PERCENTILE-BASED APPROACHES TO MODELING THE ASSOCIATION BETWEEN SCHOOL DAY ENERGY EXPENDITURE AND TWO-YEAR CHANGE IN BMI AMONG ELEMENTARY SCHOOL-AGED CHILDREN

A Thesis

by

GILSON DAGBEGNON HONVOH

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN PUBLIC HEALTH

Chair of Committee,	Carmen D. Tekwe
Committee Members,	Roger S. Zoh
	Hongwei Zhao
	Mark E. Benden
	Guoyao Wu
Head of Department	Jennifer Horney

December 2017

Major Subject: Biostatistics

ABSTRACT

Activity-permissive learning environments, such as school-based behavioral interventions, are increasingly employed to combat the prevalence of childhood obesity. These interventions introduce stand-biased desks into classrooms as a means of increasing school day physical activity behavior among children when compared to traditional desks and chairs. Overweight and obesity among children are defined based on age- and sex-adjusted body mass indexes (BMI) in the upper percentile ranges. However, most studies assessing impacts of interventions on BMI rely on traditional linear regression models designed to assess intervention effects on children within "normal" BMI percentile ranges, limiting assessments of how interventions affect children at higher risks for overweight and obesity. Thus, statistical approaches that permit evaluations of intervention effects across the full distribution of BMI are more desirable for determining their impacts on subjects at higher risks for developing overweight or obesity. In this thesis, we first investigate if increasing energy expenditure obtained at a prior time point such as baseline can be used as a non-invasive screening tool for predicting future obesity risk. Secondly, we describe the use of conditional functional quantile regression models to study the relationship between school day energy expenditure, a function-valued covariate, and BMI. Through empirical comparisons, we determine if results obtained from current standard approaches used in obesity research such as the multiple linear regression or functional linear regression provide notably different results from those obtained from the conditional functional quantile regression model.

ACKNOWLEDGEMENTS

Foremost I would like to express my sincere gratitude to my thesis advisor Dr. Carmen Tekwe for her motivation, patience and encouragement while working on this research. I am also greatly appreciative of her support throughout my years at Texas A&M School of Public Health. Her multiple advice will continue to guide me during the remaining of my academic studies and beyond.

I would also take the opportunity to thank the rest of my committee: Dr. Mark Benden, Dr. Hongwei Zhao, Dr. Guoyao Wu and Dr. Roger Zoh for their insightful comments and critics and their expertise, which greatly helped me strengthen my research methodology and write the present thesis.

Last but not least, I would like to thank the staff in the department of Epidemiology & Biostatistics and the School of Public Health in general for the assistance they continuously provided me with in solving diverse academic issues while attending the program.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Drs. Carmen Tekwe [advisor], Hongwei Zhao and Roger Zoh of the Department of Epidemiology & Biostatistics, Dr. Mark Benden of the Department of Environmental and Occupational Health and Dr. Guoyao Wu of the Department of Animal Science.

The data analyzed was collected under the supervision of Dr. Mark Benden for the stand-biased desk study.

All other work for the thesis was completed by the student, under the advisement of Dr. Carmen Tekwe of the Department of Epidemiology & Biostatistics.

Funding Sources

This work was made possible in part by the 2015 Research Enhancement Development Initiative (REDI), School of Public Health, Texas A&M University Health Science Center.

Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Texas A&M University Health Science Center.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

1. INTRODUCTION

About 90% of children diagnosed with type 2 diabetes are either overweight or obesity patients (Liu et al., 2010). While it is well known that obesity results from a chronic imbalance between energy expenditure and energy intake, as well as from environmental exposures and genetic predisposition, the exact role of reduced energy expenditure in obesity development is unclear (Bandini et al., 2004). To combat this growing epidemic among children, behavioral researchers are increasingly interested in employing school-based interventions as targeted interventions designed to reduce sedentary behavior among children. An example of such behavioral school-based interventions is the activity-permissive learning environment (APLE) (Wechsler et al., 2000; Benden et al., 2014; Lanningham-Foster et al., 2008). Activity-permissive learning environments introduce stand-biased desks into classrooms as a means of increasing physical activity among school-aged children. By reducing sedentary behavior, physical activity behavior is encouraged during the school day, and devices such as accelerometry devices are used to assess the behavioral patterns of physical activity. The devices provide estimates of school day energy expenditure (SDEE), the total amount of energy or calories expended by the body to perform physical activity during the school day.

Overweight and obesity in children are defined based on age- and sex- adjusted body mass indexes (BMI) in the upper percentile ranges. However, most studies assessing impacts of interventions on BMI rely on traditional linear regression models designed to assess intervention effects on children within the "normal" BMI percentile range, limiting assessments of how interventions affect children at higher risks for overweight and obesity. Thus, statistical approaches that permit evaluations of covariates effects across the full distribution of BMI are preferable for assessing their effects on subjects at higher risks for developing overweight or obesity (Koenker and Bassett, 1978). Quantile regression is a statistical technique used to estimate effects of predictors on quantile functions of a response. Examples of quantile functions include the median, the 85th or the 95th percentiles of the outcome. A drawback to the use of classical mean regression models in modeling BMI as an outcome is that these methods provide incomplete answers to questions related to extreme values of its distribution. Additionally, covariates such as SDEE and age may influence the quantile functions differently. Therefore, statistical approaches that allow one to determine the effects of covariates across the full spectrum of the conditional quantile functions of BMI are preferable in obesity studies (Bottai et al., 2014; Koenker and Bassett, 1978).

