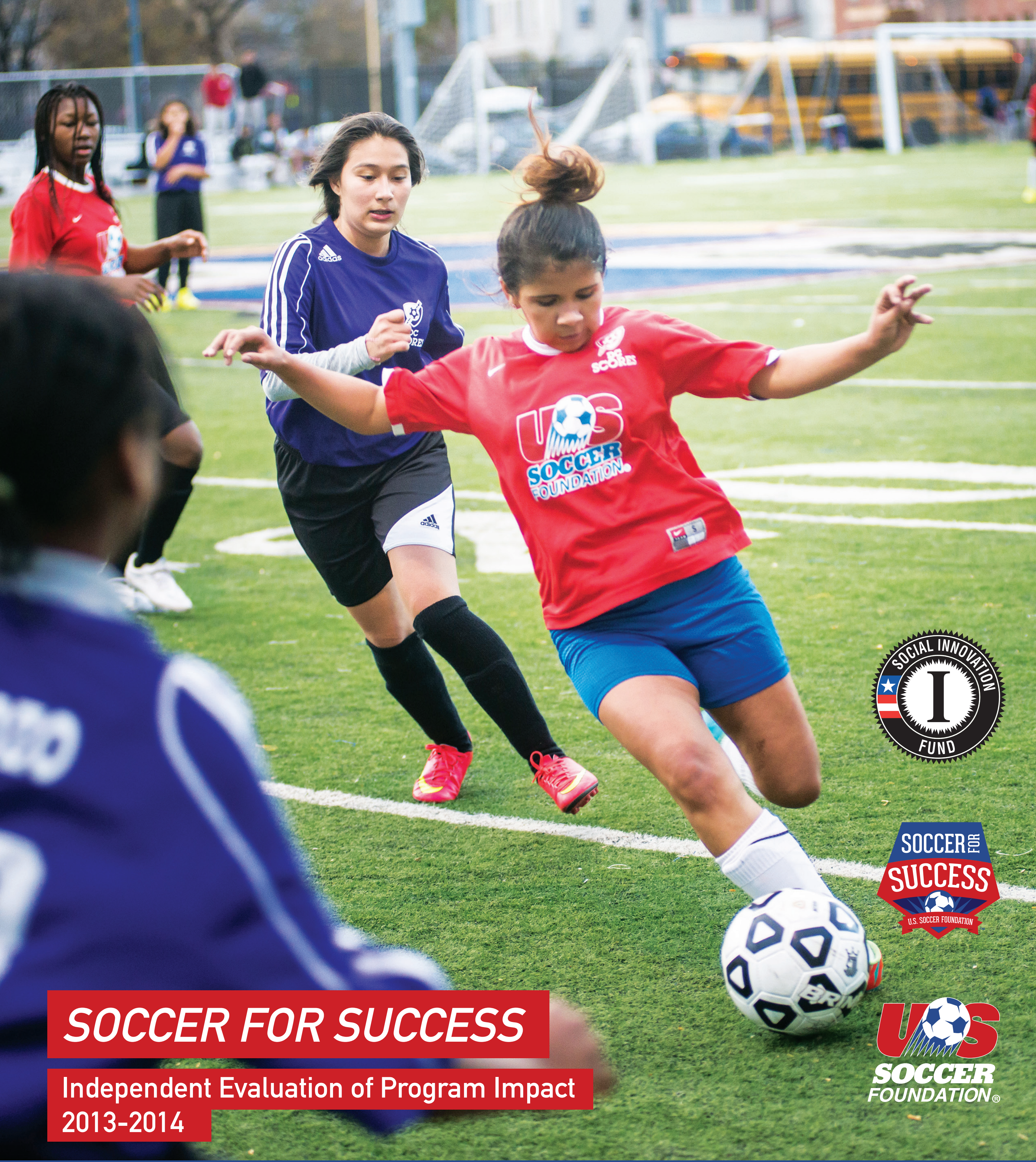


SOCCER FOR SUCCESS

Independent Evaluation of Program Impact 2013-2014

PREPARED BY:

OCTOBER 21, 2014

Dr. Danielle Hollar, PhD, MHA, MS
Executive Director
Healthy Networks Design & Research
www.HealthyNetworksDesignandResearch.org

Zach Riggle
Program Officer
U.S. Soccer Foundation
www.ussoccerfoundation.org

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The U.S. Soccer Foundation's mission is to enhance, assist, and grow the sport of soccer in the United States with a special emphasis on underserved communities. The *Soccer for Success* program advances the Foundation's mission through free afterschool soccer programming that serves children in grades K-8 in underserved urban communities. It represents an innovative approach to leveraging the appeal and growing popularity of youth soccer toward national efforts to reduce childhood obesity. Participating children enjoy learning and improving soccer skills in a low-pressure environment that differs from traditional competitive soccer leagues while also learning about healthy dietary and exercise habits. Consequently, participants increase levels of physical activity and important nutrition-related concepts which lead to improved physical fitness and reduced obesity.

During the 2013-2014 academic year the U.S. Soccer Foundation contracted Healthy Networks Design & Research to independently evaluate the impact of the *Soccer for Success* program. The evaluation consisted of a one year quasi-experimental design involving 16 randomly-assigned intervention and 14 control sites located in five cities, spread throughout the United States. The primary obesity-related outcome measures for the evaluation were age- and gender-specific body mass index (BMI) percentile, waist circumference (WC), and the Progressive Aerobic Cardiovascular Endurance Run (PACER) fitness test (completed laps) to gauge aerobic capacity/fitness.

This report is based upon work supported by the Social Innovation Fund (SIF), a program of the Corporation for National and Community Service (CNCS). The Social Innovation Fund combines public and private resources to grow the impact of innovative, community-based solutions that have compelling evidence of improving the lives of people in low-income communities throughout the United States.

SUMMARY OF RESULTS

The study results show that all obesity-related outcomes included in the design improved among *Soccer for Success* children during just one school year. Specifically, children in the *Soccer for Success* program experienced statistically significantly greater health improvement with respect to BMI percentile, waist circumference, and PACER test as compared to equivalent children in other afterschool programs in the same neighborhoods.

BACKGROUND AND RATIONALE

Between 1980 and 2008, rates of obesity among American children (ages 2-19) nearly tripled (Ogden & Carroll, 2010). While just 6.5% of children ages 6-11 were classified as obese in 1980, 19.6% were obese in 2008. Similarly, the United States experienced an increase in obesity from 5.0% to 18.1% among adolescents (ages 12-19) in the same time period. In total, 17% of America's children (roughly 12.5 million young people between the ages of 2 and 19) are at or above the 95th percentile for body mass index (BMI) according to the CDC's BMI-for-age-growth charts. Millions more are classified as overweight or at-risk for obesity (i.e. between the 85th and 95th BMI-for-age percentile).

While national trends in childhood obesity have begun to plateau in recent years, some subgroups continue to experience increases. Specifically, prevalence rates continued to increase between 2003 and 2007 among children living below 100% of the federal poverty level, African-American and Hispanic children, and publicly insured children, as overall national rates remained statistically unchanged (Bethell et al., 2010). These children and their families may lack access to viable recreational infrastructure in their neighborhoods, resources for afterschool and community-based programming that supports active lifestyles, or reliable and meaningful health information, including information about good nutrition.

