Stockholms universitet

Quantile Regression Coefficient Models to Estimate Continuous Outcomes in Epidemiologic Studies

Julia Eriksson

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2018:15
Matematisk statistik
Juni 2018

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2018:15**
http://www.math.su.se

Quantile Regression Coefficient Models to Estimate Continuous Outcomes in Epidemiologic Studies

Julia Eriksson*

June 2018

Abstract

This thesis describes quantile regression coefficient models (QRCM), a method recently developed by Frumento and Bottai [3]. This method is an extension of quantile regression, which can estimate conditional quantiles of a continuous outcome variable given covariates. QRCM specifies the coefficients of a quantile regression model as parametric functions of the order of the quantile. This thesis illustrates the use of QRCM in a study of the distribution of body mass index (BMI), defined as weight divided by height squared (kg/m^2). The data were collected by the National Health and Nutrition Examination Survey (NHANES) between the years $2015 - 2016$. The sample consisted of 8419 individuals living in the United States. QRCM were used to estimate conditional quantiles of BMI given four explanatory variables: age, race, height and gender. All these predictors appeared to be important for accurate estimations of BMI quantiles. QRCM enabled estimating reference values that could be used when assessing the BMI value of any given individual in an epidemiological or a clinical setting.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: juliaagnes.eriksson@gmail.com. Supervisor: Ola Hössjer and Disa Hansson.

Sammanfattning

Denna uppsats beskriver kvantilregressionskoefficientsmodeller (QRCM), en metod nyligen utvecklad av Frumento och Bottai [3]. Dessa modeller är en förlängning av kvantilregression, som kan uppskatta betingade kvantiler av en kontinuerlig utfallsvariabel givet kovariater. QRCM anger koefficienterna av en kvantilregressions modell som en parametrisk funktion i kvantilordningens ordning. Denna uppsats illustrerar användandet av QRCM i en studie av fördelningen av kroppsmasseindex, (BMI). BMI är definierat som vikten delat med längden i kvadrat (kg/m^2). Datan samlades in av National Health and Nuitrition Survey (NHANES) från åren $2015 - 2016$. Stickprovet bestod av 8419 individer bosatta i USA. QRCM användes för att uppskatta betingade kvantiler av BMI givet fyra förklarande variabler: ålder, etnicitet, längd och kön. Alla dessa prediktorer tycktes vara viktiga för exakta skattningar av BMI kvantiler. QRCM möjliggjorde uppskattning av referensvärden, som kan användas vid bedömning av BMI värden av någon given individ i epidemiologiska eller kliniska förhållanden.

Acknowledgements

I would like to thank my external supervisor, Professor Matteo Bottai at Karolinska Institutet, without whom I would not have been able to write this thesis. Thank you for introducing me to quantile regression, for given me a topic and data to base my thesis on, and for your time and guidance throughout the process of writing.

A warm thanks to the Unit of Biostatistics at the Institute of Environmental Medicine (IMM) at Karolinska Institutet for welcoming me to your unit and for the provision of a desk space.

At last I want to thank Ola Hössjer and Disa Hansson, my supervisors at Stockholm University, for your useful guidance and inputs to help me write this thesis.

Contents

List of Tables

List of Figures

1 Introduction

Quantile regression is an extension of linear models and is used to complement ordinary least squares regression. The story behind the development of quantile regression goes back a long time. In 1757 the priest and mathematician Boscovich fitted a straight line to a number of observations with two variables using least sum of absolute deviations. This method was later used and developed by Laplace [17]. Almost two centuries later Harris [16] found in 1950 that the problem with minimum absolute deviations can be transformed to Wagner's [28] theories on the linear programming problem, developed in 1959. This method has been used frequently because it has become popular to use robust methods and extreme value modeling. Quantile regression methods are introduced by Cameron and Trivedi [6] in 2005, Hao and Naiman [15] in 2007, and Wooldridge [30] in 2010.

In this thesis we use the definition of quantile regression presented by Koenker and Bassett [20] in 1978. They introduced quantile regression as an extension to ideas that were already existing. They suggested estimating the conditional quantile function models as functions of observed covariates.

There are advantages in using quantile regression instead of the more common ordinary least square (OLS) regression, as explained by Lê Cook and Manning [23] in 2013. When using OLS we study only the mean of the distribution and not the upper and lower tails of the distribution. An advantage of using quantile regression is that it is more robust when dealing with data that have skewed error terms. With quantile regression one can obtain a complete picture of the distribution as one can study the tails of the distribution as well as all quantiles in-between.

Frumento and Bottai [3, 4] extended quantile regression and developed quantile regression coefficient models (QRCM). The hallmark of this method is that the regression coefficients are modelled as parametric functions in the order of the quantiles. Frumento and Bottai found that using this method might be preferable with respect to efficiency and parsimony and it might help expand the potential of statistical modeling.

The data used in this thesis are from the National Health and Nutrition Examination Survey (NHANES) [7]. The data contain the explanatory variables gender, age, race and height; and the dependent variable body mass index (BMI). BMI values is used when examining body weight of a person. The BMI of a person is defined by a person's weight and height, BMI=kg/m^2, and therefore it is a measurement quite easy to obtain. In clinical and epidemiological settings, BMI values are often compared with reference values and growth charts.

In this thesis we use Frumento and Bottai's approach on data from the National Health and Nutrition Examination Survey [7] to estimate reference values and use them to assess BMI values of individuals of any given quantile.

1.1 Structure of the thesis

The structure of this thesis is as follows. Section 2 describes the basic properties of quantile regression and quantile functions along with a few examples of application from earlier studies in different fields. Section 3 introduces quantile parametric models developed by Frumento and Bottai [3] in 2016. Section 4 presents the analysis of the data from NHANES [7], on BMI and four explanatory variables. In section 5 we discuss the limitations of this study and future work on the subject.

2 Background and Methods

This section outlines the theory and concepts that will be used in the thesis.

2.1 Quantile functions and Cumulative Distribution Functions

The cumulative distribution function (CDF) of a random variable X is defined as the probability that the observed value X is smaller or equals a value that we call x,

$$F_X(x) = P(X \leq x).$$

For a continuous random variable X with a continuous CDF function, the probability density function (PDF) of X is defined as the derivative of the CDF,

$$f_X(X) = \frac{dF_X(x)}{dx}.$$

The quantile function (QF) is defined as the inverse of the CDF, i.e.

$$Q(p) = F^{-1}(p) = \inf\{x : F(x) > p\}.$$

The quantiles and their ranks can be estimated through an optimization problem as described by Koenker.[19]

2.2 Quantile regression

This section follows the material presented in the book by Koenker [19]. Quantiles can be obtained by solving a minimization problem, but let us first describe how to estimate the expected value from a sample. To obtain an estimate of the expected value $E(Y)$, from a sample $Y_1, ..., Y_n$ of independent observations of random variables $Y_1, Y_2, \ldots, Y_n$ with distribution function F_Y, one can solve the minimization problem,

$$\min_{\mu \in \Re} \sum_{i=1}^{n} (y_i - \mu)^2.$$

Suppose that some of the variation in Y is explained by a vector of covariates $x = (x_1, ...x_q)^T$. To find the conditional mean $E(Y|x)$, $\mu(x) = x^T\beta$, we can estimate β by solving the same minimization problem as above, but replace μ with $x^T\beta$, i.e.

$$\min_{\beta \in \Re^q} \sum_{i=1}^{n} (y_i - x_i^T\beta)^2,$$

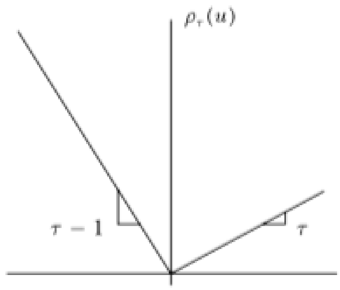
where $x_i = (x_{i1}, ..., x_{iq})^T$ is a vector of covariates for observation i.

The same procedure follows when we want to find the p sample quantile, this is obtained by solving the minimization problem

$$\min_{\alpha \in \Re} \sum_{i=1}^{n} q_p(y_i - \alpha),$$

where $q_p(\cdot)$ is the function of the tilted absolute value, with the pth sample quantile as solution. This is illustrated in Figure 1.

Figure 1: Quantile regression q function. The slope of the line to the right represents the pth quantile (denoted τ in the picture), whereas the slope of the line to the left is the $p-1$th quantile (denoted $\tau-1$ in the picture). The notion denoted ρ_τ represents the $q_p(\cdot)$ function.



We can solve the conditional quantile function for the pth quantile $Q_y(p|x) = x^T\beta(p)$ as

$$\min_{\beta \in \Re^q} \sum_{i=1}^{n} q_p(y_i - x_i^T\beta).$$

A regression model with independent and identically distributed error terms u_i is written as follows

$$Y_i = \beta_0 + x_i\beta_1 + u_i,$$

where β_0 is the intercept term. Then the quantile functions of Y can be written as

$$Q_y(p|x) = \beta_0 + x\beta_1 + F_u^{-1}(p),$$

where F_u is the common distribution function of the error terms u_i. By integrating the quantile of the error term into the intercept, we can rewrite the quantile function as

$$Q(p) = \beta_0(p) + \beta_1(p)x.$$

When estimating the parameters of a linear regression model, we need to find the partial derivatives of the objective function with respect to the intercept and the effect parameters. If a covariate is dependent on more than one coefficient in the model, the interpretation would be that when there is a change in a covariate it will lead to changes on all coefficients associated with the covariate. We have that

$$Q_{h(Y)}(p|X = x) = h(Q_Y(p|X = x)),$$

which holds for a monotone transformation $h(\cdot)$. It then follows that

$$Q_{h(Y)}(p|X = x) = x^T\beta(p),$$

if h is chosen as the link function for which the conditional quantile function of $h(Y)$ rather than Y, is a linear combination $x^T\beta(p)$ of the covariates. Then find the partial derivative for x_j,

$$\frac{\delta Q_{h(Y)}(p|X = x)}{\delta x_j} = \frac{\delta h^{-1}(x^T\beta)}{\delta x_j}.$$

An example with a logarithmic link function, would then be

$$Q_{\log(Y)}(p|X = x) = x^T\beta(p),$$

with the following derivative

$$\frac{\delta Q_{h(Y)}(p|X = x)}{\delta x_j} = \exp(x^T\beta)\beta_j.$$

2.3 Examples of quantile regression applications

Quantile regression is used in many fields and some empirical examples are described in this section.

2.3.1 Economics

The Nobel prize winner Angus Deaton introduced the concept of quantile regression in his book *The Analysis of Household Surveys* (Deaton, 1997) [12]. He used quantile regression when studying data from food share and total expenditure in Pakistan between the years 1984-85. In previous experience made on the topic it has been shown that the budget share a household spends on food can be approximated as a linear function of the logarithm of household expenditure per capita. The survey made by Deaton studied 9119 households on which he calculated the food share conditional on the logarithm of household experiment per head. In the result he observed the 10th, 50th and the 90th quantiles of the quantile regression. The observed slopes of the three quantiles differ where the 10th quantile had a slope of -0.121, the 50th (median) a slope of -0.094 and the upper 90th quantile had a slope of -0.054. The slopes have a wide variation, meaning that there is an increasing conditional variance of the regression in households that are wealthier. He noted that there is a large difference between the 10th and the 90th quantile among the wealthier households, Deaton concludes that it means that the households with more money tend to spend less of it on food but he also notes that there is a higher variability of tastes among them.

2.3.2 Epidemiology

A study by Zhang et al [32] in 2006 investigated the impact smoking has on the sleep architecture (a cyclical pattern of sleep that is affected by age and disorders). Quantile regression was used in the article to examine the parameters of sleep architecture and to study the differences between the groups: never smoker, former smoker and current smoker. After parameterization was made on the models, the regression coefficients compared the relative proportions of each sleep step between the reference group never smokers and the former or current smokers respectively.

2.3.3 Ecology

When studying statistical distributions of ecological data there is more than one rate of change that explains the relationship between the outcome and predictor variables. This is called unequal variation. To be able to estimate multiple rate of change, quantile regression is used, then the ecologists will learn more about all the factors affecting the organisms.[5] An example of quantile regression used in ecology is a study by Dunham et al [13] in 2002, where they used quantile regression to investigate if there are limiting relationships between the standing crop of cutthroat trout and measures of stream channel morphology. The quantile regression models showed an inverse relationship between the variation of fish density and the ratio between width and depth of streams. The study also indicated that there was no relation found between variation of fish density and the width and depth of the stream alone.

2.3.4 Medicine

Quantile regression is commonly used in medicine, and an example is a study of growth charts of children in Finland (Wei et al., 2006)[29]. Data of height and weight was collected from 2514 Finnish children, between the ages of 0 and 20 the children had been measured an average of 20 times. Their conclusion of the study was that using quantile regression when observing independent estimation growth curves is a flexible approach.