2. MOTIVATING EXAMPLE AND DATA

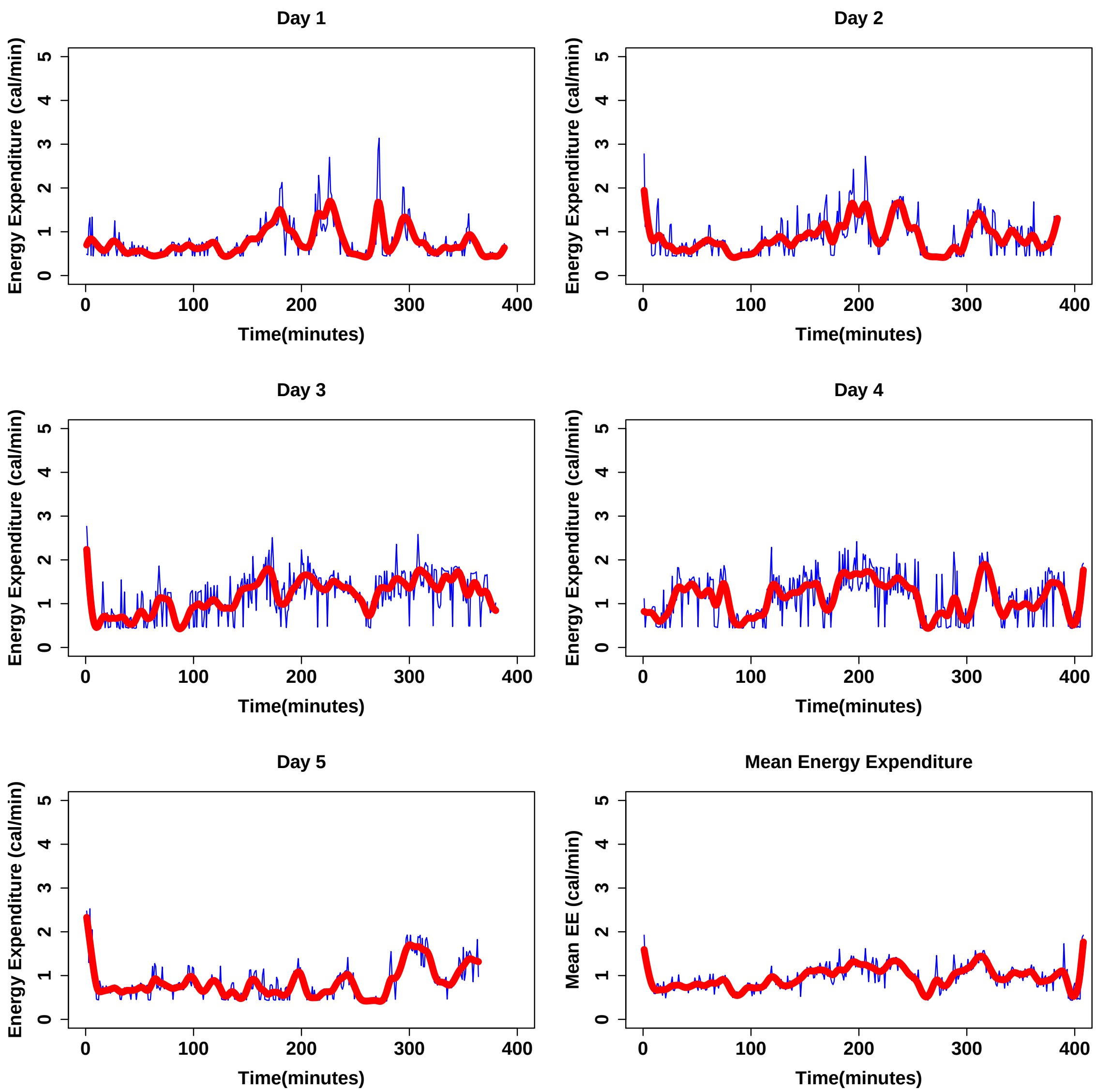
2.1 Motivating Example

Our current work was motivated by a problem in childhood obesity research. In a recent study, stand-biased desks were introduced to three elementary schools in a Texas school district as a means of increasing physical activity. A research question of interest was to examine if measures of school day energy expenditure obtained at baseline can be used as a non-invasive screening tool for future risk of obesity development. The recruited children were given BodyMedia SenseWear® Armband devices (BodyMedia, PA) to measure their energy expenditure during school hours, while sex- and age- adjusted BMI was used as an indicator for obesity. Physical activity monitoring devices are designed to measure the intensity of physical activity. Data from these devices are collected either at the second or minute level over multiple days resulting in high dimensional longitudinal data that appear as curves. Thus, SDEE data are collected over time and can easily be represented by curves rather than scalar valued summary numbers (Tekwe et al., 2017; Assaad et al., 2014; Tekwe et al., 2013; Augustin et al., 2012). Functional data analysis focuses on the analyses of experimental data collected as curves, functions or images and treats the curves as the unit of statistical analysis (Silverman and Ramsay, 2005).

Parametric regression approaches have been considered in functional data settings (Eubank, 1999). In these settings, the exact forms of the regression curves are assumed known. For example, nonlinear or polynomial mixed effects models can be used to parametrically model the effects of curves on an outcome. However, a limitation of parametric approaches to curve fitting is the requirement of strong parametric assumptions regarding the shapes of the curves. Thus, semi- and nonparametric approaches are standard approaches to analyzing functional data. These approaches provide more flexibility for fitting curves to data since they do not require a specific parametric form. Additionally, their abilities to easily accommodate the high dimensionality of functional data is desirable.

