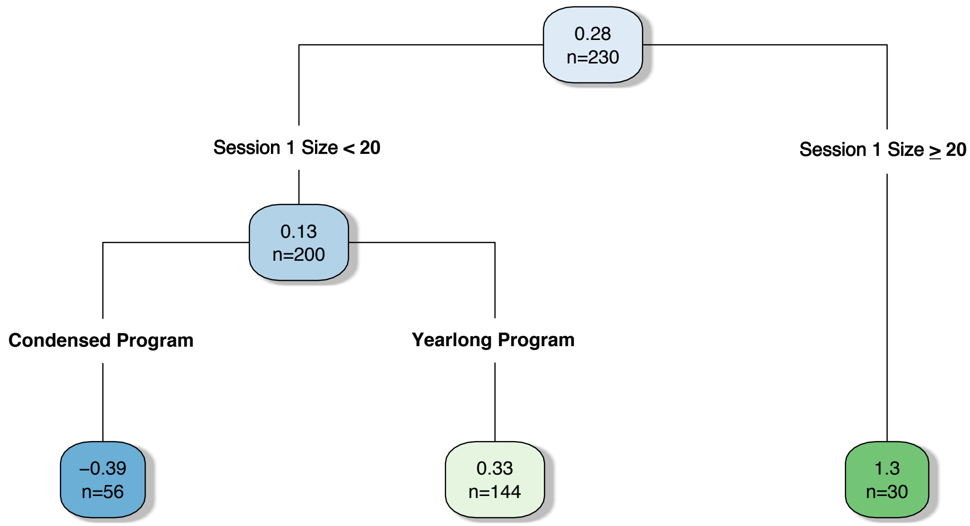


**Figure 5.** Regression tree of the outcome, "Change in child BMI," within the **Yearlong vs. Condensed Program** dataset and with inclusion of the alternate session 1 size variable. Boxed nodes represent splitting of the data along a single variable, where the top number represents the mean value for the outcome and the lower number represents the number of individuals within that group. The shades of the nodes visually represent the mean value of the outcome from low (blue) to high (green).