Physical activity has been proposed and promoted as a key factor in the fight against childhood obesity. The White House Task Force on Childhood Obesity declared that youth participation in physical activity is a "critical national priority" and one of the "nation's leading health indicators" for the next decade (White House Task Force on Childhood Obesity, 2010). Further, empirical evidence suggests that children who regularly participate in sports or other activities outside of school, and limit their time spent watching television or playing video games, substantially reduce their odds of being overweight or obese (Bethell et al., 2010). Accordingly, efforts such as First Lady Michelle Obama's *Let's Move* campaign and the NFL's *Play 60* have emphasized regular physical activity as part of a comprehensive approach to improve children's weight-related health outcomes. Unfortunately, most children still do not receive the recommended amount of daily physical activity. In fact, less than half of all American high school students report regularly engaging in 60 minutes of moderate to vigorous physical activity (Eaton et al., 2012). If healthy leisure time habits are not instilled in school-aged children, it is expected that these trends will continue in the future.

In addition to regular physical activity, efforts to teach and foster healthy eating habits are also needed to address the nation's obesity problem. Despite an emphasis in recent years on balanced nutrition, healthy in-school food offerings and healthy snack alternatives for children, consumption patterns continue to suggest relatively high sugar, fat, sodium, and "empty calorie" intake. Trends suggest that children are consuming increasing amounts of fruit and non-fruit juices, sugar-sweetened beverages, and salty snacks, and decreasing amounts of milk, vegetables, whole grain breads, and eggs (Roblin, 2007). These patterns may be especially pronounced amongst poor and minority youth who lack access to quality food environments and knowledge of good eating habits.

Overweight and obese children face immediate and long-term health challenges, such as psychosocial problems, hypertension, high cholesterol, and diabetes (Freedman et al., 2007). Additionally, these children are more likely to suffer negative academic outcomes including disengagement, grade retention, and absenteeism (Bethell et al., 2010), all of which predict high school dropout rates and limited gainful employment prospects.

Recent research has estimated the projected high health care costs associated with unchanged trends in obesity. Wang and colleagues (2011) suggest that if obesity prevalence continues to increase, even at a relatively modest rate, adult rates may reach roughly 50% for men and women combined by 2030. The authors associate this projected increase with millions of additional cases of diabetes and heart disease, and hundreds of thousands of cases of cancer amounting to trillions of dollars in additional health care costs in the next two decades.

Fortunately, the authors also note that projections are amenable to change. Modeling based on a scenario in which the United States realizes a 1% reduction in BMI across the population (i.e., about 2.2 lbs. of weight loss for an average weight adult) reveals the possibility of avoiding 2.1-2.4 million cases of diabetes, 1.4-1.7 million cases of cardiovascular disease, and 73,000-127,000 cases of cancer over the next 2 decades. Because some coordinated efforts among elementary-aged children in the school setting have shown statistically significantly greater improvements in weight, blood pressure, as well as academic achievement among children participating in nutrition and physical activity programming as compared to those who did not (Hollar et al. 2010a; Hollar et al. 2010b,c), there is evidence that programming for young children can help address this important public health issue. Thus, the U.S. Soccer Foundation decided to take action to address childhood obesity through the creation and testing of the *Soccer for Success* program. As described below, this afterschool model results in statistically significant improvements in key obesity measures among children in the program as compared to children who are not in the program. Thus, the model shows great potential to improve the health trajectories of thousands upon thousands of young people.

SOCCER FOR SUCCESS PROGRAM

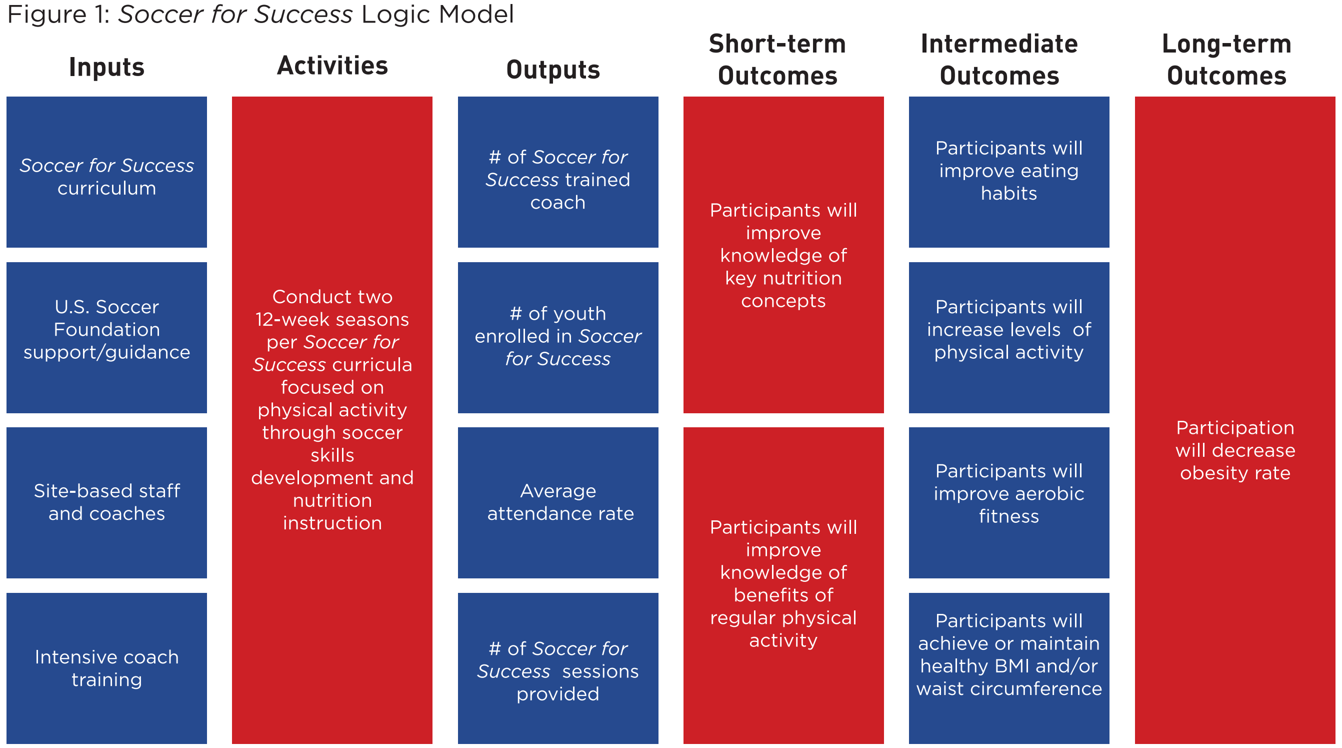
The U.S. Soccer Foundation developed the *Soccer for Success* program specifically to combat childhood obesity, promote healthy eating and exercise habits, and foster positive youth development among children in grades K-8 living in underserved urban communities. It represents an innovative approach to leveraging the appeal and growing popularity of youth soccer toward national efforts to reduce childhood obesity. Participating children enjoy learning and improving soccer skills in a low-pressure environment that differs from traditional competitive soccer leagues, while also learning about healthy dietary and exercise habits. Consequently, participants increase levels of physical activity and important nutrition-related concepts which lead to improved physical fitness and reduced obesity. The program operates 90 minutes per day, three days a week, for 24 weeks over the course of the school year. *Soccer for Success* programs host a 12-week season during the Fall and an additional 12-week season during the Spring.

EVALUATION METHODS

PROGRAM THEORY

Fundamentally, the *Soccer for Success* program is rooted in Social Learning Theory (Bandura, 1977). This theory holds that individuals learn by observing within the social context in which they live, involving several types of instruction: live, verbal, and symbolic. Learning is facilitated or impeded by factors such as attention, retention, reproduction, and motivation. *Soccer for Success* programming includes live (i.e. modeled) and verbal instruction delivered by trained coaches. Coaches are also seen as influential role models, another important concept in Bandura's theory. Repetitive, consistent instruction, through exciting and engaging soccer content, at least three times per week for 24 weeks, emphasizing four critical domains of *Soccer for Success* - physical activity, nutrition, mentorship, and family engagement – maximizes attention, retention, reproduction, and motivation. A logic model, including short-term, intermediate, and long-term outcomes, was created to guide the evaluation of *Soccer for Success* (Figure 1).