As an example, Figure 2.1 illustrates energy expenditure data gathered about every minute over five school days for a randomly selected student from our motivating example. Data like these are often summarized as a scalar-valued summary statistic such as the mean energy expenditure or the total energy expenditure in their statistical analyses (Dorminy et al., 2008; Benden et al., 2014; Wendel et al., 2016; Augustin et al., 2012). Other approaches include summarizing the data from observations taken per minute to hourly mean energy expenditures and subsequently applying standard regression approaches, such as polynomial mixed effect models Tekwe et al., 2013. However, more complex statistical data reduction techniques such as a functional principal components analysis (FPCA) or polynomial basis expansions for approximating the mean of the curves data have also been used (Silverman and Ramsay, 2005; Tekwe et al., 2017; Assaad et al., 2014; Zhou et al., 2008; James et al., 2000). Polynomial basis expansions approximate curves by describing their shapes by a few main features. Thus, an advantage of using polynomial splines is that they summarize the information contained within the curves into basis functions that adequately capture the patterns of the curves. Unlike summary statistics, such as the mean that accounts for only one source of variation in the data, each basis function accounts for a different source of variation in the data (Assaad et al., 2014). An example of such basis functions includes the B-splines (de Boor, 1978; Eilers and Marx, 1996; Rice and Wu, 2001). B-splines do not assume a specific form for the shape of the curves but rather they assume that the individual curves can be approximated by spline functions with random coefficients (Rice and Wu, 2001). In Figure

Figure 2.1: Plot of School Day Energy Expenditure and Mean Energy Expenditure Over Five Days for a Randomly Selected Subject Included in the Stand-Biased Desk Study. Blue lines indicate the observed energy expenditure while the red lines represents the smoothed version of the overall mean energy expenditure.



2.1, nonparametric smoothing was used to approximate the mean of the school day energy expenditure. By smoothing the mean, we uncover underlying patterns in the data while also retaining some of its important features (Eubank, 1999; Rice and

Wu, 2001).

The objectives of the current manuscript are two-fold. First, we examine if measures of SDEE obtained at a prior time such as baseline can be used as a non-invasive screening tool for future obesity risk, such as two years post-baseline. Secondly, we describe the use of conditional functional quantile regression models to study the relationship between SDEE and BMI, by treating SDEE as a function-valued covariate while adjusting for relevant socio-demographic variables. Through empirical comparisons, we determine if results obtained from standard approaches used in obesity research such as the multiple linear regression or functional linear regression provide notably different results from those obtained from the conditional functional quantile regression model. To the best of our knowledge, this is the first comparative analysis focused on determining the usefulness of SDEE as a non-invasive screening tool for obesity risk based among elementary school-aged children.

2.2 The Stand-Biased Desk Study

The stand-biased desk study was conducted from 2012 to 2014 in three elementary schools within the College Station Independent School District (CSISD) (Benden et al., 2014). The study has been described elsewhere, but briefly, at the beginning of the 2012-2013 academic year, 24 teachers from three elementary schools were recruited and randomly assigned to the use of either stand-biased desks (Stand2learn LLC College Station, TX, USA, stand-biased desk and stool (models S2LK04 and S2LS04, respectively)) or traditional desks (Model 2200 FBBK Series by Scholar Craft Products, Birmingham, AL), and chairs (9000 Classic Series, by Virco Inc., Torrance, CA, USA)) for in-class activities (Benden et al., 2014). A total number of 374 students from second through fourth grades were assented and included in the study at baseline. Each student's height and weight were obtained at the start of each semester by trained research assistants to calculate their BMI. The study participants were required to wear calibrated BodyMedia SenseWear® Armband devices (BodyMedia, PA) during the school hours for a week for each semester from fall 2012 to spring 2014. The devices recorded subject-specific steps counts and caloric energy expenditure per minute while worn. Of the 374 recruited students, 193 students completed the study, while the remaining either graduated from elementary school or their parents retracted their consent from the study. Five students with large proportions of missing data were excluded from the study. Thus, our final analytic sample size was 188. The study was approved by the Texas A&M IRB.

2.3 Current Approaches to Assessing the Relationship Between Objective Measures of Physical Activity and BMI

In this section, we discuss some approaches for assessing the relationship between objective measures of patterns of physical activity behavior and BMI. Wendel et al. (2016) recently used classical linear regression to analyze covariate effects on the average change in BMI percentiles from baseline. Their results indicated that the use of standing desks significantly reduced the average change in BMI percentiles when compared to conventional desks ($p = 0.04$) (Wendel et al., 2016). However, analyzing the average change in BMI percentiles does not indicate how the use of standing desks affects BMI percentile changes that are below or above the average change. Benden et al. (2014) have also shown that children who used the standing desks had a significantly higher mean energy expenditure of 0.16 kcal/min ($p < 0.0001$) when compared to students who used conventional desks. Their analyses focused on the impact of the standing desks use on the average SDEE using hierarchical linear mixed effects models (HLMEM) (Benden et al., 2014). A limitation of the use of HLMEM is that it focuses on the mean SDEE, thus only assessing impacts on the

”normal” SDEE. The HLMEM used also failed to account for the functional nature of SDEE.