2.4 Restricted cubic splines

Restricted cubic splines are a transformation of an explanatory variable used to model the non-linear relationship between an outcome and an explanatory variable. Restricted cubic splines are used when studying non-linearity or when the relationship between the variables is too non-linear for them to be summarized meaningfully by a linear model. The method of restricted cubic splines proceeds as follows: First subdivide the range of the predictor values into subintervals with a number of knots. Then fit the regression curves/lines in between these knots. The optimal number of knots to use depends on how big the sample is. If working with a small sample, three knots are optimal to use, this to ensure that there is enough observations between the knots as needed in order to fit each polynomial. With large samples, five or more knots can be used. The location of the knots is less important than the number of them, and it is specified in advance depending on the quantile of the continuous variable. The regression curves/lines must meet at the knots in a way to that they will join smoothly. This means that for the polynomials with degree n, the spline functions are continuous at the knots, as well as the first $n-1$ derivatives. For instance, the first and second derivatives of the cubic splines are continuous at the knots, this gives them a smooth shape. Cubic splines do not have to use as many degrees of polynomials as higher order splines to still have an inflection and have flexibility when fitting the data. To fit a linear model with k knots, $k+1$ coefficients are used, whereas fitting a cubic spline needs $k+3$ coefficients plus one for the intercept.

Restricted cubic splines are often preferred to cubic splines because they constrain the fitting curve to be linear at either end of the range of the predictor variable, where ordinary cubic splines may show erratic behaviour.

3 Quantile parametric models

This section closely follows the material written in 2016 by Frumento and Bottai [3]. QRCM define the coefficients of a quantile regression model as parametric functions of the order of the quantile. This approach generally has advantages in efficiency, parsimony and it might help expand the potential of statistical modeling.

3.1 Quantile regression coefficient models

To explain this approach, we start by introducing some definitions. Denote the conditional functions of PDF, CDF and QF by $f(y|\boldsymbol{x}), F(y|\boldsymbol{x})$ and $Q(p|\boldsymbol{x})$ respectively, where Y is the variable of interest conditional on the vector $\boldsymbol{x}$ with dimension q, of observed covariates. Assume that for any $p \in (0,1)$, there exists a q-dimensional column vector $\beta(p)$ such that

$$Q(p|\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}(p), \tag{1}$$

where $\boldsymbol{\beta}(p)$ is defined as a function of p. Moreover, $\beta(p)$ depends on a finite-dimensional parameter $\boldsymbol{\theta}$, such as

$$\boldsymbol{\beta}(p|\boldsymbol{\theta}) = \boldsymbol{\theta}\boldsymbol{b}(p),$$

where $\boldsymbol{b}(p) = [b_1(p), ..., b_k(p)]^T$ is a set of k known functions of p, and $\boldsymbol{\theta}$ is a $q \times k$ matrix with entries θ_{jh}. The quantile regression coefficient that is associated with the j-th covariate is $\beta_j(p|\boldsymbol{\theta}) = \theta_{j1}b_1(p) + ... + \theta_{jk}b_k(p)$, $j = 1, ..., q$. The conditional quantile function can then be written as

$$Q(p|\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{x}^T\boldsymbol{\theta}\boldsymbol{b}(p). \tag{2}$$

To allow regression coefficients to be functions of different subsets of $\boldsymbol{b}(p)$, some entries of $\boldsymbol{\theta}$ can be set to 0. To describe how the regression coefficients depend on the order of the quantile, the value of p is varied. Then the vector $\beta(p)$ will be a set of quantile coefficient functions that describe this dependency.[4]

3.1.1 Examples of conditional quantile regression.

In equation (1) we saw the quantile regression model and now we are going to use the same model but with conditionals. The conditional quantile regression model for a single covariate x can be written as

$$Q(p|x, \boldsymbol{\theta}) = \beta_0(p|\boldsymbol{\theta}) + \beta_1(p|\boldsymbol{\theta})x,$$

where we have used that $j = 0$ for the coefficient associated with the constant term in x and $h = 0$ for the coefficient associated with the constant term in $\boldsymbol{b}(p)$.

***Example*1**
In the first example we assume that $\beta_0(p|\boldsymbol{\theta}) = \theta_{00} + \theta_{01}p$ and $\beta_1(p|\boldsymbol{\theta}) = \theta_{10} + \theta_{11}p$. The quantile function is then a uniform distribution and the support can be obtained by setting $p = 0$ and $p = 1$ where the endpoints of the support are linear functions of x. We note that if $\theta_{01} + \theta_{11}x > 0$, for all x, then $Q(p|x, \boldsymbol{\theta})$ is monotonically increasing. This will make it simple to control for quantile crossing. When $\theta_{11} = 0$ homoscedasticity is imposed and when there is no intercept, $\theta_{00} = \theta_{01} = 0$, it describes a zero-flat model where all quantiles have value 0 at $x = 0$. The model is determined by the choice of the "basis" $\boldsymbol{b}(p)$, which must be defined in advance, and by the restrictions that are imposed on θ. In the first example the restrictions are the following

$$\boldsymbol{b}(p) = \begin{pmatrix} 1 \\ p \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\theta} = \begin{pmatrix} \theta_{00} & \theta_{01} \\ \theta_{10} & \theta_{11} \end{pmatrix}.$$

This model can be used for a bounded outcome (when the measurements are in a finite interval) and then θ_{00} would be the lower bound of the support for $x = 0$.

***Example*2**
In the next example we assume that the outcome is not bounded. If $\beta_0(p|\boldsymbol{\theta}) = \theta_{00} + \theta_{01}z(p)$ and $\beta_1(p|\boldsymbol{\theta}) = \theta_{10} + \theta_{12}p$, where β_1 is linear in p, and β_0 depends on the quantile function of a standard normal distribution, $z(p)$, then this model does not correspond to any known distribution for $Y|x$, except when $x = 0$. If we were to put $\theta_{12} = 0$ the model assumptions will be that of a standard linear regression, with the coefficients $\beta_0 = \theta_{00}$, $\beta_1 = \theta_{10}$ and the residual standard deviation $\sigma = \theta_{01}$. In this second example the model restrictions are

$$\boldsymbol{b}(p) = \begin{pmatrix} 1 \\ z(p) \\ p \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\theta} = \begin{pmatrix} \theta_{00} & \theta_{01} & 0 \\ \theta_{10} & 0 & \theta_{12} \end{pmatrix}.$$

This model describes an unbounded outcome where θ_{00} is the median outcome when $x = 0$.

3.1.2 Estimator of the coefficient functions

We are now going to define an estimator which generalizes ordinary quantile regression and it is developed to estimate the coefficient functions. To be able to estimate the p-th quantile regression coefficients in model (1), we must minimize the following objective function:

$$L_n(\boldsymbol{\beta}(p)) = n^{-1} \sum_{i=1}^{n} (p - \omega_{p,i})(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}(p)), \tag{3}$$

where $\omega_{p,i} = I(y_i \leq \boldsymbol{x}_i^T \boldsymbol{\beta}(p))$, and $I(\cdot)$ is the indicator function. The gradient is a discontinuous function of $\beta(p)$ which takes the form

$$S_n(\boldsymbol{\beta}(p)) = n^{-1} \sum_{i=1}^{n} \boldsymbol{x_i}(\omega_{p,i} - p). \tag{4}$$

The Matrix $\boldsymbol{\theta}$ in the quantile function (2) can be estimated as the minimizer of the integrated objective function

$$\overline{L}_n(\boldsymbol{\theta}) = \int_0^1 L_n(\boldsymbol{\beta}(p|\boldsymbol{\theta}))dp. \tag{5}$$

This is the same as to find the zeros of its gradient

$$\overline{S}_n(\boldsymbol{\theta}) = \int_0^1 S_n(\boldsymbol{\beta}(p|\boldsymbol{\theta}))\boldsymbol{b}(p)^T dp, \tag{6}$$

which we call the *integrated loss minimization* (ILM).

The integrated objective function (5) is obtained by marginalizing the objective function (3) over the interval (0,1), hence it can be said to be an average loss function. Using this approach allows us to estimate the entire quantile process instead of estimating a discrete set of variables.

Now we define some quantities

$$B_h(p) = \int_0^p b_h(u)du, \qquad \overline{B}_h = \int_0^1 B_h(u)du \qquad \text{and}$$

$$b_h'(p) = \frac{db_h(p)}{dp}, h = 1, ..., k.$$

We let $\boldsymbol{B}(p), \overline{\boldsymbol{B}}$ and $\boldsymbol{b}'(p)$ denote the corresponding k-dimensional vectors. Now we can rewrite the integrated loss function (5) and the integrated gradient function (6) can be written as

$$\overline{L}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} y_i(p_i - 0.5) + \boldsymbol{x}_i^T \boldsymbol{\theta}[\overline{\boldsymbol{B}} - \boldsymbol{B}(p_i)], \tag{7}$$

and

$$\overline{S}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \boldsymbol{x_i}[\overline{\boldsymbol{B}} - \boldsymbol{B}(p_i)]^T. \tag{8}$$

Here, $p_i = F(y_i|\boldsymbol{x}_i, \boldsymbol{\theta})$ corresponds to the CDF of y_i evaluated at θ, which is the same as saying that it corresponds to the order of the quantile such that $\boldsymbol{x}_i^T \boldsymbol{\beta}(p_i|\boldsymbol{\theta}) = y_i$, where p_i is a function of $\boldsymbol{\theta}$. The sample mean of $\boldsymbol{B}(p_i)$ is approaching its integral $\overline{\boldsymbol{B}}$ when $\overline{S}_n(\boldsymbol{\theta})$ approaches 0, which it does when $p_1, ..., p_n$ are evenly spaced within the unit interval. The integrated gradient defined in equation (8) is a $q \times k$ matrix and if p_i is a smooth function then $\overline{S}_n$ will be a smooth function of $\boldsymbol{\theta}$ as well.

3.1.3 Asymptotic properties of the estimator

In this section we are going to look further into the asymptotic behaviour of the proposed estimator. To do this we will apply the standard theory of extremum estimators. We denote the parameter space Θ as the set of possible parameter values and $\boldsymbol{\theta}_0 \in \Theta$ is called the population parameter that satisfies the condition $F(\boldsymbol{x}_i^T \boldsymbol{\beta}(p|\boldsymbol{\theta}_0)) = p$. We denote $\hat{\boldsymbol{\theta}}$ as the minimizer of the integrated loss function (6) which we base on a sample size n. We let p_i^0, which has a standard uniform distribution by definition, be the value of p_i evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta_0}$ which then corresponds to the true CDF value of the data. It follows from these conditions that for any function g, $E[g(p_i^0)] = \int_0^1 g(p)dp$ and as $E[\boldsymbol{B}(p_i^0)] = \overline{\boldsymbol{B}}$ we have unbiasedness. The conditions of consistency and asymptotic normality usually follow as well under straightforward mild conditions.

4 Data analysis

We utilized QRCM to analyze BMI as a function of age, race, gender and height.

4.1 Description of data

The data were collected by the National Health and Nutrition Examination Survey (NHANES)[7], which is an American program of studies whose mission is to examine the status of adults and children in the United States in terms of the health and nutritional perspectives. To collect the desired information they perform surveys, physical experiments and interviews.

The data spans between the years 2015-2016. During this period there were 15327 people selected to be part of NHANES, 9971 completed the interviews and of these, 9544 were examined.

4.2 Variables

The original data from the NHANES are divided into two sets. One set is data from the interview survey [8], which contains 47 variables. The other set is data from the examination survey [9], which contains 26 variables. We used the variables gender, age and race from the interview survey and the variables height and BMI from the examination survey. Since there was a different amount of people who were examined and who took part of the interview, the number of observations in the corresponding data sets differs as well.

4.2.1 Gender

The first variable in the data set is gender, a binary variable where 1 indicates males and 2 for females. The number of male participants was 4892 and the remaining 5079 participants were females.

4.2.2 Age

The variable age represents the age in years of the participants when the survey was made. The age ranges from 0 to 80 years, where the value 80 represents the participants of 80 years or more. The number of participants who were between the ages 0 and 79 was 9595 and the number of participants who were aged 80 years or more was 376.

4.2.3 Race

Race is divided into the following 5 categories, with the number of participants representing this race in brackets: Mexican American (1921), other Hispanic (1308), non-Hispanic white (3066), non-Hispanic black (2129) and other race including multi-racial (1547). All these races add up to the total number of participants; there are no missing data on the variable race.

4.2.4 Height

The variable for height is measured in centimetres and the height span is from 80.7 to 202.7 cm. Height was measured on 8769 participants with 775 missing variables. Height was measured in the participants from the age of 2 years.

4.2.5 Body mass index

The last variable in the data set is BMI, which stands for Body Mass Index. It is calculated as weight in kilograms divided by height squared in meters with values rounded to one decimal digit. The span of BMI in the data set is from 11.5 to 67.3 kg/m^2. Since the height was not obtained for all participants, there are missing values for BMI as well. Out of 9544 who were examined, there are results from 8756 participants and the remaining 788 are missing values.

4.3 Missing data

The variables have a different number of valid observations. The number of sampled individuals differ between the variables, since some participants who did the survey did not undergo the examination. There are 9971 observations for the variables gender, age and race, and 9544 observations for the variables height and BMI. We included only the observations from the participants that were younger than age 80 years. This is because the participants with an age of 80 years or more received the value 80 years. When performing the data analysis we want values of all variables for the participants. The participants with completed data were removed from the data set. This will result in a data set containing 8419 observations.