Figure 1: *Soccer for Success* Logic Model



RESEARCH QUESTIONS

- Does a greater percentage of *Soccer for Success* participants achieve or maintain a healthy BMI than youth receiving standard afterschool programming (those in the control group)?
- Does a greater percentage of *Soccer for Success* participants achieve or maintain a healthy waist circumference than youth receiving standard afterschool programming (those in the control group)?
- Do *Soccer for Success* participants experience a greater average increase in PACER test scores than youth receiving standard afterschool programming (those in the control group)?
- Are there differences in the outcomes achieved among participants who have higher BMI at baseline compared with those who have lower baseline BMI?
- Do subgroups (by gender, race/ethnicity, age) of *Soccer for Success* participants demonstrate different BMI percentile, waist circumference, PACER test, nutrition knowledge, or eating behavior outcomes than subgroups of youth receiving standard afterschool programming (those in the control group)?

DESIGN

To investigate the impact of the *Soccer for Success* program, the U.S. Soccer Foundation contracted with Healthy Networks Design & Research to serve as the external evaluator for the study component of the project. Led by Dr. Danielle Hollar, the evaluation utilized a quasi-experimental design to assess the impact of *Soccer for Success* programming in a controlled-study design, and took place during the 2013-2014 program year.

The evaluation targeted a strong level of evidence by employing a one-year, quasi-experimental design whereby data were collected at baseline and follow-up (early Fall 2013 and late Spring 2014). The methodology included 6 randomly-assigned intervention and 14 control sites located in five cities, spread throughout the United States: Buffalo, NY (4 intervention and 4 control sites); Denver, CO (4 intervention and 2 control sites); Detroit, MI (4 intervention and 2 control sites); Los Angeles, CA (1 intervention site and 1 control site); and Seattle, WA (3 intervention and 3 control sites). The five cities were selected on 1) ability to assess impact on racial/ethnic population (Table 1), 2) successful leadership of year one *Soccer for Success* activities, and 3) ability to recruit new children for the evaluation. Two groups of children in grades K-5 who participated in both baseline and follow-up data collection periods were included in the evaluation; 712 intervention children in the *Soccer for Success* program, and 522 control children who did not participate in *Soccer for Success*, but were in afterschool programs in the same neighborhood.

The protocol for selecting control sites was based on a process tested by key staff of HNDR during the Healthier Options for Public Schoolchildren (HOPS) Study (Hollar et al, 2010; Hollar et al 2010a,b). This system ensures that children in control data collection sites were similar in key characters: Gender, socio-economic status (SES), and racial characteristics of participants at the evaluation sites was representative of the United States at large.

Using a random number generator, all sites in each city were entered into the randomization program, which assigned a subset of sites into treatment/interventions. Control locations were selected based on matching the racial/ethnic and socio-economic characteristics of children in non-physical activity afterschool programs in the same neighborhood as intervention sites. All intervention and control sites that participated in data collection for the valuation study were operating under Institutional Review Board (IRB) approval, via Sterling, IRB, Atlanta, GA (www.SterlingIRB.com).

Table 1. Locations, Diversity, and Gender of Evaluation Sample

Impact Evaluation Locations	African-American	Asian	Caucasian	Hispanic/ Latino	Multi-Ethnic	Female
Brotherhood Crusade (Los Angeles, CA)	15%	0%	0%	81%	4%	33%
Colorado Rapids Youth Soccer Club (Denver, CO)	18%	0%	3%	74%	5%	44%
Detroit PAL (Detriot, MI)	57%	0%	2%	31%	10%	42%
Independent Health Foundation (Buffalo, NY)	53%	7%	14%	9%	17%	43%
Washington Youth Soccer (Seatlle, WA)	12%	12%	19%	35%	22%	36%

EVALUATION MEASURES

Based on the intermediate outcomes listed in the logic model, the primary measures of interest for the evaluation are health outcomes of age- and gender-specific BMI percentile (a calculated variable of weight in kilograms [kg] divided by height in meters squared [m2] that takes into consideration gender and age) and waist circumference. Additionally, data from the PACER fitness test was collected (completed laps). Regarding anthropometric measures specifically, baseline and follow-up anthropometric measures of all children in academic and control sites were conducted at the beginning (baseline) and at end (follow-up) of the *Soccer for Success* 2013-2014 academic year. The three core health indicators for the evaluation are described below:

- BMI percentile categories – children were classified according to their BMI percentile for age and gender in accordance with the CDC: (1) normal weight (BMI < 85th percentile); (2) Overweight (BMI > 85th percentile but < 95th percentile); and (3) Obese (BMI > 95th percentile). Height without shoes measured to the nearest 0.01 inch, weight with light clothing and without shoes measured to the nearest 0.01 pound, provides data for computations.
- Waist Circumference – waist circumference, measured to the nearest 0.01 cm was collected using a non-stretchable plastic tape measure. Waist circumference was measured at the navel at the end of gentle exhalation.
- Aerobic Capacity/Fitness – the PACER test involves running laps back and forth across a 20-meter course in time to music played from a tape or CD. The number of completed laps was recorded for this variable.

STATISTICAL ANALYSIS OF IMPACTS

Due to the nature and implementation of *Soccer for Success*, the cities included in this evaluation were not randomly selected, and children signed up to participate in the programing. However, sites within each city were randomly selected for inclusion, and control groups with similar demographic and SES characteristics, in the same neighborhoods as the *Soccer for Success* program, were selected for comparison. Children who did not have complete data for the measures of primary concern, BMI percentile, waist circumference, and PACER test, at baseline and follow-up, were not included in the analyses. Although data are collected at the individual level, the unit of assignment and the unit of analysis were at the site level. Accordingly, cluster randomization was taken into account. With cluster randomization, the mean response under each experimental condition is subject to two sources of variation: cluster to cluster and across individuals within a cluster. Approaching the analytical plan from an individual-level only, rather than a cluster-level, would not take into account the between-cluster variation and can cause an inflation of type I errors where any intervention effect may become confounded with the natural cluster-to-cluster variability. Thus, the hierarchical linear models (HLM) were applied to 1) improve estimation of effects within individual units, and 2) partition the variance and covariance components between individual level and site level.

Initially a univariate analysis consisted of simple frequency statistics for all demographic variables. Wilcoxon Signed Ranks Test and repeated measures ANOVAs were conducted to evaluate the change of dependent variables (BMI percentile category, BMI percentile, waist circumference, and PACER test score) from pre-test to post-test. Lastly, the HLM approach was applied to test the effects of independent variables on the following dependent variables: (1) BMI percentile, (2) waist circumference, and (3) PACER test score. The independent variables included (a) intervention versus control group (ARM), (b) age, (c) gender, (d) race/ethnicity, and (e) geographic location (city). Among them, geographic location (city) was a random effect, others were fixed effects. All statistical analyses were conducted using SPSS 22. All tests were two-tailed with the significant level of 0.05.