Trinh et al. (2013) studied the effect of physical activity at baseline on 3-year change in BMI among elementary school-aged children in Australia. The authors provided little evidence to indicate that baseline physical activity is predictive of future risk of obesity. These conclusions were drawn based on applications of classical regression methods that treated objective measures of physical activity as a summary statistic. Specifically, the summary statistic used was the average count of steps per minute (Trinh et al., 2013). The use of summary statistics to describe the intensity of physical activity fails to account for its diurnal patterns (Davis and Fox, 2007; Augustin et al., 2012; Tekwe et al., 2013; Valenti et al., 2016; Tekwe et al., 2017). Thus, approaches that allow assessments of diurnal patterns have been considered as alternatives. Recently, Tekwe et al. (2017) used functional principal components methods to analyze energy expenditure data. This approach allowed assessments of its diurnal patterns on obesity-related outcomes. Augustin et al. (2012) also considered semi-parametric approaches to describe the patterns of physical activity. The authors used a histogram of the distribution of physical activity as a predictor in their regression model. Using data from the Avon Longitudinal Study of Parents and Children (ALSPAC), they established that their approach provided a better fit than summary statistics-based methods.

3. MODEL SPECIFICATIONS

In this section, we provide descriptions of the models considered. To fit the various models considered, we first obtained socio-demographic adjusted residuals for our outcome of interest. The residuals adjusted for baseline measures of BMI, age, sex, ethnicity, treatment assignment, as well as potential interactions between the demographic variables. Ethnicity was categorized into four categories, namely, white, Hispanic, black, and Asian/native American. Mean hourly SDEE were obtained by averaging observations obtained per minute to the hourly level across the five days that the devices were worn. The adjusted residuals were treated as the outcomes in our models while SDEE was the independent variable. All the statistical analyses were easily implemented using the R software (R Core Team, 2016).

3.1 Multiple Linear Regression Model (MLRM)

In the first model considered, we obtained a summary scalar-valued form for SDEE after obtaining its overall mean. The following model,

$$Y_i = \alpha_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \ldots, 188, \tag{3.1}$$

was specified for the MLRM. The $Y_i = residual(BMI_i)$ represents the response, α_0 is a scalar-valued intercept, β_1 represents the coefficient on SDEE. Usual assumptions associated with linear regression models were assumed. Thus, we assume $\varepsilon_i \sim N(0, \sigma_u^2)$. While the implementation of the MLRM is straightforward, it is limited in our application. It assesses impacts of covariates on the conditional expectation of Y_i, and information in SDEE is summarized into a scalar-valued covariate, leading to potential loss of information.

3.2 Functional Linear Regression Model (FLRM)

In the functional linear regression model (FLRM), the outcome considered was also $Y_i = residual(BMI_i)$. However, SDEE was treated as a function-valued covariate. The FLRM allowed a scalar-valued outcome and a function-valued covariate. Let $\{Y, X(t)\}$ be a pair of variables where Y is a scalar-valued random variable, and $X(t)$ a random function assumed to be square integrable and defined on $[0, 1]$ such that $X_i = \{X_i(t), t\epsilon[0, 1]\}$. The functional linear regression model for the ith subject at time t is specified as

$$Y_i \;=\; \alpha_0 + \int_0^1 \beta_1(t)X_i(t)dt + \varepsilon_i \quad i = 1, \ldots, 188, \tag{3.2}$$

where α_0 is a scalar-valued intercept and $\beta_1(t)$ is a functional coefficient. The $X_i(t)$ represents a function-valued covariate and $\varepsilon_i \sim N(0, \sigma_u^2)$. To implement the model, we first represent the functional component with polynomial splines. Then, $\beta_1(t)$ becomes $\beta_1(t) \approx \sum_{k=1}^{K_n} \gamma_k b_k(t)$, where γ_k are unknown spline coefficients and $\{b_k(t)\}_{k=1}^{K_n}$ are a set of known spline basis functions. The term K_n indicates the number of basis functions used to approximate the curve associated with $\beta_1(t)$. The explanatory variable, $X_i(t)$, can also be expressed as $X_{ik} = \int_0^1 X_i(t)b_k(t)dt$. The reparameterized model becomes

$$Y_i \;\approx\; \alpha_0 + \sum_{k=1}^{K_n} \gamma_k X_{ik} + \varepsilon_i \quad i = 1, \ldots, 188. \tag{3.3}$$

The reparameterized functional linear regression model in (3.3) also reduces to a multiple linear regression model. An advantage of using splines is their flexibility in capturing the patterns associated with the functional coefficient $\beta_1(t)$. This model can be easily fitted using standard software such as R (R Core Team, 2016) or SAS

(Cary, NC). To assess the effect of baseline SDEE on BMI values at two years post baseline using the FLRM, we obtained $\widehat{\beta(t)} \approx \sum_{k=1}^{K_n} \widehat{\gamma}_k b_k(t)$, while inferences were based on bootstrap point-wise confidence intervals.