4.4 Analysis of the data

To perform the data analysis we used the R program with the package named *qrcm*. The package was developed by Frumento in 2017 [14]. The integrated loss minimization (ILM), described in Section 3.1.2, is implemented in the qrcm package and it was used in the data analysis. The R code used in the data analysis is reported in Appendix A3. To decide the number of splines to use we looked at the significance of the splines. If they were not significant we decreased their number by one, and if all splines were significant we added one. However, no more than 4 splines were used. We divided men and women and analyzed them separately. For each gender we also studied the different races separately. We show the results of the analyses in graphs. The regression model was:

$$\begin{aligned} Q(p) = \beta_0 + \beta_1 \text{HeightSpline1} + \beta_2 \text{HeightSpline2} + \beta_3 \text{HeightSpline3} \\ + \beta_4 \text{AgeSpline1} + \beta_5 \text{AgeSpline2} + \beta_6 \text{AgeSpline3}. \end{aligned} \tag{9}$$

The β coefficients are estimated as

$$\beta_i(p) = \theta_{i0} + \theta_{i1} \log(p) + \theta_{i2} \log(1-p) + \theta_{i3} slp1 + \theta_{i4} slp2,$$

where the variables *slp* indicate Legendre polynomial bases. The β coefficients for the regression model were the following

$$\begin{aligned}
\beta_0(p) &= \theta_{00} + \theta_{01} \log(p) + \theta_{02} \log(1-p) + \theta_{03} slp1 + \theta_{04} slp2 \\
\beta_1(p) &= \theta_{10} + \theta_{11} \log(p) + \theta_{12} \log(1-p) + \theta_{13} slp1 + \theta_{14} slp2 \\
\beta_2(p) &= \theta_{20} + \theta_{21} \log(p) + \theta_{22} \log(1-p) + \theta_{23} slp1 + \theta_{24} slp2 \\
\beta_3(p) &= \theta_{30} + \theta_{31} \log(p) + \theta_{32} \log(1-p) + \theta_{33} slp1 + \theta_{34} slp2 \\
\beta_4(p) &= \theta_{40} + \theta_{41} \log(p) + \theta_{42} \log(1-p) + \theta_{43} slp1 + \theta_{44} slp2 \\
\beta_5(p) &= \theta_{50} + \theta_{51} \log(p) + \theta_{52} \log(1-p) + \theta_{53} slp1 + \theta_{54} slp2 \\
\beta_6(p) &= \theta_{60} + \theta_{61} \log(p) + \theta_{62} \log(1-p) + \theta_{63} slp1 + \theta_{64} slp2.
\end{aligned} \tag{10}$$

4.5 Results

We obtained estimates for the β coefficients. In Table 1 the estimates for females are reported with the standard error within the parenthesis. The

p-values represent significance of the null hypothesis that the value of θ in each column, is 0. All β coefficients are significantly different from zero. For example, using equation (10) we can read that the estimated quantile regression coefficient for spline 3 of height is

$$\begin{aligned}\hat{\beta}_3(p) = &-1.7 \cdot 10^{-6} - 1.8 \cdot 10^{-7} \log(p) + 5.3 \cdot 10^{-7} \log(1-p) \\ &+ 5.2 \cdot 10^{-7} slp1 - 4.9 \cdot 10^{-8} slp2.\end{aligned}$$

Table 1: Summary of estimates of β coefficients for Females. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Inter-cept	82 (51)	3.1 (16)	-15 (37)	-4.9 (48)	1.8 (26.9)	$<2{\cdot}10^{-16}$
Height Spline 1	-0.8 (0.6)	-0.1 (0.2)	0.2 (0.5)	1.2 (0.6)	$-9.8{\cdot}10^{-4}$ (0.3)	$<2{\cdot}10^{-16}$
Height Spline 2	$2.2{\cdot}10^{-3}$ $(1.6{\cdot}10^{-3})$	$2{\cdot}10^{-4}$ $(5{\cdot}10^{-4})$	$-6.4{\cdot}10^{-4}$ $(1.3{\cdot}10^{-3})$	$-5.0{\cdot}10^{-5}$ $(1.6{\cdot}10^{-3})$	$4.4{\cdot}10^{-5}$ $(8.8{\cdot}10^{-4})$	$<2{\cdot}10^{-16}$
Height Spline 3	$-1.7{\cdot}10^{-6}$ $(1.3{\cdot}10^{-6})$	$-1.8{\cdot}10^{-7}$ $(3.8{\cdot}10^{-7})$	$5.3{\cdot}10^{-7}$ $(9.6{\cdot}10^{-7})$	$5.2{\cdot}10^{-7}$ $(1.3{\cdot}10^{-6})$	$-4.9{\cdot}10^{-8}$ $(6.8{\cdot}10^{-7})$	$<2{\cdot}10^{-16}$
Age Spline 1	0.2 (0.2)	$1.2{\cdot}10^{-2}$ $(4.6{\cdot}10^{-2})$	$-4.2{\cdot}10^{-2}$ (0.1)	0.1 (0.2)	$1.6{\cdot}10^{-2}$ $(8.8{\cdot}10^{-2}$	$<2{\cdot}10^{-16}$
Age Spline 2	$-1.5{\cdot}10^{-3}$ $(1.5{\cdot}10^{-3})$	$-2.2{\cdot}10^{-4}$ $(4.1{\cdot}10^{-4})$	$-5.5{\cdot}10^{-5}$ $(1.1{\cdot}10^{-3})$	$-7.5{\cdot}10^{-4}$ $(3.4{\cdot}10^{-6})$	$-4.6{\cdot}10^{-4}$ $(1.9{\cdot}10^{-6})$	$<2{\cdot}10^{-16}$
Age Spline 3	$4.1{\cdot}10^{-6}$ $(3.6{\cdot}10^{-6})$	$8.4{\cdot}10^{-7}$ $(10{\cdot}10^{-7})$	$6.8{\cdot}10^{-7}$ $(2.5{\cdot}10^{-6})$	$7.8{\cdot}10^{-7}$ (0.3)	$1.5{\cdot}10^{-6}$ (0.2)	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	$5.1{\cdot}10^{-5}$	$<2{\cdot}10^{-16}$	$2.9{\cdot}10^{-10}$	$2.9{\cdot}10^{-10}$	

From the estimates in Table 2 we obtain that the estimated quantile regression coefficient for spline 2 of age is

$$\begin{aligned}\hat{\beta}_5(p) = &-1.5 \cdot 10^{-3} - 2.2 \cdot 10^{-4} \log(p) - 5.5 \cdot 10^{-5} \log(1-p) \\ &- 7.5 \cdot 10^{-4} slp1 - 4.6 \cdot 10^{-4} slp2.\end{aligned}$$

The remaining summaries of ILM estimates are shown in Appendix A1.

Table 2: Summary of estimates of β coefficients for Males. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Intercept	86 (38)	10 (10)	12 (26)	32 (32)	16 (20)	$<2{\cdot}10^{-16}$
Height Spline 1	-0.8 (0.4)	-0.1 (0.1)	-0.2 (0.3)	-0.4 (0.4)	-0.2 (0.2)	$<2{\cdot}10^{-16}$
Height Spline 2	$1.8{\cdot}10^{-3}$ ($1.1{\cdot}10^{-3}$)	$2.8{\cdot}10^{-4}$ ($3.1{\cdot}10^{-4}$)	$4.0{\cdot}10^{-4}$ ($8.0{\cdot}10^{-4}$)	$1.2{\cdot}10^{-3}$ ($9.8{\cdot}10^{-4}$)	$6.6{\cdot}10^{-4}$ ($6.0{\cdot}10^{-4}$)	$<2{\cdot}10^{-16}$
Height Spline 3	$-1.2{\cdot}10^{-6}$ ($8.4{\cdot}10^{-7}$)	$-1.8{\cdot}10^{-7}$ ($2.3{\cdot}10^{-7}$)	$-3.2{\cdot}10^{-7}$ ($6.0{\cdot}10^{-7}$)	$-9.1{\cdot}10^{-7}$ ($7.4{\cdot}10^{-7}$)	$-4.9{\cdot}10^{-7}$ ($4.4{\cdot}10^{-7}$)	$<2{\cdot}10^{-16}$
Age Spline 1	-0.1 (0.2)	$-7.9{\cdot}10^{-2}$ ($5.4{\cdot}10^{-2}$)	-0.2 (0.1)	-0.2 (0.2)	-0.3 ($9.6{\cdot}10^{-2}$)	$<2{\cdot}10^{-16}$
Age Spline 2	$1.6{\cdot}10^{-3}$ ($1.7{\cdot}10^{-3}$)	$5.6{\cdot}10^{-4}$ ($4.8{\cdot}10^{-4}$)	$1.2{\cdot}10^{-3}$ ($9.5{\cdot}10^{-4}$)	$1.2{\cdot}10^{-3}$ ($1.3{\cdot}10^{-3}$)	$1.8{\cdot}10^{-3}$ ($8.2{\cdot}10^{-4}$)	$<2{\cdot}10^{-16}$
Age Spline 3	$-3.7{\cdot}10^{-6}$ ($4.1{\cdot}10^{-6}$)	$-1.0{\cdot}10^{-6}$ ($1.6{\cdot}10^{-6}$)	$-2.0{\cdot}10^{-6}$ ($2.2{\cdot}10^{-6}$)	$-2.3{\cdot}10^{-6}$ ($3.0{\cdot}10^{-6}$)	$-3.3{\cdot}10^{-6}$ ($1.9{\cdot}10^{-6}$)	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	$5.1{\cdot}10^{-10}$	$<2{\cdot}10^{-16}$	$2.9{\cdot}10^{-10}$	$2.9{\cdot}10^{-10}$	

For the QRCM estimates we considered the 25th, 50th and 75th quantiles. We performed this for each gender and race separately in the data analysis. The results in Table 3 and Table 4 are from the estimates of all males and all females, leaving the remaining estimates to Appendix A2.

The estimates in Table 3 and Table 4 represent the values for the male and female populations respectively. In the regression model in equation (9) the estimates for the β coefficients are the values in the tables. For example, the estimates for the 25th quantile are

$$
\begin{aligned}
Q(0.25) = 66.1 - 0.5 HeightSpline1 + 1.2 \cdot 10^{-3} HeightSpline2 \\
-7.4 \cdot 10^{-7} HeightSpline3 + 0.2 AgeSpline1 \\
-0.9 \cdot 10^{-3} AgeSpline2 + 8.3 \cdot 10^{-7} AgeSpline3.
\end{aligned}
$$

In Figure 2 and Figure 3 we have plotted the estimates for the intercepts and the splines. From the figures we can observed that the first splines for both age and height are significant, whereas the coefficients for the second and third splines in both cases have very small values. The reason for this can be due to the fact that we have a large data set. A statistically insignificant spline can however be clinically significant. The estimated values are difficult to interpret.

Table 3: QRCM estimates of the 25th, 50th and 75th quantiles for Males

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	66.1	6.3	53.7	78.5
	0.50	78.4	6.8	65.1	91.7
	0.75	96.4	11.6	73.7	119.2
Height Spline 1	0.25	-0.5	0.1	-0.7	-0.4
	0.50	-0.7	0.1	-0.8	-0.5
	0.75	-0.9	0.1	-1.2	-0.7
Height Spline 2	0.25	$1.2{\cdot}10^{-3}$	$0.2{\cdot}10^{-3}$	$0.8{\cdot}10^{-3}$	$1.6{\cdot}10^{-3}$
	0.50	$1.6{\cdot}10^{-3}$	$0.2{\cdot}10^{-3}$	$1.1{\cdot}10^{-3}$	$2.0{\cdot}10^{-3}$
	0.75	$2.2{\cdot}10^{-3}$	$0.4{\cdot}10^{-3}$	$1.5\ {\cdot}10^{-3}$	$2.9{\cdot}10^{-3}$
Height Spline 3	0.25	$-7.4{\cdot}10^{-7}$	$1.4{\cdot}10^{-7}$	$-1.0{\cdot}10^{-7}$	$-4.6{\cdot}10^{-7}$
	0.50	$-1.0{\cdot}10^{-6}$	$1.6{\cdot}10^{-7}$	$-1.3{\cdot}10^{-6}$	$-6.9{\cdot}10^{-7}$
	0.75	$-1.5{\cdot}10^{-6}$	$2.7{\cdot}10^{-7}$	$-2.0{\cdot}10^{-7}$	$-9.8{\cdot}10^{-7}$
Age Spline 1	0.25	0.2	$3.7\ {\cdot}10^{-2}$	0.2	0.3
	0.50	0.3	$4.1{\cdot}10^{-2}$	0.2	0.3
	0.75	0.2	0.1	0.1	0.3
Age Spline 2	0.25	$-0.9{\cdot}10^{-3}$	$0.3{\cdot}10^{-3}$	$-.1.6{\cdot}10^{-3}$	$-0.3{\cdot}10^{-3}$
	0.50	$-1.1{\cdot}10^{-3}$	$0.4{\cdot}10^{-3}$	$-1.8{\cdot}10^{-3}$	$-0.4{\cdot}10^{-3}$
	0.75	$-0.4{\cdot}10^{-3}$	$0.5{\cdot}10^{-3}$	$-1.4{\cdot}10^{-3}$	$0.5{\cdot}10^{-3}$
Age Spline 3	0.25	$8.3{\cdot}10^{-7}$	$7.76{\cdot}10^{-7}$	$-6.9{\cdot}10^{-7}$	$2.4{\cdot}10^{-6}$
	0.50	$1.0{\cdot}10^{-6}$	$8.1{\cdot}10^{-7}$	$-5.8{\cdot}10^{-7}$	$2.6{\cdot}10^{-6}$
	0.75	$-4.3{\cdot}10^{-7}$	$1.1{\cdot}10^{-6}$	$-2.6{\cdot}10^{-6}$	$1.8{\cdot}10^{-6}$

Figure 2: Plotted estimates of splines and intercept for Males.