EVALUATION RESULTS

The evaluation, including the key variables and statistical procedures described, showed the *Soccer for Success* program having a statistically significant impact on obesity-related measures of participating children, as compared to controls. Demographic information, as well as the key results, are presented below. In four of the five cities (Buffalo, Denver, Detroit, Los Angeles) a data collection team comprised of *Soccer for Success* administrators and coaches collected data at control and intervention sites. All data collection teams were trained by HNDR staff prior to data collection. In Seattle, HNDR collected pre-data and a data collection team comprised of *Soccer for Success* administrators and coaches collected post-data with assistance from HNDR staff.

AMOUNT OF DATA COLLECTED

Key variables of interest collected included age, gender, height, weight, body mass index (BMI) percentile, waist circumference, and PACER test.

- Age – children's age was collected via administrative data, and used in calculation of BMI percentile and BMI classification.
- Gender – children's gender also was collected via administrative data, and used in calculations of BMI percentile and BMI classification.
- Height and Weight – children's height and weight were collected by site staff and recorded as inches (height) and pounds (weight). These measures also were used to calculate BMI percentile and BMI classification.
- BMI Percentile and BMI Percentile Categories – children were classified according to their BMI percentile for age and gender in accordance with the CDC: (1) normal weight (BMI < 85th percentile); (2) Overweight (BMI > 85th percentile but < 95th percentile); and (3) Obese (BMI > 95th percentile). Height without shoes measured to the nearest inch, weight with light clothing and without shoes measured to the nearest pound, provided data for computations.
- Waist Circumference – waist circumference, measured to the nearest 0.01 cm was collected using a non-stretchable plastic tape measure. Waist circumference was measured at the navel at the end of gentle exhalation.
- Aerobic Capacity/Fitness - The PACER test involves running back and forth across a 20-meter course in time to music played from a tape or CD. Beeps on the sound track indicate when a person should reach the ends of the course. The test begins at a slow pace, and each minute the pace increases. A participant continues running until the pace can no longer be maintained. The number of completed laps is recorded for this variable.

DEMOGRAPHIC CHARACTERISTICS OF PARTICIPANTS

Descriptive statistics, expressed as frequencies and percentages, of key participant demographics are outlined in Tables 13-17. As shown, participants were recruited from one of five major US cities with the greatest percentage of participants coming from Detroit (37.4% [Control]; 29.1% [Intervention]). Just over half of the participants in the control group were female (50.4%), whereas there were slightly more males in the intervention group (61.9%). Due to limited observations of ethnicities, this variable was recoded into three groups (i.e., African-American, Hispanic/Latino, Other).

Table 2. Participants by City

City	Control Group		Intervention Group	
	Frequency	Percent (%)	Frequency	Percent (%)
Buffalo	56	10.7	61	8.6
Denver	87	16.7	148	20.8
Detriot	195	37.4	207	29.1
Los Angeles	123	23.6	180	25.3
Seattle	61	11.7	116	16.3
Total	522	100	712	100

Table 3. Participants by Site

Site	Control Group		Intervention Group	
	Frequency	Percent (%)	Frequency	Percent (%)
Babcock Boys and Girls Club	6	1.1	0	0
Bennett Elementary	90	17.2	48	6.7
Bow Lake	26	5	35	4.9
Budlong Elementary School	0	0	180	25.3
Butler Boys and Girls Club	12	2.3	0	0
Cole Arts & Science Academy vvc	0	0	49	6.9
Dossin Elementary-Middle School	32	6.1	42	5.9
Elmwood Village Charter	0	0	11	1.5
Garden Place Academy	32	6.1	28	3.9
Gildo Rey	20	3.8	42	5.9
Greenwood Academy	0	0	50	7
Jessie Whaley Maxwell Elementary School	55	10.5	21	2.9
JFK Recreation Center @ Lanigan	12	2.4	0	0
Masten Clubhouse Elementary-Middle School	0	0	17	2.4
PS 54 Blackman	26	5	14	2
Roberto Clemente Academy	34	6.5	40	5.6
Starr Detroit Academy	39	7.5	77	10.8
Thorndyke	15	2.9	39	5.5
Watts Learning Center	123	23.6	0	0
William C. Baird Clubhouse	0	0	19	2.7
Total	522	100	712	100

Table 4. Participant Gender

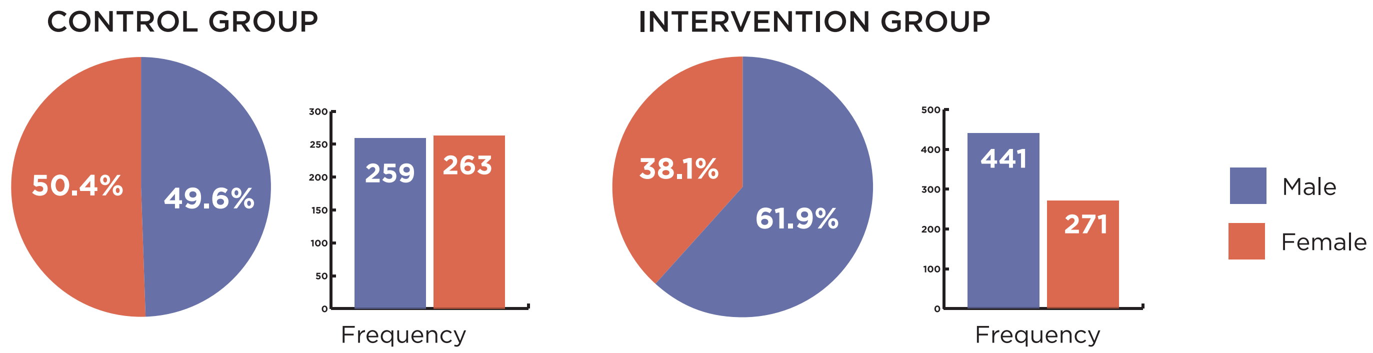


Table 5. Participant Ethnicity

Race/ Ethnicity	Control Group		Intervention Group	
	Frequency	Percent (%)	Frequency	Percent (%)
African-American	248	47.7	224	31.5
American Indian/Native Alaskan	1	0.2	1	0.1
Asian	5	1	13	1.8
Caucasian	14	2.7	31	4.4
Haitian	0	0	1	0.1
Hispanic/Latino	237	45.6	396	55.8
Pacific Islander	4	0.8	11	1.5
Other/Multi-Ethnic	11	2.1	33	4.6
Total	520	100	710	100
Missing	2		2	

Table 6. Recoded Participant Ethnicity

Race/ Ethnicity recode	Control Group		Intervention Group	
	Frequency	Percent (%)	Frequency	Percent (%)
African-American	248	47.7	224	31.5
Hispanic/Latino	237	45.6	396	55.8
Other	35	6.7	90	12.7
Total	520	100	710	100
Missing	2		2	

BASELINE EQUIVALENCE ANALYSIS

In order to assess the equivalence of participant demographics across intervention groups, a series of crosstabulations with Pearson's chi square were conducted between group and city, gender, and ethnicity. Results indicate that there were slightly more participants from Seattle in the intervention group (16.3%) compared to the control group (11.7%). There were also more females in the control group (50.4%) compared to the intervention group (38.1%). Regarding ethnicity, there were more participants who identified as African-American in the control group (47.7%) than the intervention group (31.5%). Conversely, there was a greater proportion of participants who identified as Hispanic/Latino and other in the intervention group (55.8% and 12.7%, respectively) compared to the control group (45.6% and 6.7%, respectively). *As shown in analyses below, none of these differences were found to have an effect on the dependent variables.*