3.3 Conditional Functional Quantile Regression Model (CFQRM)

The conditional functional quantile regression model (CFQRM) was applied with SDEE as a functional covariate at the 10th, 25th, 50th, 85th, 95th and 99th percentiles of the outcome variable $Y_i = residual(BMI_i)$. Following the expansion of the functional covariate using polynomial splines as presented in the FLRM case, the reparameterized model becomes

$$Q_\tau\{Y_i \mid X_i(t)\} \approx \sum_{k=1}^{K_n(\tau)} \gamma(\tau)_k X_{ik} \quad i = 1, \ldots, 188, \tag{3.4}$$

where $Q_\tau\{Y_i \mid X_i(t)\}$ represents the τth conditional quantile function for the response, $Y_i|X_i(t)$, while $\gamma(\tau)_k$ represents the kth unknown spline coefficient associated with the τth quantile. The RQ function available in the QUANTREG package in R was used to fit the model (Koenker et al., 2017). The CFQRM can also be easily fitted in SAS using PROC QUANTREG. Similar to the functional linear regression model, we also obtained $\widehat{\beta_\tau(t)}$ from our estimated coefficients $\widehat{\gamma(\tau)_k}$ by the expression $\widehat{\beta_\tau(t)} \approx \sum_{k=1}^{K_n} \widehat{\gamma(\tau)_k} b_k(t)$. Inferences for the CFQRM were also based on bootstrap point-wise confidence intervals.

The number of basis functions, K_n and $K_n(\tau)$ associated with models (3.3) and (3.4) respectively, control the smoothness of the functional covariate (Yao et al., 2005). Thus, selection of the number of basis functions is a critical step when considering nonparametric approaches for fitting curves. In our applications, we considered 4 to 7 basis functions for each model. In R, the AIC function in the LM package and the AIC.rq function in the RQ package provide AIC values, which were obtained and compared to select the best fitting number of basis functions for the FLRM and the CFQRM at each quantile.

4. RESULTS

4.1 Descriptive Statistics

Table 4.1 presents the descriptive statistics for our motivating example. The mean BMI at baseline (Fall of Year 1) was 17.18 kg/m^2 (sd = 2.79), while the mean BMI at the end of the study (Spring of Year 2) was 17.55 kg/m^2 ($sd = 3.23$). Gender was equally distributed (94 boys and 94 girls) in the study sample, and the average age of the enrolled students at baseline was 7.75 ($sd = 0.73$) years. About 75% of the students were whites, 8% Hispanics, 7.5% blacks, and 9.5% Asians/native Americans.

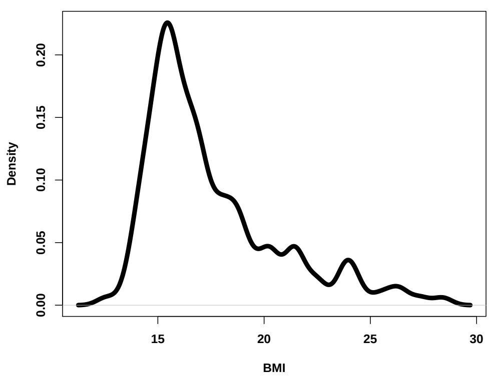
Table 4.1: Descriptive Statistics for the Study Sample (n=188). "Other"=Asians/Native Americans, SDEE= school day energy expenditure, s.d.=standard deviation of the mean.

Variable	Mean(s.d.)/ N(%)
BMI at baseline (kg/m^2)	17.18(2.79)
BMI in Spring Year 2 (kg/m^2)	17.55(3.23)
Average SDEE (cal/min)	2.48(1.34)
Age (years)	7.75(0.73)
Whites	141(75.00 %)
Hispanics	15(7.98 %)
Blacks	14(7.45 %)
Other	18(9.57 %)
Boys	94(50.00 %)
Girls	94(50.00 %)

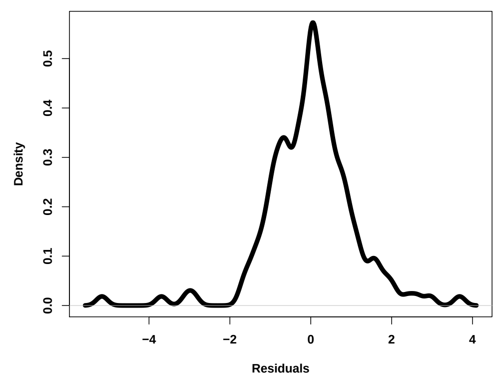
In Figure 4.1a and Figure 4.1b, we provide density plots of the BMI 24 months post-baseline and the residuals obtained after adjusting for baseline covariates. The plots illustrate the skewness of both the BMI and the residuals distributions. The shapes of these curves indicate a possible violation of the normality assumption in linear regression models.

Figure 4.1: Density Plot of BMI Distribution 24 Months Post Baseline (Figure 2a) and Density Plot of the Baseline Covariates Adjusted Residuals (Figure 2b). The skewness of the distribution of the BMI outcome and the adjusted residuals are outlined in the two plots.

(a) BMI distribution of the study population 24 months post baseline.



(b) Distribution of the residuals adjusted for baseline covariates.



4.2 Results from MLRM

The MLRM requires summarizing the high dimensional measures of SDEE per subject to a scalar-valued measure. We obtained this summary scalar by computing the arithmetic mean of all measures of SDEE by subject. From the MLRM results, we concluded that mean SDEE was significantly associated with BMI 24 months post-baseline ($\widehat{\beta_1(t)} = 0.54$, 95% CI: 0.04-1.05, $p = 0.0356$). Thus, application of the MLRM indicated that overall mean energy expenditure obtained at baseline could be used as a predictor of future values of the conditional mean of BMI after adjusting for the socio-demographic covariates ($p = 0.0356$). The MLRM produced an AIC of 472. While the use of the overall mean SDEE to represent patterns of school day physical activity behavior results in loss of information, functional regression models correct for this loss of information by using the full profile of the function-valued covariate in the estimation process (Augustin et al., 2012; Durá et al., 2012; Ratcliffe et al., 2002).