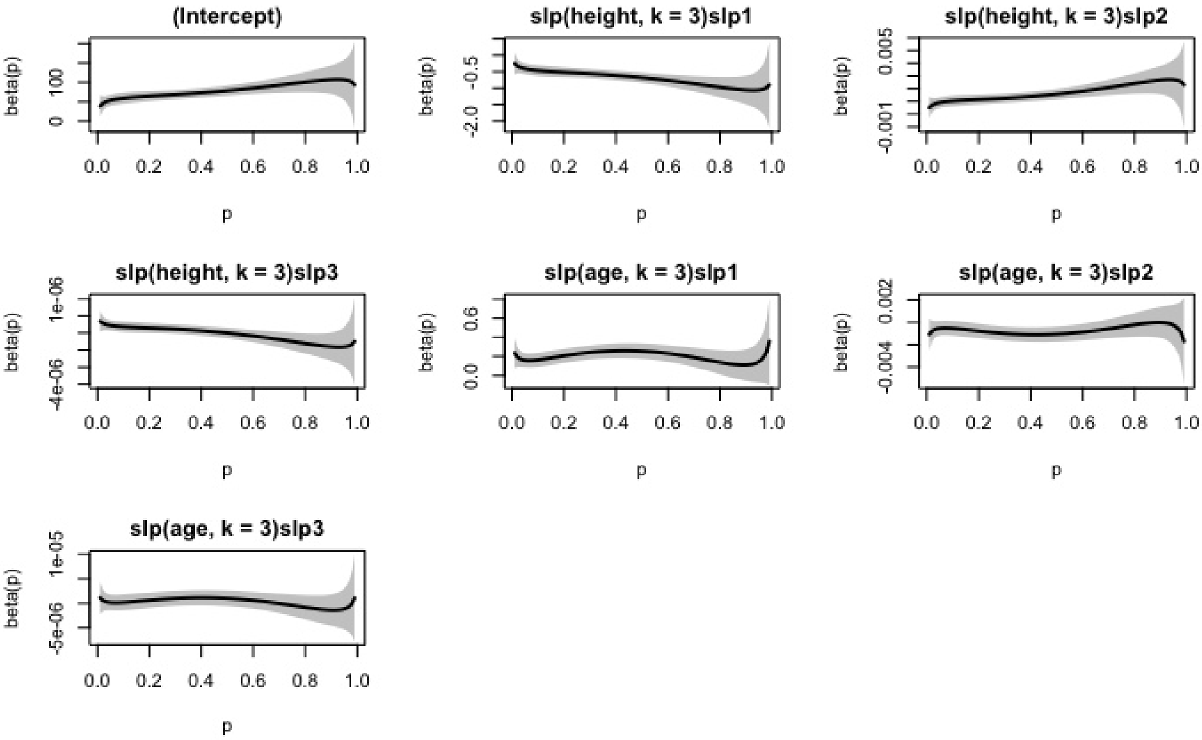
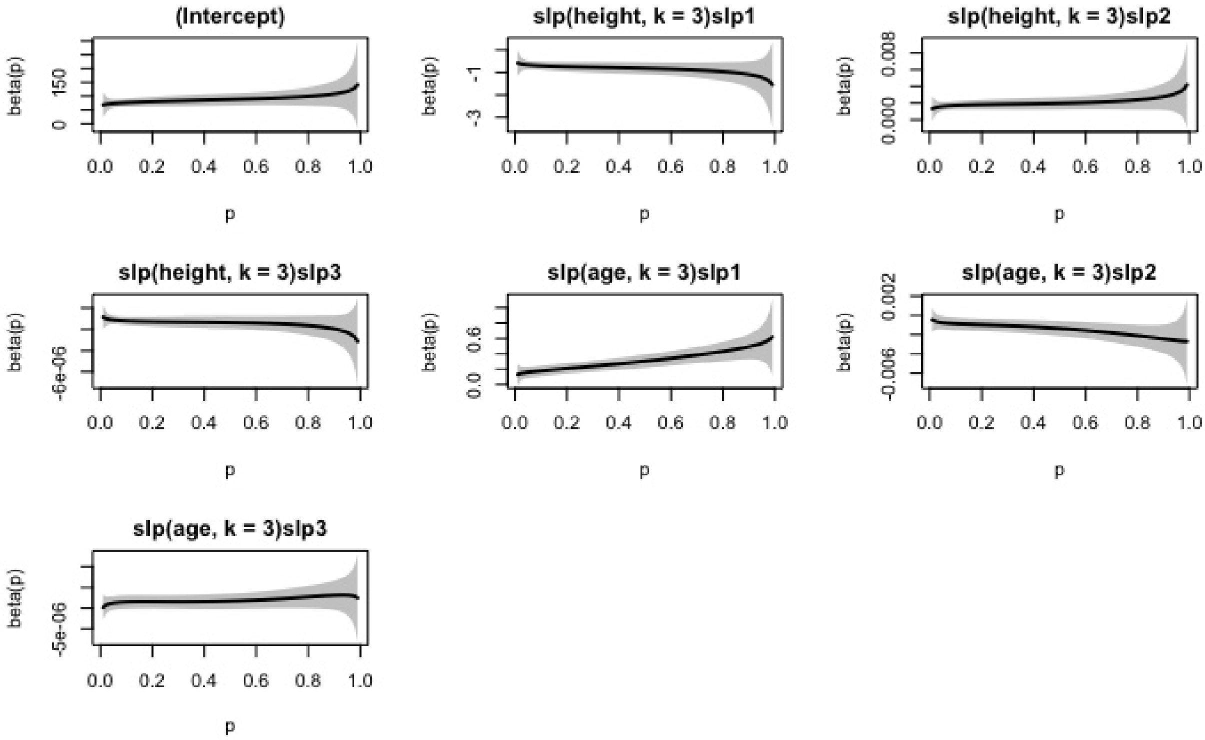


Table 4: QRCM estimates of the 25th, 50th and 75th quantiles for Females

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	81.4	9.6	62.6	100.2
	0.50	87.8	11.1	66.0	109.6
	0.75	96.3	15.2	66.5	126.1
Height Spline 1	0.25	-0.8	0.1	-1.0	-0.5
	0.50	-0.8	0.1	-1.1	-0.6
	0.75	-0.9	0.2	-1.3	-0.6
Height Spline 2	0.25	$1.8{\cdot}10^{-3}$	$0.3{\cdot}10^{-3}$	$1.2{\cdot}10^{-3}$	$2.5{\cdot}10^{-3}$
	0.50	$2.0{\cdot}10^{-3}$	$0.4{\cdot}10^{-3}$	$1.2{\cdot}10^{-3}$	$2.3{\cdot}10^{-3}$
	0.75	$2.3{\cdot}10^{-3}$	$0.5{\cdot}10^{-3}$	$1.2{\cdot}10^{-3}$	$3.3{\cdot}10^{-3}$
Height Spline 3	0.25	$-1.3{\cdot}10^{-7}$	$2.5{\cdot}10^{-7}$	$-1.8{\cdot}10^{-6}$	$-7.9{\cdot}10^{-7}$
	0.50	$-1.3{\cdot}10^{-6}$	$3.0{\cdot}10^{-7}$	$-1.9{\cdot}10^{-6}$	$-7.6{\cdot}10^{-7}$
	0.75	$-1.6{\cdot}10^{-6}$	$4.2{\cdot}10^{-7}$	$-2.4{\cdot}10^{-6}$	$-7.3{\cdot}10^{-7}$
Age Spline 1	0.25	0.2	$3.6{\cdot}10^{-2}$	0.2	0.3
	0.50	0.3	$4.4{\cdot}10^{-2}$	0.2	0.4
	0.75	0.4	0.1	0.3	0.5
Age Spline 2	0.25	$-1.0{\cdot}10^{-3}$	$0.3{\cdot}10^{-3}$	$-1.7{\cdot}10^{-3}$	$-0.4{\cdot}10^{-3}$
	0.50	$-1.4{\cdot}10^{-3}$	$0.4{\cdot}10^{-3}$	$-2.1{\cdot}10^{-3}$	$-0.6{\cdot}10^{-3}$
	0.75	$-2.0{\cdot}10^{-3}$	$0.5{\cdot}10^{-3}$	$-3.0\ {\cdot}10^{-3}$	$-0.9{\cdot}10^{-3}$
Age Spline 3	0.25	$1.5{\cdot}10^{-6}$	$8.1{\cdot}10^{-7}$	$-6.3{\cdot}10^{-8}$	$3.1{\cdot}10^{-6}$
	0.50	$1.7{\cdot}10^{-6}$	$9.4{\cdot}10^{-7}$	$-1.7{\cdot}10^{-7}$	$3.5{\cdot}10^{-6}$
	0.75	$2.5{\cdot}10^{-6}$	$1.3{\cdot}10^{-6}$	$2.0{\cdot}10^{-8}$	$4.9{\cdot}10^{-6}$

Figure 3: Plotted estimates of splines and intercept for Females



4.6 Comparison with growth charts and reference values

Growth of height and BMI in children is often evaluated using reference values and growth charts. Parents can then easy measure the height of their child and compare the result with the reference value of a child of the same gender and age. The 2000 CDC growth charts for the United States [22] are recommended when estimating the growth of children from newborns up to age 20 in the United States. These growth charts are a revised version of the National Center for Health Statistics (NCHS) growth charts that has been used since 1977. The growth charts are divided by gender, into two sets. The first growth charts are for infants who are newborn up to the age of 36 months. The second set of growth charts, the growth charts of our interest, are for children from the age of 2 to 20. There are also different types of growth charts. For the age span between 2-20 years, there are growth charts for weight-for-age, stature-for-age and BMI-for-age. The two types of growth charts that we are going to take a closer look at are growth charts for stature-for-age and BMI-for-age. We will compare the CDC growth charts with the data used in this thesis and compare the graphs.

Reference values differ from country to country and to obtain a reliable reference value it is necessary to compare reference values for your country.

The process of calculating these reference values were divided into two parts [22]. The first part is called the curve smoothing stage. Sample weights were applied in order to calculate the weighted empirical percentile points. The weights were calculated from the midpoint of the age groups separately. Then the estimates of the weighted empirical percentiles were obtained for the percentiles 3rd, 5th, 10th, 25th, 50th, 75th, 90th, 95th and 97th. The curves were then smoothed, resulting in charts with a parametric form. Each percentile had a corresponding estimated parameter. The second stage was the transformation stage, where the smoothed curves obtained from the first stage were approximated. The result of this stage are percentile curves which are very alike the smoothed curves. A modified LMS estimation was used to approximate the curves and the method produced the transformation parameters which are lambda, mu and sigma, hence the shortening LMS. From these percentile curves it is possible to obtain additional percentiles and z-scores. Z-scores were calculated using the formulas below [26].

$$Z = \frac{(\frac{BMI}{M})^L - 1}{LS} \qquad \text{for} \qquad L \neq 0,$$

$$Z = \frac{1}{S} \cdot \log_e(\frac{BMI}{M}) \qquad \text{for} \qquad L = 0.$$

The z-scores are then used for the LMS transformation equation which is stated below [22],

$$X = M(1 + LSZ)^{1/L} \qquad \text{for} \qquad L \neq 0,$$

$$X = M \exp(SZ) \qquad \text{for} \qquad L = 0.$$

The letter M represents the median, S is for the generalized coefficient of variation, X is the physical measurement and L is the Box-Cox transformation which corresponds to the degree of skewness. The z-score obtained corresponds to a percentile which can be obtained from a normal distribution table.

To analyse results of BMI it is necessary to know what the values stands for. The BMI ranges for adults with an age over 20 are shown in Table 5 [31].