To assess for baseline differences in BMI category by intervention group, another series of crosstabulations with Pearson's chi square were conducted, yielding no significant relationship. What this suggests is that there was equal representation of BMI levels across treatment groups – thus, there was no difference in baseline BMI percentile between intervention and control groups. More children in the *Soccer for Success* program, as compared to the nation at large, were in "overweight" and/or "obese" BMI categories. Specifically, 54.6% of children in the intervention group and 49.7% of children in the control group were in the "overweight" and/or "obese" BMI percentile categories combined, as compared to the national average of 33.3% for these combined categories. This is somewhat expected because the *Soccer for Success* program specifically targeted children who are higher risk for obesity. To further assess baseline differences between groups, a series of independent sample t-tests were conducted to test for differences in age, height, weight, waist circumference, and PACER test scores by intervention group.[1] Results revealed that participants in the control group were slightly taller (M = 53.22, SD = 4.93) than those in the intervention group (M = 52.27, SD = 4.66). While there were differences in height, the primary outcome measure of BMI is normed based on gender, height, and weight, which controls for these differences of height. Additionally, those in the control group had higher PACER test scores (M = 2.72, SD = .56) than those in the intervention group (M = 2.62, SD = .61).[1,2]

Table 7a. Baseline Equivalence Analyses Results

		Control Group (N=522)	Intervention Group (N=712)	
City	Buffalo	10.7%	8.6%	p value= .004 Cramer's V= .111
	Denver	16.7%	20.8%	
	Detriot	37.4%	29.1%	
	Los Angeles	23.6%	25.3%	
	Seattle	11.7%	16.3%	
Gender	Male	49.6%	61.9%	p value< .001 Cramer's V= .123
	Female	50.4%	38.1%	
Race/Ethnicity	African-American	47.7%	31.5%	p value< .001 Cramer's V= .173
	Hispanic/Latino	45.6%	55.8%	
	Other	6.7%	12.7%	
BMI Percentile Category	Underweight	2.8%	1.8%	p value= .109 Cramer's V= .070
	Health Weight	50.3%	45.4%	
	Overweight	20.2%	20.2%	
	Obese	26.7%	32.6%	

Table 7b. Baseline Equivalence Analyses Results (continued)

Group		N	Mean	Std. Deviation	T test	P value
Age at Pre-rest	Control Group	517	8.51	1.658	1.073	0.284
	Intervention Group	712	8.40	1.855		
Height Pre-test	Control Group	512	53.2172	4.93051	3.428	0.001
	Intervention Group	712	52.2690	4.65747		
Weight Pre-test	Control Group	519	81.0015	28.90516	1.041	0.298
	Intervention Group	712	79.3477	26.45916		
Waist circumference Pre-test	Control Group	518	26.7543	4.95744	-0.097	0.923
	Intervention Group	711	26.7718	4.75820		
Pacer Pre-test	Control Group	513	17.8908	11.60481	2.847	0.004
	Intervention Group	697	16.7891	12.74879		

[1]Waist circumference and PACER test scores were transformed using logarithm transformation to meet the normality assumption of t-test.
[2]While control children has "higher" results for PACER per the analyses, because the data were transformed, that actually means that they did not perform as well as the intervention children.

DIFFERENTIAL ATTRITION ANALYSIS

Differential Attrition Analysis was used to assess the potential biases due to non-consent or non-response. There are 1,538 children who did the baseline test. Among them, 1,234 children did the follow up test, and 304 children dropped. The dropout rate is 19.8%. In order to examine potential differences in dropout rates, a series of crosstabulations with Pearson's chi square were conducted between attrition (coded as completer vs. dropout) and city, gender, race/ethnicity (recoded), and intervention group. Results yielded a significant effect between dropout and city, with the highest proportion of dropouts in Buffalo (48.0%) followed by Denver (32.3%), Seattle (14.9%), Detroit (10.5%), and Los Angeles (1.9%). Another significant effect was between dropout and race/ethnicity, with the highest proportion of dropouts in Other (18.8%), followed by African American (10.8%), and Hispanic/Latino (9.1%). The dropout rates of African American and Hispanic/Latino were not significantly different. There was not a significant effect between dropout and age, gender, or treatment group, indicating that dropout appeared to be unaffected by these demographic factors.

POWER ANALYSIS FINDINGS

In addition to a prior power analyses, observed power was also calculated for primary analyses. This was conducted in order to get a precise estimate of the obtained power so that the likelihood of a Type II error could be examined.

After attrition, the current sample size is a total of 31 sites with 1234 individuals across 22 different data collection sites. That is, 16 sites with 712 subjects in the treatment group and 15 sites with 522 subjects in the control group. The standardized effect (Cohen's D) of .35 was found in the present sample. Thus the present sample size was adequate for analysis with an expected power of .90 and alpha of .05. To account for the correlated structure we use the methods provided by Hox (2002) to test this sample size for clustering by taking into account the intra-class correlation. The intraclass correlation in the present sample is .103, thus an effective sample of 240 or 15 subjects at each site (15-16 sites per treatment arm) is adequate. An achieved power of .96 was found in the present sample.

MISSING DATA ANALYSIS FINDINGS

Patterns of missing data were examined, indicating that just over 1% of the data was missing (1.18%). Further examination indicated that the data was missing at random. Overall, this suggests is that missing data was likely not a function of participant characteristics or other extraneous variables. Since the rate of missing data was very low, listwise or pairwise deletion was used on missing data. T test and Wilcoxon Signed Ranks Test used pairwise deletion, others used listwise deletion.

UNIVARIATE ANALYSIS FINDINGS

BMI Percentile Category

Wilcoxon Signed Ranks Tests were conducted to compare the BMI percentile categories cross pre- and post-tests for control and intervention groups. *The result shows that participants' BMI categories improved significantly in intervention group, p < .001, but the effect size was relatively small (r = -.11). The participants in control group did not change significantly, p = .882.*

Table 8. BMI Percentile Category Results

	Pr -				Pos -			
	Control Group		Interven ion Group		Control Group		Interven ion Group	
	Freq	(%)	Freq	(%)	Freq	(%)	Freq	(%)
Underw ight	11		11	1.	14	.	16	.
Normal W ight	256	.7	311	4	253	1.	345	4 .7
Overw ight	103	2 .4	147	.7	87	17.7	154	21.7
Obese	135	2 .7	240	3 .	137	27.	194	27.4
Total	505	100	709	10 .1	491	100	709	10 .1
Mis ing	17		3		31		3	

*Percentages may not total 100 due to rounding.