4.3 Results from FLRM

Energy expenditure measures obtained at baseline were summarized using four basis functions in the FLRM considered. The final number of basis functions were selected by comparing the AIC values for the FLRM under varying number of basis functions. The computed AIC ranged between 444 and 447, with the lowest value of 444 achieved for four basis functions. The basis functions were subsequently used as explanatory variables for SDEE to fit the FLRM. Once the model was fitted, SDEE was considered statistically significant when all estimated coefficients of the basis functions yielded small p-values. Based on our estimations, we did not find evidence of a relation between SDEE and BMI adjusted for baseline covariates after the two-year period ($p > 0.05$ for all of the estimated coefficients), see Table 4.2.

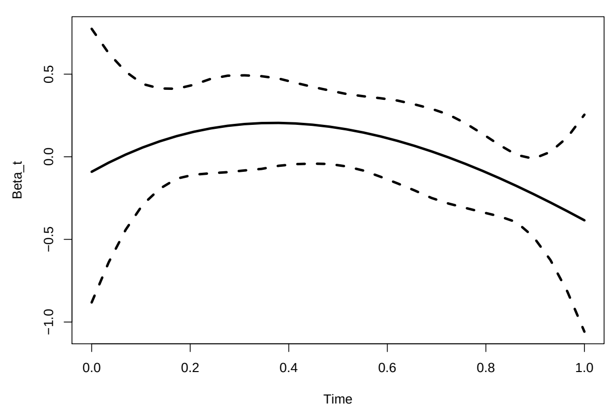
Figure 4.2 shows the estimated functional coefficient, $\widehat{\beta(t)}$ along with its 95% confidence interval. The functional coefficient was estimated from a linear combination of the estimated spline coefficients $\widehat{\gamma_k}$ and the basis functions $b_k(t)$ using: $\widehat{\beta(t)} \approx \sum_{k=1}^{k=4} \widehat{\gamma_k} b_k(t)$. Point-wise bootstrap confidence intervals were also obtained at the 95% confidence level. The estimated functional coefficient illustrates the curvilinear pattern for the energy expenditure estimate over the time, indicating that the pattern of physical activity is not constant across time. The AIC obtained from the FLRM was 444, which was smaller than the value of 472 obtained for MLRM. Thus,

Table 4.2: Results from the FLRM to Assess the Effect of Energy Expenditure (as a Functional Independent Variable) on Two-Year Change in BMI. S.E.=Standard Error. $\widehat{\gamma}_1,\ldots,\widehat{\gamma}_4$ are estimates for the spline coefficients of energy expenditure.

	Estimate	S.E.	P-value
$\widehat{\gamma}_1$	-0.11	0.38	0.771
$\widehat{\gamma}_2$	0.42	0.82	0.605
$\widehat{\gamma}_3$	0.23	0.74	0.753
$\widehat{\gamma}_4$	-0.40	0.31	0.192

the FLRM provides both a better fit and more flexibility in the estimations when compared to the MLRM in our application. The confidence intervals also support the conclusion that there is insufficient evidence in our data to indicate that SDEE is predictive of the BMI values at two years post-baseline among the children.

Figure 4.2: Plot of the Conditional Mean Function of the Energy Expenditure Estimate and its Confidence Interval. The plain line indicates the conditional quantile function while dashed lines represent the upper and lower bound of the confidence interval for each quantile function.



4.4 Results from CFQRM

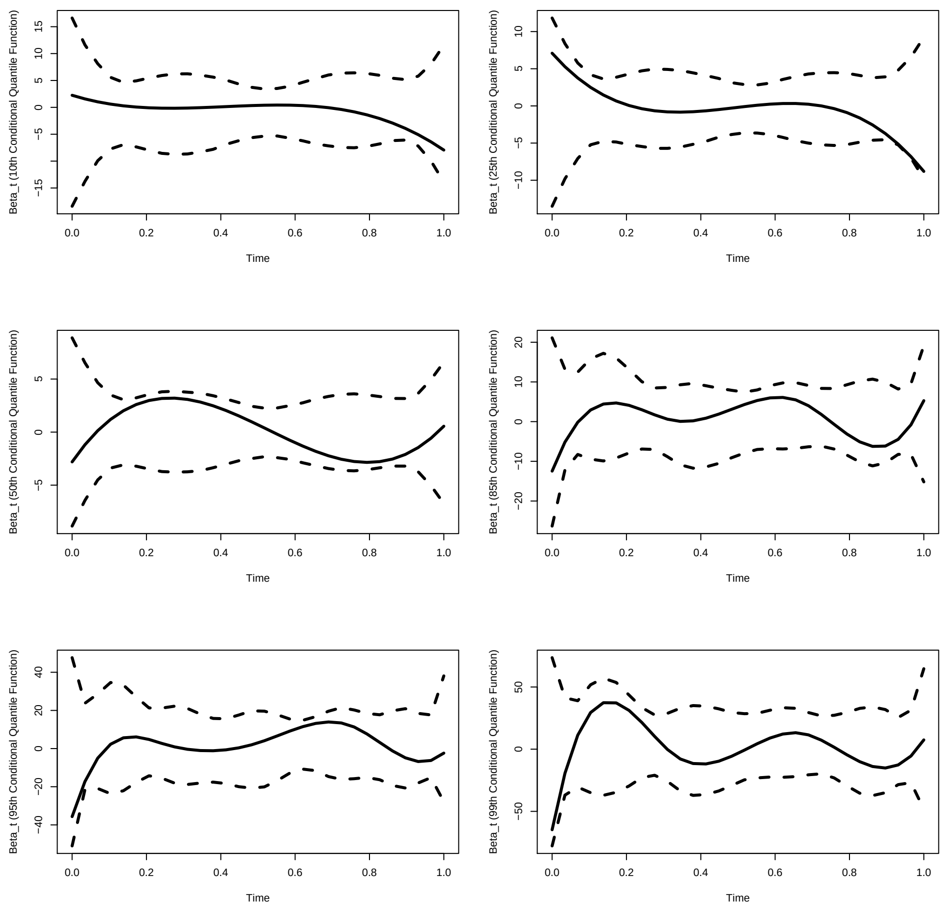
At each quantile, measures of energy expenditure were reduced into linear combinations of splines and basis functions. Similar to the FLRM, the final numbers of basis functions were selected by comparing the AIC values computed under varying number of basis functions at each quantile. The AIC comparisons led to the choice of $K_n = 4$ at the 10th, 25th and 50th quantiles, $K_n = 6$ at the 85th and 99th quantiles, while $K_n = 7$ was selected at the 95th quantile. The AIC values were higher for extreme quantiles. Table 4.3 displays the results and AICs produced by the CFQRM at each quantile. We did not detect any statistical significant association between SDEE and the conditional quantile functions of our response across all the quantile regressions considered ($p > 0.05$ for most spline coefficients).