Table 5: The World Health Organisation BMI Ranges of Nutritional Status

BMI	Nutritional Status
Below 18.5	Underweight
18.5-24.9	Normal weight
25.0-29.9	Pre-obesity
30.0-34.9	Obesity class I
35.0-39.9	Obesity class II
Above 40	Obesity class III

In the growth chart in Figure 4, we observe the BMI-for-age percentiles for girls between the ages 2 to 20. The growth chart has been applied to 9 empirical percentiles. From the lowest 5th percentile to the 95th percentile. It is important that the 85th percentile is included in the BMI growth chart because this percentile is a cutoff recommend to use to single out children who are at the risk of obesity according to Barlow and Dietz [2]. Studying the graph in Figure 4 we can approximate the BMI value of the 50th percentile for a 5-year-old to be 15.2, for an 11-year-old to 17.5 and for a 20-year-old to 21.9. The children who are above the 85th percentile, that is for example an 8-year-old with a BMI of 18.7 or a 14-year-old with a BMI of 23.3, and have complications of obesity are considered to receive treatment to prevent obesity. The children who have values that are above the 95th percentile will receive treatment even though they do not have complications with obesity, according to Barlow and Dietz [2].

In Figure 5 we observe the 10th, 25th, 50th , 75th, 85th and 95th percentiles of BMI-age of Girls aged 2-20 years old from the data used in this thesis. From Figure 5 we can roughly read that the BMI value for a 5-year-old from the 50th percentile is 17, for an 11-year-old it is about 19 and for a 20-year-old the value is approximately 23. If we study the values of the 85th percentile 8-year-olds we see that it corresponds to a BMI of 22, for 14-year-olds it corresponds to a BMI of 25.

Figure 4: CDC Growth Charts of BMI-Age of Girls 2-20 years

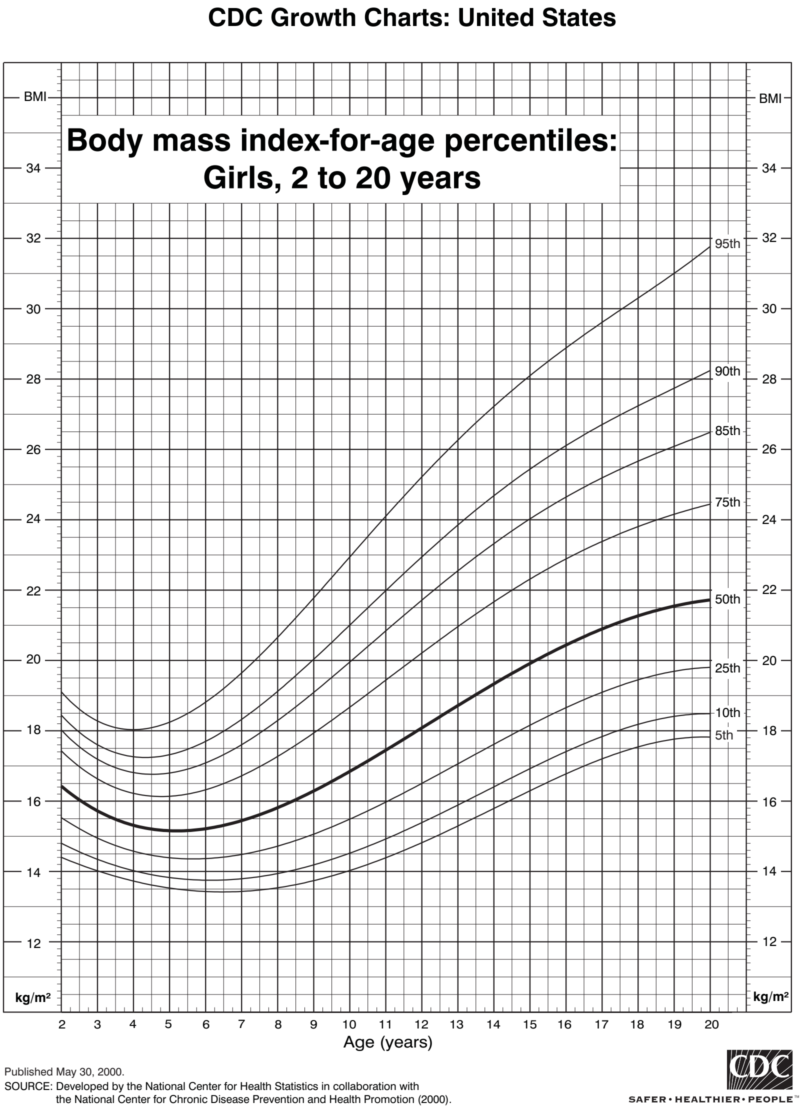
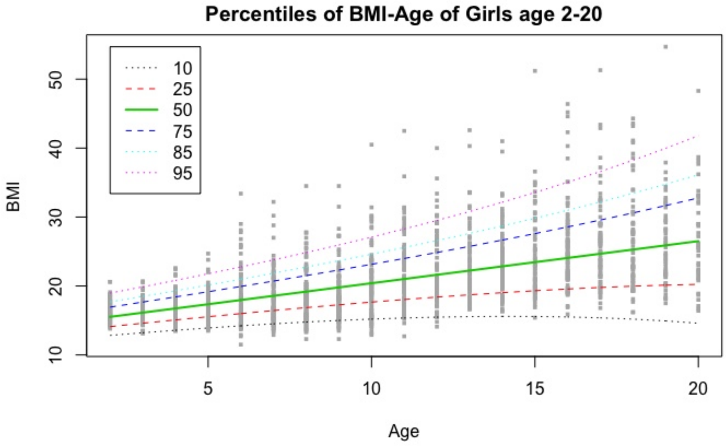


Figure 5: Percentiles of BMI-Age of Girls 2-20 years



In the growth chart in Figure 6 we observe the BMI-for-age percentiles for boys between the ages 2 to 20. The percentiles in this growth chart are the same as for the girls, from the lowest 5th percentile to the 95th percentile. From the graph we find that the BMI value of the 50th percentile for a 5-year-old is about 13.7, for a 11-year-old about 14.5 and for a 20-year-old the value is approximately 19. All these values are lower than the same for the girls. The boys who are above the 85th percentile, for example a 13-year-old with a BMI of 22 or a 9-year-old with a BMI of 18.5, and have complications of obesity are considered to receive treatment to prevent obesity.

In Figure 7 we observe 6 percentiles of BMI-age of boys aged 2-20 years old from the data used in this thesis. From Figure 7 we can roughly read that the BMI value for a 5-year-old in the 50th percentile is 17, for an 11-year-old it is about 18 and for a 20-year-old the value is approximately 23. If we study the values of the 85th percentile for 8-year-olds we see that it corresponds to a BMI of 23, 15-year-olds it corresponds to a BMI of 27.

Figure 6: CDC Growth Charts of BMI-age of Boys 2-20 years

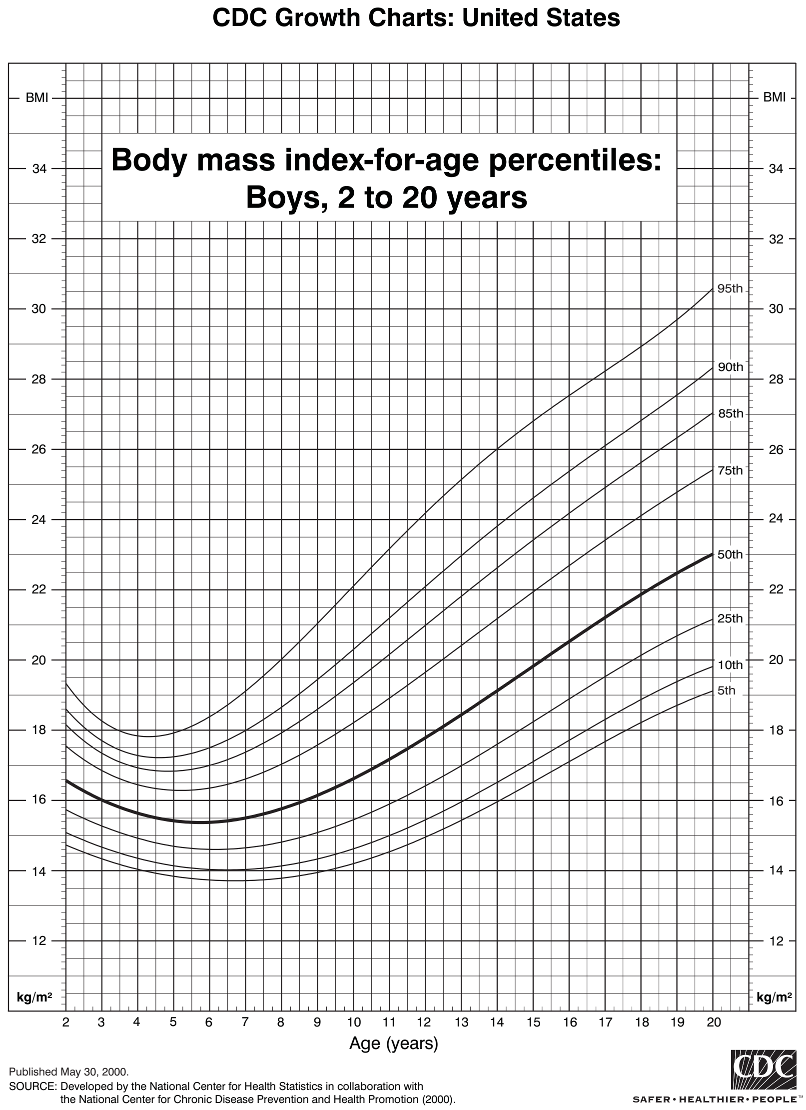
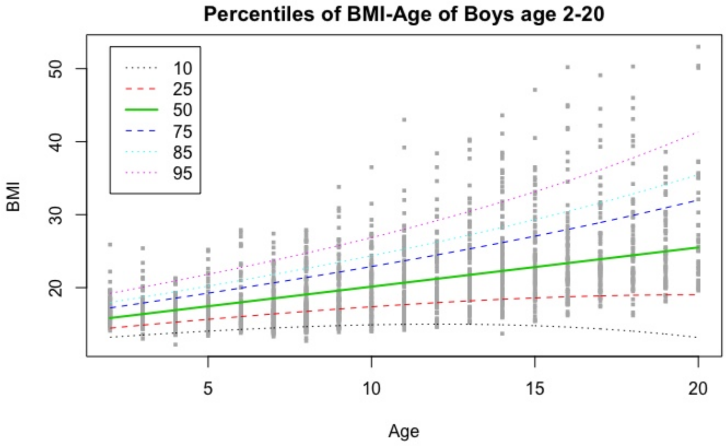


Figure 7: Percentiles of BMI-Age of Boys 2-20 years



In the growth chart in Figure 8 we observe the Stature-for-age percentiles for girls between the ages 2 to 20. In this growth chart 7 percentiles are included. From the graph we find that the height of the 50th percentile for a 5-year-old is 108 cm, for an 11-year-old it is about 145 cm and for a 20-year-old the value is approximately 163 cm.

In Figure 9 we observe percentiles of Height-age of girls aged 2-20 years old from the data used in this thesis. In this graph we observe that the height of the 50th percentile for a 5-year-old is 110 cm, for an 11-year-old it is 137 cm and for a 20-year-old it is 178 cm. In this data set we can conclude, by comparing Figure 8 and Figure 9 that the girls are taller in the 50th percentile than the heights from the growth chart.

Figure 8: CDC Growth Charts of Height-Age of Girls 2-20 years

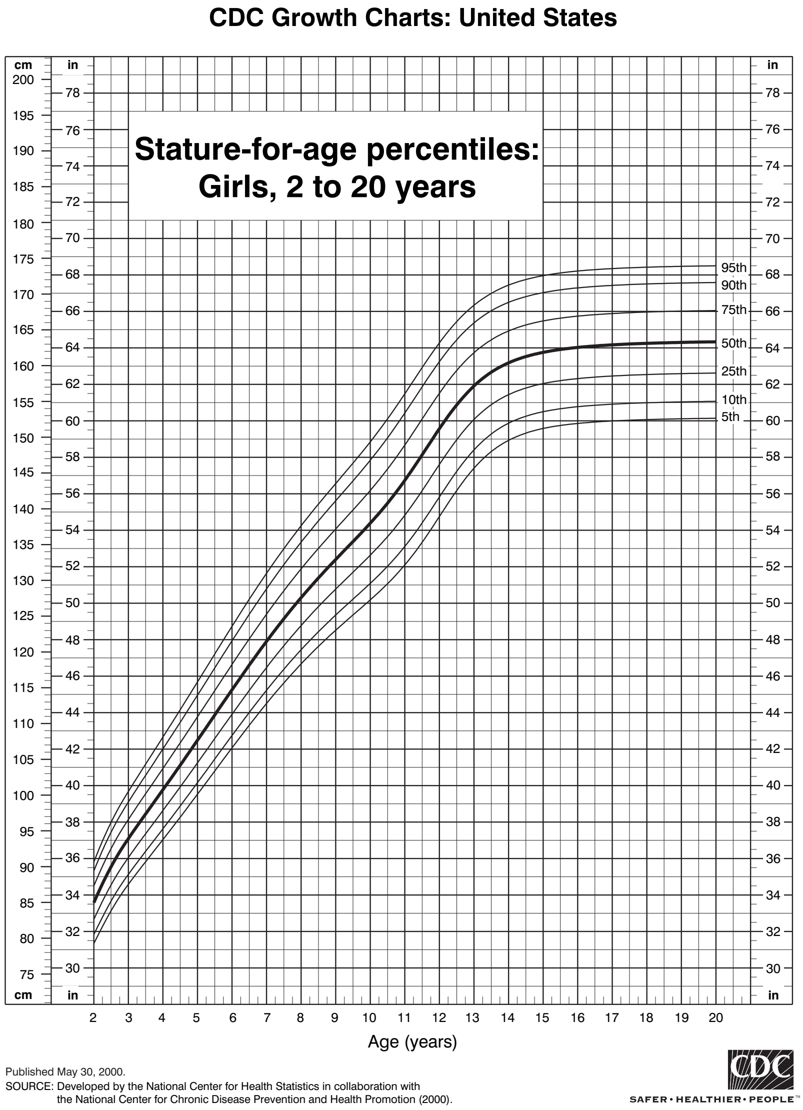
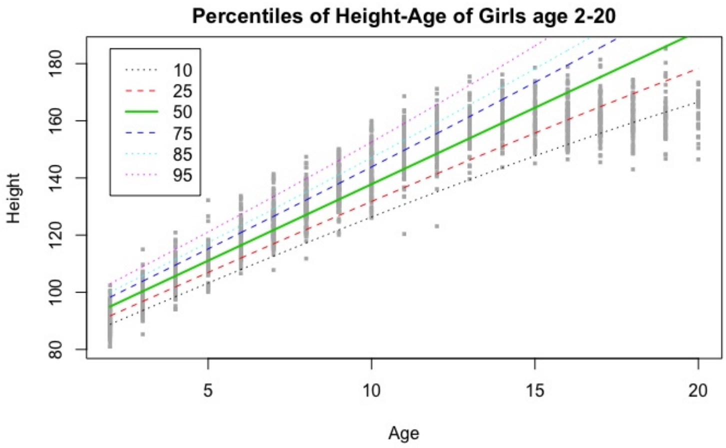