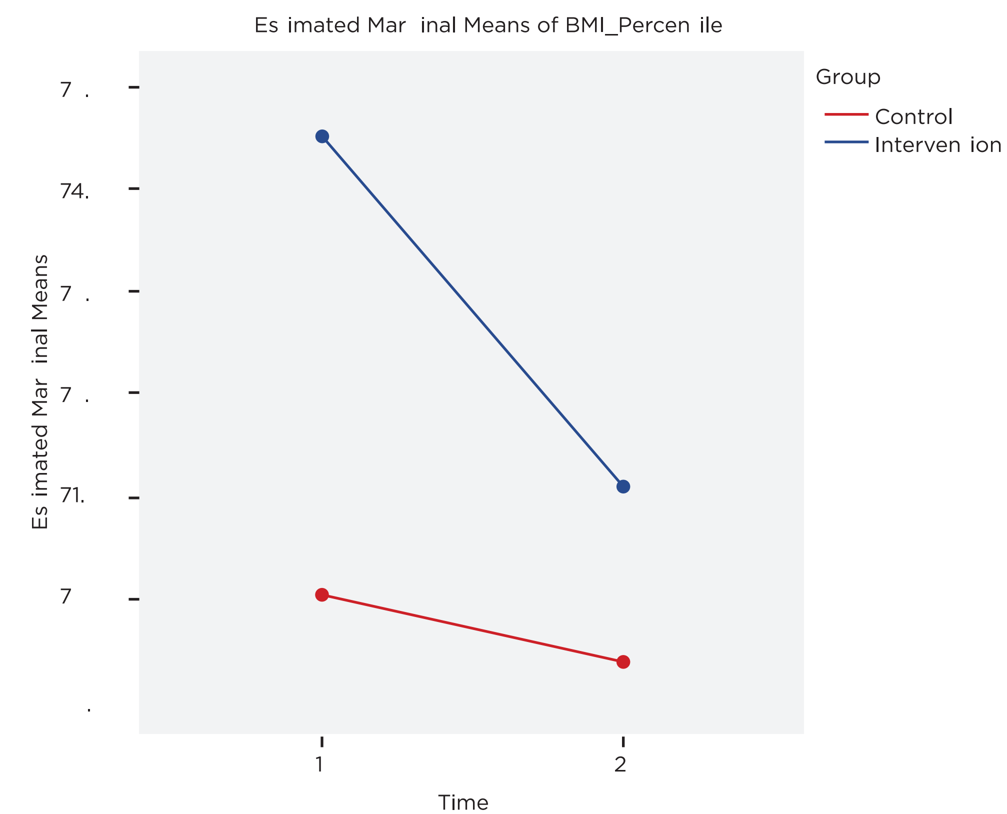
Exact BMI Percentile

Repeated measures ANOVA was conducted to test the change of BMI percentile from pre-test to post-test. The result shows that the interaction of treatment (control vs. intervention) and time (pre-test vs. post-test) was significant, p = .001, and the effect size is small (partial n^2 = .009). *The BMI percentiles of participants in intervention group decreased significantly more than those in the control group.* The line graph below shows this difference in improvement in BMI percentile for both groups – note the much greater decrease among the intervention group.

Table 9. BMI Percentile Results

Group		Mean	St . via ion
BMI Percen ile Pr -	Interven ion	74.4994	7. 45
	Control	7 .	2 . 7702
BMI Percen ile Pos -	Interven ion	71. 724	2
	Control	6 . 14	.14258

Figure 2:



Waist Circumference (WC)

Repeated measures ANOVA was conducted to test the change of waist circumference from pre-test to post-test1. The result shows that the interaction of treatment (control vs. intervention) and time (pre-test vs. post-test) was significant, p= .001, and the effect size is small (partial n^2 = .010). *The waist circumference of participants in intervention group decreased, while the waist circumference of participants in control group increased.* The graph below shows the difference in waist circumference between groups.

Table 10. Waist Circumference Results

Group		Mea in Inches (o i inal)	St . via ion (o i inal)	Mean (transformed)	St . via ion (transformed)
W ist ircumference Pr -	Interven ion	2 .7718	4.7582	. 724	.1711
	Control	2 .7543	4. 744	. 707	.1706
W ist ircumference Pos -	Interven ion	2 . 7	4.73447	. 7	.1746
	Control	2 .	. 4	. 78	.17338

[1]Waist circumference was transformed using logarithm transformation to meet the normality assumption.

Figure 3:

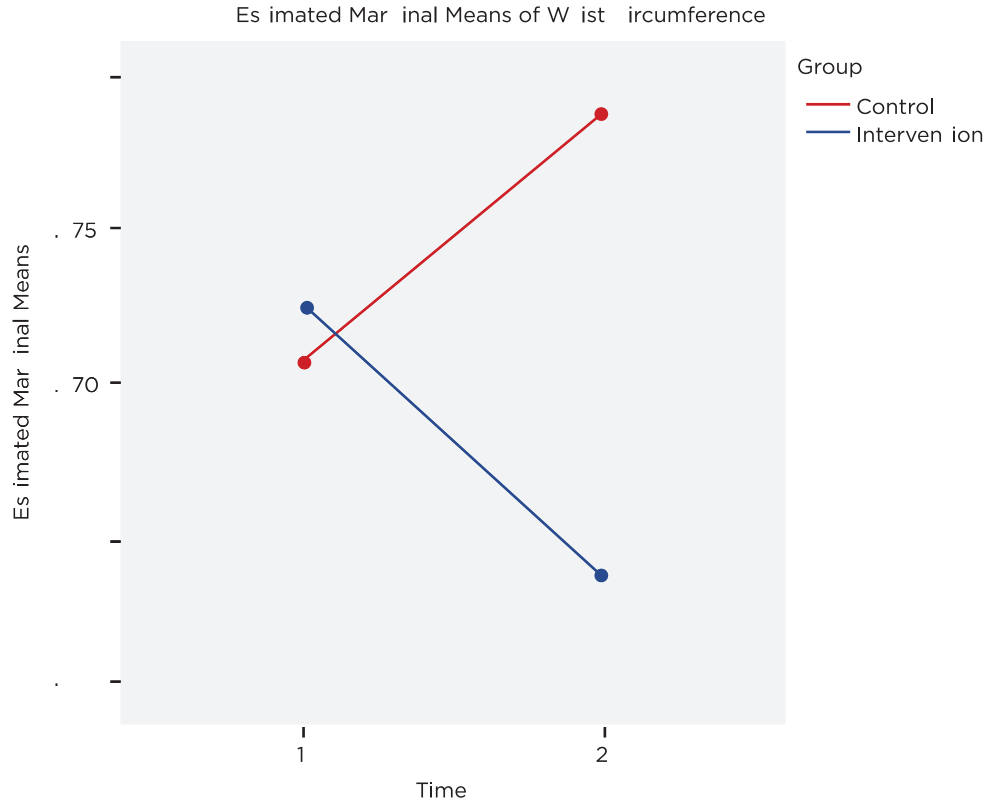
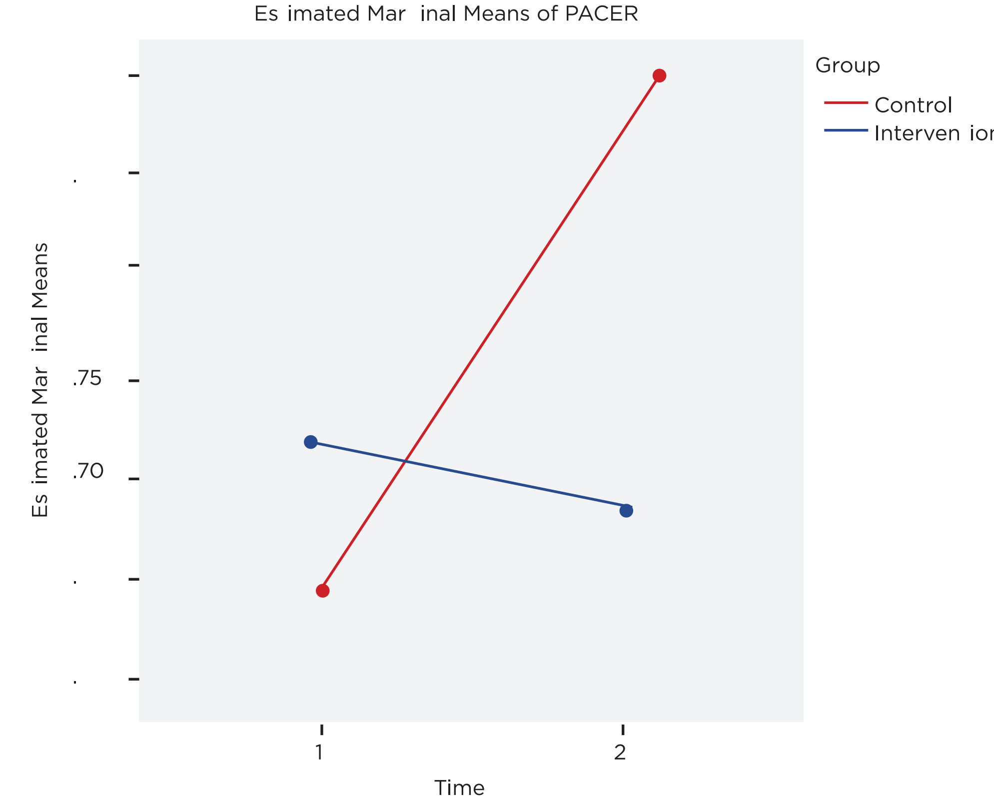