Figure 4.3 provides plots of the estimated functional coefficients on SDEE and their corresponding 95% point-wise confidence intervals. The plots also illustrate the patterns of physical activity behavior across time under each quantile regression. Each of the six quantile functions shows a different pattern.

Table 4.3: Results from the CFQRM to Assess the Effect of SDEE (as a Function-Valued Covariate) on the Percentiles of BMI Two Years Post Baseline. Est.= spline coefficient estimate, S.E.=standard error. AIC= Akaike Information Criterion. $\widehat{\gamma_1}$,...,$\widehat{\gamma_7}$ are estimates for the spline coefficients of energy expenditure.

		$\widehat{\gamma_1}$	$\widehat{\gamma_2}$	$\widehat{\gamma_3}$	$\widehat{\gamma_4}$	$\widehat{\gamma_5}$	$\widehat{\gamma_6}$	$\widehat{\gamma_7}$
10th quantile	Est.	2.25	-4.94	7.82	-7.94	–	–	–
	S.E.	5.34	14.05	14.45	4.93	–	–	–
	P	0.675	0.726	0.589	0.109	–	–	–
	AIC	549						
25th quantile	Est.	7.09	-11.82	11.88	-8.81	–	–	–
	S.E.	6.04	15.52	15.71	5.52	–	–	–
	P	0.242	0.448	0.451	0.112	–	–	–
	AIC	482						
50th quantile	Est.	-2.81	14.25	-11.68	0.55	–	–	–
	S.E.	7.86	17.40	15.12	5.40	–	–	–
	P	0.721	0.414	0.441	0.918	–	–	–
	AIC	446						
85th quantile	Est.	-12.46	15.12	-11.94	20.08	-18.52	5.28	–
	S.E.	12.63	16.86	17.76	17.65	15.27	7.94	–
	P	0.325	0.371	0.503	0.257	0.227	0.507	–
	AIC	500						
95th quantile	Est.	-35.66	16.91	-2.97	-1.84	28.16	-16.40	-2.38
	S.E.	9.61	12.70	13.36	11.19	14.11	12.41	7.33
	P	0.000	0.185	0.825	0.870	0.048	0.189	0.746
	AIC	558						
99th quantile	Est.	-64.76	107.29	-73.73	62.47	-44.63	7.46	–
	S.E.	25.22	39.52	36.28	35.38	28.94	17.09	–
	P	0.011	0.007	0.044	0.079	0.125	0.663	–
	AIC	601						

Figure 4.3: Plot of the Estimated Functional Coefficients and Their Corresponding 95% Point-Wise Bootstrap Confidence Intervals at the 10^{th}, 25^{th}, 50^{th}, 85^{th}, 95^{th} and 99^{th} Quantiles. For each plot, the plain line indicates the conditional quantile function while dashed lines represent the upper and lower bound of the confidence interval for each quantile function.



5. DISCUSSIONS AND CONCLUSIONS

Three regression approaches were used to assess the impact of baseline school day energy expenditure on BMI values at two years post-baseline. The use of splines in the models considered enabled flexible assessments of the impact of patterns of physical activity behavior across time. Unlike the spline methodology employed, the use of the overall mean school day energy expenditure to represent physical activity behavior at baseline resulted in loss of information. While the MLRM and FLRM enable the evaluation of covariates on the conditional mean of BMI, the CFQRM enables assessment of its impact across its full distribution. MLRM indicated that the average hourly energy expenditure was associated with BMI two years post-baseline, while the FLRM and the CFQRM did not. AIC values ranged between 444 and 601 under the FLRM and CFQRM, while the MLRM had an AIC of 472. This study illustrates the use of various statistical approaches to assess the relationship between a function-valued covariate and BMI. Overall, we recommend the use of quantile based regression approaches for assessing behavioral interventions on BMI.