Figure 9: Percentiles of Height-Age of Girls 2-20 years



In the growth chart in Figure 10 we observe the Stature-for-age percentiles for boys between the ages 2 to 20. From Figure 10 we find that the height of the 50th percentile for a 5-year-old is 109 cm, for an 11-year-old it is about 144 cm and for a 20-year-old the value is approximately 173 cm. The heights for a 5 and an 11-year-old are approximately the same as the height for the girls, whereas for a 20-year-old the height of the 50th percentile is about 10 cm more.

In Figure 11 we observe percentiles of Height-age of boys aged 2-20 years old from the data used in this thesis. From Figure 11 we observed that the height of the 50th percentile for a 5-year-old is 115 cm, for an 11-year-old it is 140 cm and for a 20-year-old the it is 185 cm. By comparing Figure 10 and Figure 11 we see that the height of the boys at the 50th percentile in this data set are taller than the heights in the growth chart.

Figure 10: CDC Growth Charts of Height-Age of Boys 2-20 years

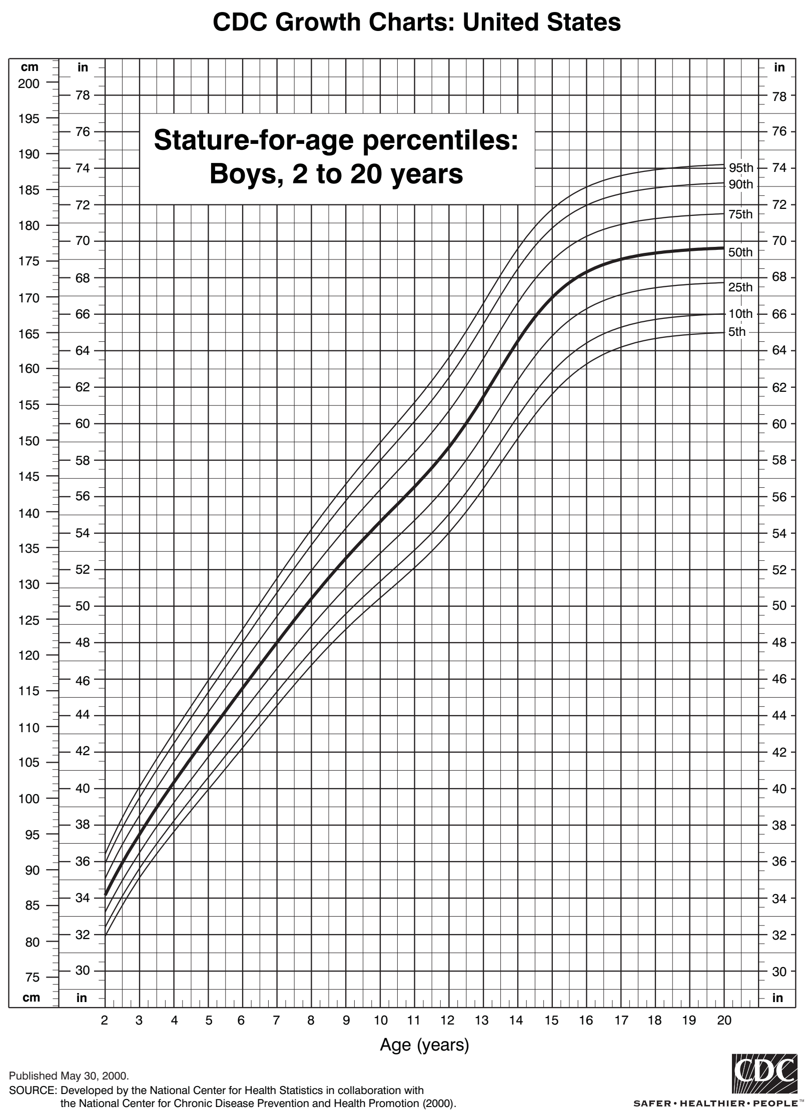
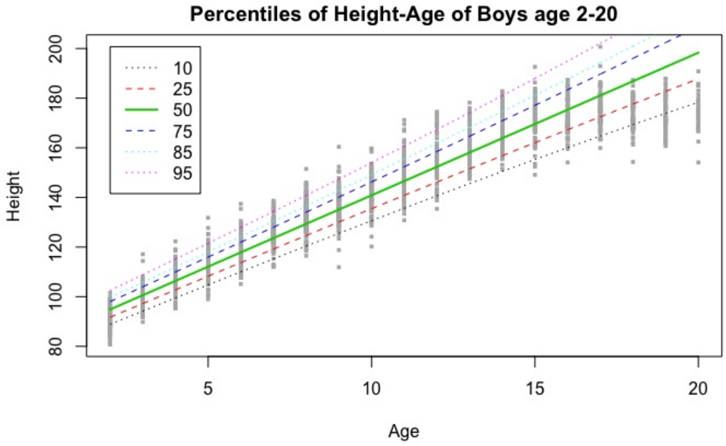


Figure 11: Percentiles of Height-Age of Boys 2-20 years



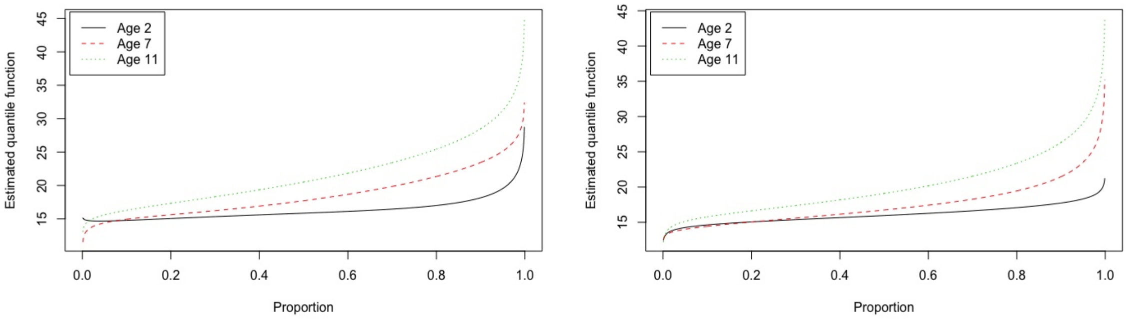
4.7 Interpretation

From the QRCM in R we obtained estimates of reference values of BMI given a specified age, race and gender for any given quantile. In this section we study articles written about BMI for different ages and ethnicities to compare their results with the reference values obtained by the QRCM. The quantiles that we chose to study are the 10th, 50th and the 95th quantiles.

In an article published in the American Public Health Association by Balistreri et al [1] in 2009, the authors studied the relationship between socioeconomic status and BMI among Hispanic children of immigrants and children of natives. They followed the children from kindergarten up to fifth grade, when the children are at an age of 10 years old. One of the results from the study was that for non-Hispanic white children the parental education had a negative effect of the child's BMI values. The parental education for Hispanic children had a weak negative effect on BMI growth. For Hispanic and white native children in kindergarten, the income had a strong negative relationship with the child's BMI value. As for children from a Hispanic immigrant family, the association between income and BMI was positive. Among the children in the data set that we used, we did not have information on the origins of the parents, education or income of the parents. We used our data to study the development of BMI for Hispanic and non-Hispanic white children for ages 2, 7 and 11. In 2013 the percentage of children aged 6-11 years with pre-obesity and obesity in the United States was 17.4% [10], for children aged 2-5 years the value was 9.4% . Worldwide the percentage of boys with obesity was 23.8% and for girls the percentage was 22.6% [25].

In the left-hand graph in Figure 12 we observe that the black straight line, representing age 2, is vertical for most of the quantiles. This is because most 2-year-olds have similar height and weights and therefore the BMI does not vary much. When we study the quantiles for ages 7 and 11 we see that the there is an upward slope of the curve, this means that the variation of BMI for those children is beginning to grow.

Figure 12: Quantiles of BMI of Hispanic (left) and non-Hispanic white children (right).



The estimates of BMI for non-Hispanic white children are shown in Table 6, for the 10th, 50th and 95th quantiles. If we study the values for a 2-year-old we notice that the increase in BMI from the 10th, with an estimated value of 14.77, to the 95th, with an estimated value of 19.55, quantiles is not very large. This small variation is due to the fact that on average, most 2-year-olds have similar heights and weights. For the ages 7 and 11 years there are larger differences between the 10th and 95th quantiles. For a 7-year-old Hispanic child the estimated BMI values of the 10th quantile is 14.95 and the corresponding values for the 95th quantile is 25.13, which is a difference of 10 BMI units. For an 11-year-old Hispanic child the estimated BMI values of the 10th quantile is 16.29 and the corresponding values for the 95th quantile is 31.23, which is a difference of 15 BMI units. The BMI ranges from Table 5 is not applied for children under the age of 20, but if a child has a BMI value of 30 or larger it is clear that the child is obese. Already at the age of 11 the 95th quantile of Hispanic children corresponds to being obese. The difference between the 50th quantiles of a 2-year-old and a 7-year-old is larger than the difference between the 50th quantiles for the 7-year-old and the 11-year-old.

Table 6: QRCM Estimates of Hispanic children with the standard errors in the parenthesis.

Hispanic Children			
Age and Height	10th quantile	50th quantile	95th quantile
2 Years, 93 cm	14.77(1.3)	15.85(0.23)	19.55(1.16)
7 Years, 123 cm	14.95(0.21)	17.71(0.30)	25.13(0.77)
11 years, 140 cm	16.29 (0.23)	20.50(0.28)	31.23(0.76)

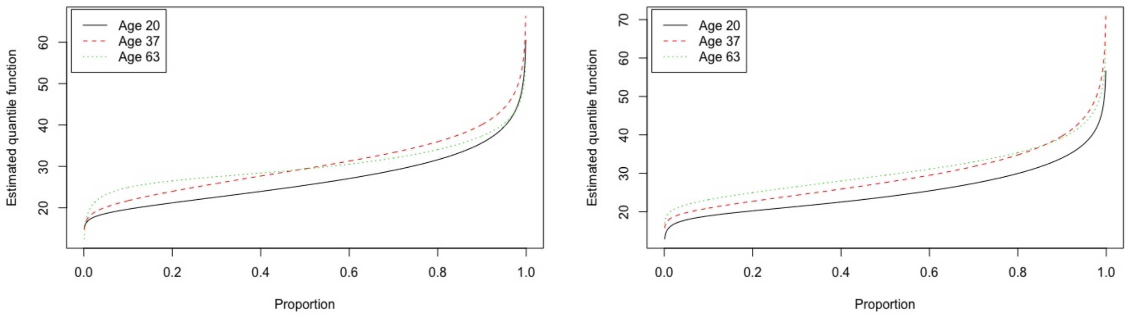
The estimates of BMI for non-Hispanic white children are shown in Table 7, for the 10th, 50th and 95th quantiles. The values for a 2-year-old with a length of 93 cm are very similar to Hispanic children, as the difference in estimated BMI for the 10th quantile,with an estimated value of 14.62, and the 95th quantile, with an estimated value of 18.32, is small. As for the estimated values of the Hispanic children, there is a larger difference between the 10th and 95th quantiles for age 7 with a length of 123 cm and for the age 11 with a length of 140 cm. For a 7-year-old non-Hispanic white child the estimated BMI value of the 10th quantile is 14.42 and the corresponding values for the 95th quantile is 23.51. For an 11-year-old Hispanic child the estimated BMI value of the 10th quantile is 15.77 and the corresponding values for the 95th quantile is 29.10. Overall, from these estimates the BMI values for non-Hispanic white children are smaller than the BMI values for the Hispanic children. The difference between the 50th quantiles for the 2-year-old and the 7-year-old is smaller for non-Hispanic white children than Hispanic children.