Figure 4:



Aerobic Capacity/Fitness

Repeated measures ANOVA was conducted to test the change of PACER laps completed from pre-test to post-test[1]. The result shows that the interaction of treatment (control vs. intervention) and time (pre-test vs. post-test) was significant, $p < .001$, and the effect size is large (partial $n^2 = .120$). *The PACER tests of participants in intervention group increased dramatically, while the PACER tests of participants in control group decreased slightly.* Finally, again note Figure 4 that shows the difference in results between intervention and control groups.

Table 11. PACER test Results

Group		Mean # of Laps (o i inal)	St . via ion (o i inal)	Mean (transformed)	St . via ion (transformed)
PACER Pr -	Interven ion	1 .7891	1 .74879	. 4	. 1007
	Control	17	11. 481	.7186	. 701
PACER Pos -	Interven ion	.7575	1 . 767		. 182
	Control	17.7826	1	. 4	. 49

[1]PACER test was transformed using logarithm transformation to meet the normality assumption.

HIERARCHICAL LINEAR MODEL FINDINGS

The changes of BMI percentile, waist circumference , and PACER test were calculated (post-pre). HLMs were conducted to test the effect of treatment (control vs. intervention) given the random effects of sites, and fixed effects of gender, age, and ethnicity. Since some sites have both control group and intervention group, the treatment (control vs. intervention) was analyzed as subject level predictor. Thus, cross-level interaction analysis was not necessary. For each outcome, the null model, a one-way analysis of variance (ANOVA), was fitted to partition the within- and between-group variance; then the full model containing all other predictors was fitted to examine the effects.

BMI Percentile

The null model shows that 1.7% of the variance in the difference of BMI percentile from pre to post is at the site level and the rest 98.3% is at individual level. The full model shows that among all the fixed effects, treatment ($p = .012$) and age ($p = .007$) were significant. *Keeping other variables constant, the BMI percentile of the intervention group decreased 2.57 more as compared to control group.* Keeping other variables constant, if a child's age increases 1 year, the change of BMI percentile decreased .63; thus, older children experienced more significant intervention effects than younger. However, the effect size of the full model is small: the predictors only explain 1.3% of the variance in the difference of BMI percentile from pre to post.

Table 12. HLM BMI Percentile Results

	Estimate	Std. Error	df	T	Sig.
Intercept	3.037154	2.564352	758.369	1.184	.237
Male	-1.406211	.821977	1175.679	-1.711	.087
Female (reference group)	*	*	*	*	*
Control Group	2.477137	.982895	233.111	2.520	.012
Intervention Group (reference group)	*	*	*	*	*
Age at Post-test	-6.25680	.233129	1113.345	-2.684	.007
African American	1.083863	1.583517	350.590	.684	.494
Hispanic/Latino	-2.11478	1.505124	462.094	-.141	.888
Other ethnicities (reference group)	*	*	*	*	*

*Cells intentionally left blank.

Waist Circumference (WC)

The null model shows that 2.1% of the variance in the difference of WC from pre to post is at the site level and the rest 97.9% is at individual level. The full model shows that among all the fixed effects, treatment ($p < .001$) was significant and age ($p=.057$) was marginally significant. *Keeping other variables constant, the WC of the intervention group decreased 0.59 more compared to control group.* Keeping other variables constant, if a child's age increases 1 year, the change of WC decreases .07; thus consistent with the results above regarding BMI percentile, *older children* were affected more strongly than younger by *SfS* interventions with respect to this outcome measure. However, the effect size of the full model is small: the predictors only explain 2.6% of the variance in the difference of WC from pre to post.

Table 13. HLM Waist Circumference Results

	Estimate	Std. Error	df	T	Sig.
Intercept	.431446	.402184	324.808	1.073	.284
Male	-.057861	.125414	1186.747	-.461	.645
Female (reference group)	*	*	*	*	*
Control Group	.799936	.156665	235.544	5.106	.000
Intervention Group (reference group)	*	*	*	*	*
Age at Post-test	-.068634	.035966	1156.936	-1.908	.057
African American	-.177402	.251259	320.392	-.706	.481
Hispanic/Latino	-.177459	.235886	442.367	-.752	.452
Other ethnicities (reference group)	*	*	*	*	*

*Cells intentionally left blank.

Aerobic Capacity/Fitness

The null model shows that 24.6% of the variance in the difference of PACER test (laps completed) from pre to post is at the site level and the rest 75.4% is at individual level. The full model shows that among all the fixed effects, only treatment ($p < .001$) was significant. *Keeping other variables constant, the PACER (laps) of the intervention group increased 3.8 more compared to control group.* However, the effect size of the full model is small: the predictors only explain 2.6% of the variance in the difference of WC from pre to post.

However, the effect size of the full model is small: the predictors only explain 0.8 % of the variance in the difference of PACER test (laps completed) from pre to post.

Table 14. HLM PACER Results

	Estimate	Std. Error	df	T	Sig.
Intercept	5.288792	2.061300	187.291	2.566	.011
Male	1.032327	.556690	1100.818	1.854	.064
Female (reference group)	*	*	*	*	*
Control Group	-2.981062	.750527	983.484	-3.972	.000
Intervention Group (reference group)	*	*	*	*	*
Age at Post-test	-.145213	.161104	1110.834	-.901	.368
African American	-.198880	1.232942	1057.227	-.161	-.198880
Hispanic/Latino	.339824	1.133907	1085.430	.300	.339824
Other ethnicities (reference group)	*	*	*	*	*

*Cells intentionally left blank.

DISCUSSION

The *Soccer for Success* program achieved its goal to combat childhood obesity, promote healthy eating and exercise habits, and foster positive youth development among children in grades K-8 living in underserved communities. The evaluation results show that all obesity-related outcomes targeted during *Soccer for Success* improved during just one school year, due to this innovative afterschool program.

LIST OF KEY RESULTS

- More children in the *Soccer for Success* program, as compared to the nation at large, were in “overweight” and/or “obese” BMI categories. Specifically, 54.6% of children in the intervention group and 49.7% of children in the control group were in the “overweight” and “obese” BMI percentile categories combined, as compared to the national average of 33.3% for these combined categories. This is somewhat expected because the *Soccer for Success* program specifically targeted children who are at a higher risk for obesity.
- The Foundation and its program partners in five states did an excellent job selecting intervention and control sites. Baseline equivalence analyses showed that although some differences in demographic variables were found, none of these differences were found to have an effect on the dependent variables.
- Baseline equivalence analyses also showed there was equal representation of BMI levels across intervention group – thus, there was no difference in baseline BMI percentile between intervention and control groups.
- The dataset was found to be valid: There was not a significant effect between dropout and age, gender, or treatment group, indicating that dropout appeared to be unaffected by these demographic factors. Additionally, power was adequate for analysis with an expected power of .90 and alpha of .05. An achieved power of .96 was found in the present sample. Finally, missing data was not likely a function of participant characteristics or other extraneous variables.
- All targeted obesity-related outcomes showed statistically significantly greater outcomes among intervention children as compared to controls:
    - Participants’ BMI categories improved significantly in the intervention group, $p < .001$. The participants in the control group did not change significantly, $p = .882$.
    - The BMI percentiles of participants in the intervention group decreased significantly more than those in the control group.
    - The waist circumference of participants in the intervention group decreased, while the waist circumference of participants in the control group increased.
    - The PACER tests of participants in the intervention group increased dramatically, while the PACER tests of participants in the control group decreased slightly.
    - Keeping other variables constant, the BMI percentile of the intervention group decreased 2.57 more as compared to the control group.
    - Keeping other variables constant, the waist circumference of the intervention group decreased by .21 inches more compared to the control group.
    - Keeping other variables constant, the PACER tests (laps completed) of the intervention group increased 3.8 more compared to the control group.
    - Older children experienced more significant intervention effects than younger children for BMI percentile and waist circumference.