Despite its strengths of the statistical methods used, we identified a few limitations to this study that might impact our results. A larger sample size would improve our results in the study. Also, some factors such as diet and socioeconomic status were not measured during the stand-biased desk study and therefore, were not taken into account in our analysis. Such factors could have some significant effects on our results. Finally, it would have been of interest for researchers to assess the predictability of SDEE on BMI beyond the two-year time frame of the stand-biased desk study, as detecting future risks of obesity at a very early point in time will permit to design and implement effective interventions.

REFERENCES

Assaad, H., Yao, K., Tekwe, C. D., Feng, S., Bazer, F. W., Zhou, L., Carroll, R. J., Meininger, C. J., and Wu, G. (2014). Analysis of energy expenditure in diet-induced obese rats. *Frontiers in Bioscience (Landmark edition)*, 19:967–985.

Augustin, N. H., Mattocks, C., Cooper, A. R., Ness, A. R., and Faraway, J. J. (2012). Modelling fat mass as a function of weekly physical activity profiles measured by actigraph accelerometers. *Physiological Measurement*, 33(11):1831–1839.

Bandini, L. G., Must, A., Phillips, S. M., Naumova, E. N., and Dietz, W. H. (2004). Relation of body mass index and body fatness to energy expenditure: longitudinal changes from preadolescence through adolescence. *The American Journal of Clinical Nutrition*, 80(5):1262–1269.

Benden, M. E., Zhao, H., Jeffrey, C. E., Wendel, M. L., and Blake, J. J. (2014). The evaluation of the impact of a stand-biased desk on energy expenditure and physical activity for elementary school students. *International Journal of Environmental Research and Public Health*, 11(9):9361–9375.

Bottai, M., Frongillo, E. A., Sui, X., O'Neill, J. R., McKeown, R. E., Burns, T. L., Liese, A. D., Blair, S. N., and Pate, R. R. (2014). Use of quantile regression to investigate the longitudinal association between physical activity and body mass index. *Obesity*, 22(5):E149–E156.

Davis, M. G. and Fox, K. R. (2007). Physical activity patterns assessed by accelerometry in older people. *European Journal of Applied Physiology*, 100(5):581–589.

de Boor, C. (1978). *A practical guide to splines*. New York, NY: Springer-Verlag.

Dorminy, C. A., Choi, L., Akohoue, S. A., Chen, K. Y., and Buchowski, M. S. (2008). Validity of a multisensor armband in estimating 24-h energy expenditure in children. *Medicine and Science in Sports and Exercise*, 40(4):699.

Durá, J. V., Belda, J. M., Poveda, R., Page, Á., Laparra, J., Das, J., Prat, J., and García, A. C. (2012). Comparison of functional regression and nonfunctional regression approaches to the study of the walking velocity effect in force platform measures. *Journal of Applied Biomechanics*, 26(2):234–239.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, pages 89–102.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. New York, NY: CRC press.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87:587–602.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, pages 33–50.

Koenker, R., Portnoy, S., Ng, P. T., Zeileis, A., Grosjean, P., and Ripley, B. D. (2017). Package 'quantreg'. *Quantile Regression. In*, 5.

Lanningham-Foster, L., Foster, R. C., McCrady, S. K., Manohar, C. U., Jensen, T. B., Mitre, N. G., Hill, J. O., and Levine, J. A. (2008). Changing the school environment to increase physical activity in children. *Obesity*, 16(8):1849–1853.

Liu, L. L., Lawrence, J. M., Davis, C., Liese, A. D., Pettitt, D. J., Pihoker, C., Dabelea, D., Hamman, R., Waitzfelder, B., and Kahn, H. S. (2010). Prevalence of overweight and obesity in youth with diabetes in usa: the search for diabetes in youth study. *Pediatric Diabetes*, 11(1):4–11.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ratcliffe, S. J., Leader, L. R., and Heller, G. Z. (2002). Functional data analysis with application to periodically stimulated foetal heart rate data. I: Functional regression. *Statistics in Medicine*, 21(8):1103–1114.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259.

Silverman, B. and Ramsay, J. (2005). *Functional Data Analysis*. New York, NY: Springer-Verlag.

Tekwe, C. D., Lei, J., Yao, K., Rezaei, R., Li, X., Dahanayaka, S., Carroll, R. J., Meininger, C. J., Bazer, F. W., and Wu, G. (2013). Oral administration of interferon tau enhances oxidation of energy substrates and reduces adiposity in zucker diabetic fatty rats. *BioFactors*, 39(5):552–563.

Tekwe, C. D., Zoh, R. S., Bazer, F. W., Wu, G., and Carroll, R. J. (2017). Functional multiple indicators, multiple causes measurement error models. *Biometrics to appear*.

Trinh, A., Campbell, M., Ukoumunne, O. C., Gerner, B., and Wake, M. (2013). Physical activity and 3-year bmi change in overweight and obese children. *Pediatrics*, 131(2):e470–e477.

Valenti, G., Bonomi, A. G., and Westerterp, K. R. (2016). Diurnal patterns of physical activity in relation to activity induced energy expenditure in 52 to 83 years-old adults. *PloS One*, 11(12):e0167824.

Wechsler, H., Devereaux, R. S., Davis, M., and Collins, J. (2000). Using the school environment to promote physical activity and healthy eating. *Preventive Medicine*, 31(2):S121–S137.

Wendel, M. L., Benden, M. E., Zhao, H., and Jeffrey, C. (2016). Stand-biased versus seated classrooms and childhood obesity: A randomized experiment in texas. *American Journal of Public Health*, 106(10):1849–1854.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Zhou, L., Huang, J. Z., and Carroll, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika*, 95:601–619.