Table 7: QRCM Estimates of non-Hispanic white children with the standard errors in the parenthesis.

Non-Hispanic white Children			
Age and Height	10th quantile	50th quantile	95th quantile
2 Years, 93 cm	14.62(0.15)	15.95(0.14)	18.32(0.22)
7 Years, 123 cm	14.42(0.11)	16.74(0.15)	23.51(0.59)
11 years, 140 cm	15.77 (0.12)	19.10(0.18)	29.10(0.55)

In an article published in the American Journal of Public Health by Kirby et al. [18] in 2012, the authors addressed the question if there was a relationship between living in a specific racial community and the risk of obesity. According to the world health organisation (WHO) [31], adults with BMI values between 25 and 30 are classified as pre-obesity or overweight and those with BMI values over 30 are obese. The results from the study by Kirby et al. showed that the Hispanic inhabitants who lived in communities in which the Hispanic population exceeded 25%, were associated with an increase of 0.55 in BMI and the odds for obesity was 21%. The same numbers for non-Hispanic whites who lived in a community with a Hispanic population of over 25% was an 0.42 increase in BMI and a 23% odds of obesity. The researchers also studied non-Hispanic whites who lived in communities with a non-Hispanic Asian population of at least 25% and this was associated with a 0.68 decrease in BMI and 28% lower odds of obesity. The left-hand graph in Figure 13 represents Hispanic men and women for the ages 20, 37 and 63. The right-hand graph represents non-Hispanic white men and women for the ages 20, 37 and 63. Studying these figures we see clearly that the BMI values are higher for people with a Hispanic ethnicity.

Figure 13: Quantiles of BMI of Hispanic (left) and non-Hispanic white (right).



The estimates for Hispanics aged 20, 37 and 63 of the 10th, 50th and 95th quantiles are shown in Table 8. For a 165 cm, 20-year-old the estimated BMI values of the 50th quantile is 25.42 which indicates pre-obesity. The estimated BMI values for the 50th quantile of a 171 cm long 37-year-old is 29.63 and for a 166 cm long 63-year-old is 29.53. Both of these values indicate pre-obesity. As seen in Figure 13 there is a big difference between the 10th and the 95th quantiles for all ages. This is also verified from Table 8 where the 10th quantile has a BMI of 19.65 and the 95th quantile a BMI of 39.56 for a 20-year-old. The difference is even bigger for the 37 and 63-year-olds where the difference is about 20 BMI units for both ages.

Table 8: QRCM Estimates of Hispanic with the standard errors in the parenthesis.

Hispanic people			
Age and Height	10th quantile	50th quantile	95th quantile
20 Years, 165 cm	19.65(0.30)	25.42(0.41)	39.56(1.07)
37 Years, 171 cm	22.90(0.37)	29.63(0.39)	43.76(1.09)
63 years, 166 cm	23.94 (0.39)	29.53(0.39)	41.72(1.16)

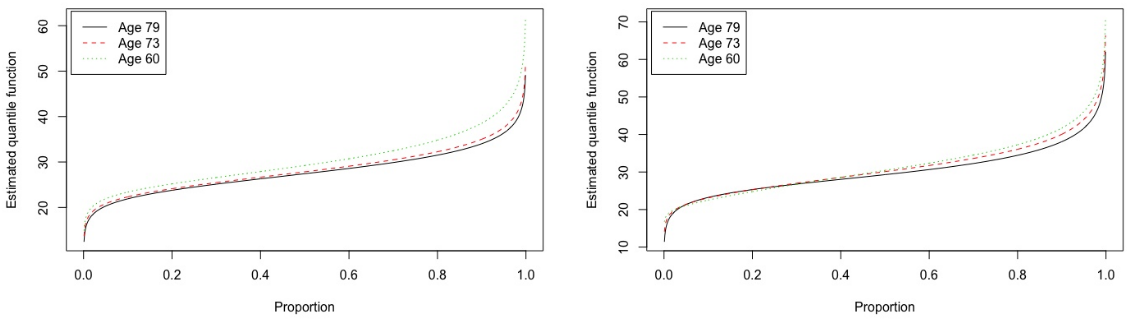
In Table 9 we have the estimates for non-Hispanics whites aged 20, 37 and 63 of the 10th, 50th and 95th quantiles. For a 168 cm long 20-year-old the estimates BMI values of the 50th quantile is 23.96 which indicates normal weight. The estimated BMI values for the 50th quantile of a 167 cm long 37-year-old is 28.10 and for a 166 cm long 63-year-old is 29.50. Both of these values indicate pre-obesity. As seen in Figure 13 there is a big difference between the 10th and the 95th quantiles for all ages. This is also verified from Table 9 where the 10th quantile has a BMI of 19.05 and the 95th quantile a BMI of 37.63 for a 20-year-old. The difference is even bigger for the 37 and 63-year-olds where the difference is over 20 BMI units for both ages. Compared with the estimates for Hispanics, a 20-year-old non-Hispanic white is normal weight and the BMI values overall are smaller. For a 37 and a 63-year-old the 50th quantiles both indicates pre-obesity, the same as for the Hispanics.

Table 9: QRCM Estimates of non-Hispanic white with the standard errors in the parenthesis.

Non-Hispanic white people			
Age and Height	10th quantile	50th quantile	95th quantile
20 Years, 168 cm	19.05(0.18)	23.96(23)	37.63(0.66)
37 Years, 167 cm	21.52(0.23)	28.10(0.29)	44.67(0.73)
63 years, 166 cm	23.00 (0.25)	29.50(0.33)	43.33(0.89)

The two graphs in Figure 14 illustrate quantiles, with BMI on the y-axis, for men and women separately who were born before, during and after World War II. The black line represents a person that was 79 years old in 2015 and hence was born before the war. The red line represents a person that was 73 years old in 2015, and therefore was born during the war and lastly the green line represents a person of age 60 who was born a decade after the war ended, in 1955.

Figure 14: Quantiles of BMI of Males (left) and Females (right) before, during and after World War II.



The values in Table 10 represents the estimates of the 10th, 50th, and 95th quantiles of Males who were born before, during and after World War II. The estimates for the 50th quantiles are 21.90 for a male born before the war, 22.33 for a male born during the war and 23.39 for a male born a decade after the war. These estimated BMI values represents pre-obesity for all ages.

Table 10: QRCM Estimates of Male before/during/after WWII with the standard errors in the parenthesis.

Males before/during/after WWII			
Age and Height	10th quantile	50th quantile	95th quantile
79 Years, 175 cm	21.90(0.61)	27.42(0.44)	36.38(0.99)
73 Years, 171 cm	22.33(0.33)	27.84(0.25)	37.60(0.59)
60 years, 179 cm	23.39(0.20)	29.24(0.22)	42.13(0.65)

In Table 11 we find the estimated BMI values of females born before, during and after World War II for the 10th, 50th and 95th quantiles. These values are all slightly higher than for the males, where the estimates of the 50th quantiles are pre-obesity or obesity. The difference between the 10th and the 95th quantiles is quite large with a difference of about 20 units. This goes for all ages in the Table 11.

Table 11: QRCM Estimates of Females before/during/after WWII with the standard errors in the parenthesis.

Female before/during/after WWII			
Age and Height	10th quantile	50th quantile	95th quantile
79 Years, 157 cm	23.22(0.66)	29.29(0.57)	41.68(1.55)
73 Years, 165 cm	23.09(0.38)	30.11(0.34)	44.00(0.99)
60 years, 163 cm	22.52(0.22)	30.43(1.70)	46.10(0.75)

There are many studies of the development of BMI from the last hundred years in the U.S. This time line has been chosen because there has been rapid increase in BMI since the beginning of the 20th century [27]. In one article written by Komos and Brabec [21] in 2010, the trend of BMI is studied using US adults. They investigated the four groups, black and white, male and female, separately, with birth cohorts between 1882-1986. In their research they found that for the cohorts in the beginning of the 20th century the BMI started to increase and after the First World War the BMI values increased rapidly for all groups. The estimated rate of difference of BMI values of black females increased by 71%. The increase in BMI was rapid after the First World War but during the Great Depression and the Second World War, they found that it decreased. Then after the Second World War there was an increase in BMI again. This increase took off about a decade after the war, due to the fact that when the war ended there was a decline in income which according to Komos and Brabec resulted in a lower availability of labour saving technologies, and this caused people to stay at home to eat their meals.

The increase of BMI in the first half of the 20th century can be seen as a positive development for the well-being, as income increase and technological development elevated the health of the people. After the Second World War the BMI values increased again, but now a large group of the population reached so high values that it became a health hazard. This unhealthy increase in BMI has a positive correlation with the launch of fast food chains. According to Lin and Gurthrie [24], in 2008 the percent expenditure of food prepared away from home, that includes restaurants, fast-foods and take-out, was 41%. This corresponded to 32% of the calorie intake. Studies show that compared to food prepared at home, food prepared away from home has a lower nutrition level. This generates an increase in the calorie intake. The study also found that adults and children who consume food prepared away from home impairs the quality of their diet. A reduced diet can lead to obesity, cancer and other health conditions. The number of people with obesity in the United States is large compared to the rest of the world. In 2013 the percentage of adults aged over 20 with pre-obesity and obesity in the United States was 70.9% [10]. The world wide the percentage of obesity was 37% [25].

Figure 15 represents quantiles of BMI for male and females for the ages 40, 50 and 60. The left hand graph illustrates the quantiles of males for the ages 40, 50 and 60. From the graph we can observe that the quantiles for age 50 are slightly higher than for the other two age groups. Table 12 gives information about the estimates of the males. As seen in the left hand graph in Figure 15, the BMI estimates for age 50 with a height of 180 cm are higher for all quantiles than for the other ages. This might be due to that the 50-year-old male is taller than the 40 and 60-year-old males. The BMI values for the 50th quantile is, for a 40-year-old with a height of 169 cm, 28.11. The same estimate for a 50-year-old is 29.36 and for a 60-year-old with height 169 cm, the BMI value is 28.63. The quantiles for the females appears roughly the same for the different ages. In Table 13 some estimates of females from the graph are stated. The 50th quantile for a 40-year-old with height 167 cm has an estimated BMI value of 29.70. The estimated BMI values of the 50th for ages 50 and 60 are both 30.60. The 20th quantiles and above indicates obesity for both female and male. We then have to keep in mind that about 71% of the U.S. population is pre-obese or obese which is why the estimates are so high.

Figure 15: Quantiles of BMI for Males (left) and Females (right) of different ages.

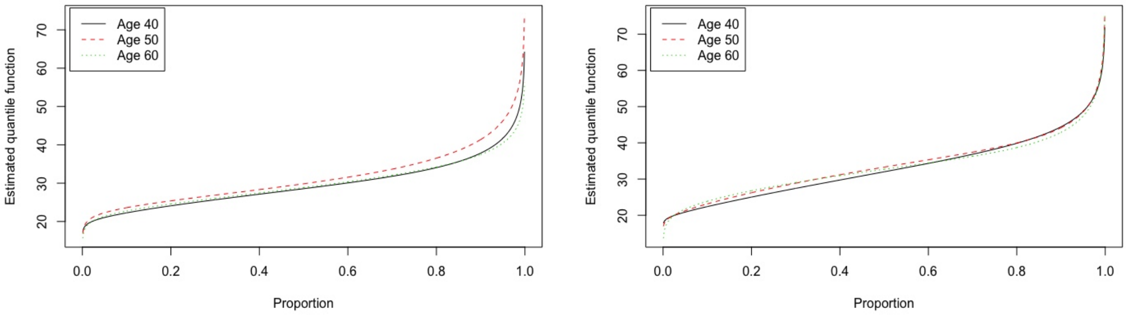


Table 12: QRCM Estimates of Males with the standard errors in the parenthesis.

Male			
Age and Height	10th quantile	50th quantile	95th quantile
40 Years, 169 cm	22.27(0.18)	28.11(0.18)	40.84(0.59)
50 Years, 180 cm	23.36(0.19)	29.36(0.20)	43.21(0.60)
60 years, 169 cm	22.93(0.19)	28.63(0.20)	40.04(0.64)

Table 13: QRCM Estimates of Females with the standard errors in the parenthesis.

Female			
Age and Height	10th quantile	50th quantile	95th quantile
40 Years, 167 cm	21.62(0.16)	29.70(0.25)	46.72(0.06)
50 Years, 170 cm	22.21(0.24)	30.60(0.32)	47.45(0.85)
60 years, 166 cm	22.57(0.24)	30.59(0.30)	46.40(0.82)

5 Discussion

In this thesis we used QRCM to estimate reference values of BMI given age, race, height and gender of any given quantile.

This section states the limitations that surfaced during the process of writing this thesis, some of my personal considerations about the results, and suggestions for future research.