LIMITATION

Although not a major limitation, based on the fact that all outcomes showed significant improvement among intervention versus control children, the assessment of nutrition knowledge change (indicated early on as a research question) ended up not being feasible. This was due to the fact that the assessment tool was not age-appropriate for the evaluation sample (grades K-5). After discussions with the U.S. Soccer Foundation, HNDR agreed to ask sites to attempt administration of the survey with grades 3-5. During site visits, HNDR staff observed many problems in administration of the survey. For example, most children did not have a literacy skill set to read the questions, so staff had to read them to them, which creates extraordinary bias in responses. Thus, HNDR requested of the Foundation to remove this variable in the evaluation. However, anecdotal evidence suggests that *Soccer for Success* children learned much about nutrition, and accordingly improved their eating habits.

KEY LESSONS

- Lesson 1: The fact that these obesity-related outcomes were achieved, despite the program only operating 90 minutes for only three days a week, and for just 24 weeks of the year, shows great promise of efficacy for afterschool programs that do not operate daily, if similar structures and processes are in place as in *Soccer for Success*.
- Lesson 2: Many contend that children need to partake in 60+ minutes of moderate to vigorous exercise daily to combat obesity. On the contrary as shown in *Soccer for Success*, a more reasonable, achievable, and sustainable programming plan including short bouts of exercise, coupled with nutrition education and nutrition snacks, during a 90 minute, 3-days-per-week program model shows significant improvements in obesity risk factors.
- Lesson 3: Lessons learned about exercise and eating habits appear to have been absorbed by *Soccer for Success* children, and adopted outside of their afterschool program – if not, it is not likely that the interventions would have shown such significant effects as compared to children in non-*Soccer for Success* afterschool programs. Likely, the enthusiasm of coaches and the structured, easy-to-use booklet style of the coaches guide for daily drills and nutrition lessons, made the lessons “stick.”

CONCLUSION

In conclusion, the success of *Soccer for Success* calls for national expansion of efficacious, coordinated afterschool efforts, such as the *Soccer for Success* program and perhaps other youth sports as well, that include physical activity/exercise, nutrition education, and healthy snacks to combat the public health issue of childhood obesity.

RECOMMENDATIONS

With regards to quantitative evaluation activities, we suggest two components to be run tangentially: 1) assessment/data collection of key obesity-related outcome measures (BMI percentile, waist circumference, and PACER test) in all new participants in the program; and 2) ongoing assessment/data collection of children in a) the 2013-2014 intervention and control sites, in a smaller substudy sample, and b) inclusion of all children in the program post their year 1 experiences to understand if health improvements experienced remain, or show better results year after year. A 3-year tracking program would be ideal.

FOOTNOTES

1. After intervention sites were randomly identified and engaged, the task was to identify and engage “matched” controls. In some cases, there were control afterschool program groups that were larger than predicted, meaning they that had more children in the programs that expected. Thus, fewer control sites were needed in some locations resulting in fewer control sites and intervention sites overall.

2. All *Soccer for Success* partners in the five cities engaged intervention and control groups. In some locations, the control group children included participants in an afterschool program in the same site (same school, Boys & Girls Club, etc.). In these cases, the site would be considered to have a large afterschool set of children in a diverse set of activities. In other locations, the control group of children included participants from “around the corner” in the same neighborhood as the *Soccer for Success* intervention site. In these instances, which were very few, the *Soccer for Success* intervention sites did not have large numbers of children in non-*Soccer for Success* activities. Thus, recruitment for control group participation was in afterschool programs nearby. This type of a situation is expected in community-based research, and was shown to not affect results – baseline characteristics between intervention and controls were not different (please refer to “Baseline Equivalence Analysis” section).

REFERENCES

Bandura, A. (1977). Social Learning Theory. New York: Prentice Hall.

Bethell, C., Simpson, L., Stumbo, S., Carle, A. C., & Gombojav, N. (2010). National, state, and local disparities in childhood obesity. Health Affairs, 29(3), 347–356. doi:doi: 10.1377/hlthaff.2009.0762

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Flint, K. H., Hawkins, J., & Wechsler, H. (2012). Youth risk behavior surveillance - United States, 2011 (Morbidity and Mortality Weekly) (pp. 1–162). Atlanta, GA: Centers for Disease Control and Prevention. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6104a1.htm

Freedman, D. S., Mei, Z., Srinivasan, S. R., Berenson, G. S., & Dietz, W. H. (2007). Cardiovascular risk factors and excess adiposity among overweight children and adolescents: The Bogalusa Heart Study. The Journal of Pediatrics, 150(1), 12–17.

Hollar, D., Lombardo, M., Lopez-Mitnik, G., Almon, M., Hollar, T.L., Agatston, A.S., Messiah, S.E. (2010a). Effective multilevel, multi-sector, school-based obesity prevention programming improves weight, blood pressure, and academic performance, especially among low income, minority children. Journal of Health Care for the Poor and Underserved, 21:93-108.

Hollar, D., Messiah, S.E., Lopez-Mitnik, G., Almon, M., Hollar, T.L., Agatston, A.S. (2010b). Effect of an elementary school-based obesity prevention intervention on weight and academic performance among low income children. American Journal of Public Health, 100:646-653.

Hollar, D., Messiah, S.E., Lopez-Mitnik, G., Almon, M., Hollar, T.L., Agatston, A.S. (2010c). Effect of a school-based obesity prevention intervention on weight and blood pressure in 6-13 year olds. Journal of the American Dietetic Association, 110:261-267.

Hox J.J. (2002). Multilevel analysis. Techniques and applications. New York: Routledge.

Ogden, C., & Carroll, M. (2010). Prevalence of obesity among children and adolescents: United States, trends 1963-1965 through 2007-2008 (pp. 1–5). Atlanta, GA: Centers for Disease Control and Prevention.

Ogden, C.L., Carroll, M.D., Kit, B.K., Flegal, K.M. (2012). Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. Journal of the American Medical Association, 307:483-490.

Roblin, L. (2007). Childhood obesity: Food, nutrient, and eating-habit trends and influences. Applied Physiology, Nutrition & Metabolism, 32, 635–645. doi:doi:10.1139/H07-046.

Wang, Y. C., McPherson, K., Marsh, T., Gortmaker, S. L., & Brown, M. (2011). Health and economic burden of the projected obesity trends in the USA and the UK. The Lancet, 378(9793), 815–825.

White House Task Force on Childhood Obesity. (2010). Solving the problem of childhood obesity within a generation (pp. 1–124). Washington, D.C.: White House. Retrieved from http://www.letsmove.gov/sites/letsmove.gov/files/TaskForce_on_Childhood_Obesity_May2010_FullReport.pdf.