The advantages of using QRCM to construct reference values are that it is an effective method to use in order to obtain accurate estimations of BMI quantiles. The QRCM method used in this thesis gives both the graphs of the quantiles for a given individual, and the estimates of the conditional

BMI quantiles given a set of covariate values. The procedure to calculate the estimates was simple as the implementation in R did not require many steps to obtain the results. From the data we could perform QRCM and from the result obtain the estimates. This method has an advantage in parsimony to calculate reference values compared to the 2000 CDC reference values. In the latter method, the calculations were more complex and the z-score value obtained that corresponded to a percentile was taken from a normal distribution.

The data set used in the analysis contained only four explanatory variables, so we were restricted to study the effect of these variables on the outcome, BMI. Of course, there are other factors that explain the BMI values, such as socioeconomic status and place of living. DNA can also have a key role in explaining extreme values of BMI, both high and low. Body mass index is measured in height and weight. Because height was included in the models, weight was an unnecessary variable to include in the data as BMI is already a measure of weight relative to height.

Frumento and Bottai [3] used Monte Carlo simulation to obtain the estimates. In these data analyses we chose not to use a bootstrap method because the sample size was large for the large-sample approximation to be reliable.

A common factor for all of the BMI estimates that we have studied is that the estimates of the 50th quantile are defined as pre-obese or obese. The data that we have used is collected from the United States and the percentage of obesity of adults is 70.9%, and this can explain why we got so high values of BMI in the 50th quantiles. If we used data from Sweden the reference values would probably be lower for all of the quantiles. This is why it is necessary to compare a person's BMI value with reference values calculated for the individual's country of living. There are cultural differences from country to country that have an impact on the residents' BMI values, such as differences in supply of fast-food chains, the economic status in the country and many other factors.

5.1 Future work

Using QRCM to estimate BMI is a good approach because it makes it possible to obtain the estimates for all quantiles. It would be a good method for health researchers to use: not only can they study the median of the population but also the other quantiles. To know the difference in BMI estimates between the different quantiles gives more information on the population. For example, the government is contemplating introducing more health class and physical education in schools and they know that for children of a specific age group, the median, are normal weight, they decide that this is not necessary since the children already are normal weight. If they were to study the 95th quantile, that is 5% of the students, and discover that they are of obesity class III and that the 10th quantile, 10% are underweight, they know that an action must be taken to help the children be normal weight.

In this thesis we have learnt about the theories of quantile regression coefficients models but not mentioned the more general quantile parametric models. Future work on this subject can be applied using parametric modelling. Using QRCM to estimate BMI can be developed further to include even more explanatory variables that could have a significant effect on BMI, for example socioeconomic status, number of meals a month that is food not prepared from home and the health status. Then even more reliable results could be obtained which will give more information about the overall health status in a country or region. To be able to use the reference values throughout the world, data needs to be collected for the inhabitants in each country and QRCM can be performed on the data to obtain accurate estimates of BMI.

References

[1] Balistreri, K.S. and Van Hook, J. (2009). Socioeconomic Status and Body Mass Index Among Hispanic Children of Immigrants and Children of Natives. *American Journal of Public Health*, Vol99, No.12: 2238-2246

[2] Barlow, S.E. and Dietz, W.H. (1998). Obesity Evaluation and Treatment: Expert Committee Recommendations Sarah E. Barlow, William H. Dietz *Pediatrics* Sep 1998, 102 (3) e29; DOI: 10.1542/peds.102.3.e29

[3] Bottai, M. and Frumento, P. (2016). Parametric Modeling of Quantile Regression Coefficient Functions. *Biometrics*, **72**, 74-84

[4] Bottai, M. and Frumento, P. (2017). Parametric Modeling of Quantile Regression Coefficient Functions with Censored and Truncated Data. *Biometrics*, **73**, 1179-1188

[5] Cade, B.S. and Noon, B.R. (2003). A gentle introduction to quantile regression for ecologists. *Front Ecol Environ The Ecological Society of America*, **1(8)**, 412-420

[6] Cameron, A.C. and Trivedi, P.K. (2005). Microeconometrics: Methods and Appliations. New York: Cambridge University Press.

[7] Centers for Disease and Control and Prevention. (2018). National Health and Nutrition Examination Survey. [online] Available at: https://www.cdc.gov/nchs/nhanes/index.htm. [Accessed 13 March 2018].

[8] Centers for Disease and Control and Prevention. (2017). National Health and Nutrition Examination - Survey 2015-2016 Data Documentation, Codebook, and Frequencies. [online] Available at: https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DEMO_I.htm. [Accessed 13 March 2018]

[9] Centers for Disease and Control and Prevention. (2017). National Health and Nutrition Examination - Survey 2015-2016 Data Documentation, Codebook, and Frequencies. [online] Available at: https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/BMX_I.htm. [Accessed 13 March 2018].

[10] Centers for Disease and Control and Prevention. (2017). National Health and Nutrition Examination - Obesity and Overweight. [online] Available at: https://www.cdc.gov/nchs/fastats/obesity-overweight.htm [Accessed 22 April 2018].

[11] Coxford, R. (2016). Restricted Cubic Spline Regression: A Brief Introduction. Institute for Clinical Evaluative Sciences, Paper 5621

[12] Deaton, A. (1997). The Analysis of Household Surveys. *Baltimore: Johns Hopkins.*

[13] Dunham, J.B, Cade, B.S. and Terrell, J.W. (2002). *Influences of spatial and temporal variation on fish-habitat relationships defined by regression quantiles.* Transactions of the American Fishers Society, **131**, 86-98

[14] Frumento, P. (2017). Quantile Regression Coefficients Modeling (qrcm). R package version 2.1 [online] Available at https://cran.r-project.org/web/packages/qrcm/index.html [Accessed 28 Mars 2018]

[15] Hao, L. and Naiman, D.Q. (2007). Quantile Regression. Thousand Oaks, CA: Sage

[16] Harris, T. (1950). Regression using minimum absolute deviations. *American Statistician* 4:14-15.

[17] Heyde, C.C. and Seneta, E. (2013). Statisticians of the Centuries. *Springer Science Business Media*, p. 82-84.

[18] Kirby, J., Liang, L., Chen, H-J. and Wang, Y. (2012). Race, Place and Obesity: The Complex Relationships Among Community Racial/Ethnic Composition, Individual Race/Ethnicity, and Obesity in the United States. *American Journal of Public Health*, 102(8): 1572-1578

[19] Koenker, R (2000)Quantile regression. *Department of Economics, University of Illinois*, Urbana-Champaign, Champaign. Version: 10 Nov. 2000. Available at:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.394.3210 &rep=rep1&type=pdf

[20] Koenker, R. and Bassett, G. Jr. (1978). Refression Quantiles. *Econometrica* Vol. 46, Nr 1, 33-50

[21] Komlos, J. and Brabec, M. (2010). The Trend of Mean BMI Values of US Adults, Birth Cohorts 1882-1986 that the Obesity Epidemic Began Earlier than Hitherto Thought. *American Journal of Human Biology* , **22**, 631-638

[22] Kromeyer-Hauschild, K., Wabitsch, M., Kunze. D., et al.(2001). Percentiles of body mass index in children and adolescents evaluated from different regional German studies. *Monatsschr Kinderh*, 149(8), 807-18.

[23] Lê Cook, B. and Manning, W.G. (2013). Thinking beyond the mean: a practical guide for using quantile regression methods for health and services research. *Shanghai Arch Psychiatry*, **25**: 55-59

[24] Lin, B-H. and Guthrie, J. (2012). Nutritional Quality of Food Prepared at Home and Away From Home, 1977-2008. EIB-105, U.S Department of Agriculture, Economic Research Service, December 2012. https://ageconsearch.umn.edu/bitstream/142361/2/EIB-105.pdf

[25] Ng, Marie et al. (2013). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* , Volume 384 , Issue 9945 , 766 - 781

[26] Nysom, K., Molgaard, C., Hutchings, B. and Michaelsen, K.F.(2001). Body mass index of 0 to 45-y-old Danes: reference values and comparison with published European reference values. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity*, 25(2), 177-84.

[27] Voxeu.org. (2010). 100 years of US obesity | VOX, CEPR's Policy Portal. [online] Available at: https://voxeu.org/article/100-years-us-obesity_ftn1 [Accessed 11 Apr. 2018].

[28] Wagner, H.M. (1959). Linear programming techniques for regression analysis.*Journal of the American Statistical Association* 54:206-212.

[29] Wei, Y., Pere, A., Koenker, R. and He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, **25**, 1369-1382

[30] Wooldridge, J.M. (2010). Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge, MA: MIT Press.

[31] World Healt Organization. 2018. Body mass index-BMI. [online] Available at: http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi. [Accessed 21 April 2018].

[32] Zhang, L., Samet, J., Caffo, B. and Punjabi N.M. (2006). Cigarette Smoking and Nocturnal Sleep Architecture. *American Journal of Epidemiology*, Vol 164, Issue 6.

Appendix

A1 ILM Estimates

In tables 14-23 the estimates of the β coefficients for respectively race for the different genders are presented.

Table 14: Summary of estimates of β coefficients for Mexican-American Males. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Intercept	190 (11)	43 (31)	9.2 (60)	-21 (78)	35 (46)	$5.0{\cdot}10^{-5}$
Height Spline 1	-2.1 (1.3)	-0.5 (0.4)	-0.2 (0.7)	0.2 (0.9)	-0.5 (0.6)	$3.1{\cdot}10^{-13}$
Height Spline 2	$5.3{\cdot}10^{-3}$ $(3.5{\cdot}10^{-4})$	$1.3{\cdot}10^{-3}$ $(9.9{\cdot}10^{-4})$	$5.0{\cdot}10^{-4}$ $(1.9{\cdot}10^{-3})$	$-1.4{\cdot}10^{-4}$ $(2.5{\cdot}10^{-3})$	$1.3{\cdot}10^{-3}$ $(1.5{\cdot}10^{-3})$	$3.1{\cdot}10^{-13}$
Height Spline 3	$-3.8{\cdot}10^{-6}$ $(2.7{\cdot}10^{-6})$	$-9.7{\cdot}10^{-7}$ $(7.5{\cdot}10^{-7})$	$-4.4{\cdot}10^{-7}$ $(1.4{\cdot}10^{-6})$	$1.4{\cdot}10^{-8}$ $(1.9{\cdot}10^{-6})$	$-9.7{\cdot}10^{-7}$ $(1.1{\cdot}10^{-6})$	$3.1{\cdot}10^{-13}$
Age Spline 1	-0.4 (0.6)	-0.2 (0.2)	-0.4 (0.3)	-0.3 (0.4)	-0.3 (0.2)	$<2{\cdot}10^{-16}$
Age Spline 2	$4.0{\cdot}10^{-3}$ $(4.9{\cdot}10^{-3})$	$1.2{\cdot}10^{-3}$ $(1.5{\cdot}10^{-3})$	$2.6{\cdot}10^{-3}$ $(2.7{\cdot}10^{-3})$	$1.6{\cdot}10^{-3}$ $(3.5{\cdot}10^{-3})$	$1.5{\cdot}10^{-3}$ $(2.0{\cdot}10^{-3})$	$<2{\cdot}10^{-16}$
Age Spline 3	$-8.5{\cdot}10^{-6}$ $(1.1{\cdot}10^{-5})$	$-2.4{\cdot}10^{-6}$ $(3.5{\cdot}10^{-6})$	$-4.7{\cdot}10^{-6}$ $(6.3{\cdot}10^{-6})$	$-2.7{\cdot}10^{-6}$ $(8.2{\cdot}10^{-6})$	$-2.1{\cdot}10^{-6}$ $(4.7{\cdot}10^{-6})$	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	0.02	$7.9{\cdot}10^{-12}$	0.4	0.4	

Table 15: Summary of estimates of β coefficients for Other Hispanic Males. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Inter-cept	-8.5 (150)	-31 (49)	320 (110)	130 (120)	8.4 (81)	$1.6{\cdot}10^{-7}$
Height Spline 1	-2.1 (1.8)	0.2 (0.6)	0.3 (1.3)	-0.5 (1.4)	-1.6 (1.0)	$<2{\cdot}10^{-16}$
Height Spline 2	$-4.3{\cdot}10^{-5}$ ($4.9{\cdot}10^{-3}$)	$-6.6{\cdot}10^{-4}$ ($1.5{\cdot}10^{-3}$)	$1.6{\cdot}10^{-3}$ ($3.4{\cdot}10^{-3}$)	$4.5{\cdot}10^{-3}$ ($3.4{\cdot}10^{-3}$)	$1.1{\cdot}10^{-3}$ ($2.6{\cdot}10^{-3}$)	$< 2{\cdot}10^{-16}$
Height Spline 3	$-1.9{\cdot}10^{-8}$ ($3.9{\cdot}10^{-6}$)	$4.5{\cdot}10^{-7}$ ($1.2{\cdot}10^{-6}$)	$-1.4{\cdot}10^{-7}$ ($2.5{\cdot}10^{-6}$)	$-3.5{\cdot}10^{-6}$ ($2.8{\cdot}10^{-6}$)	$-9.7{\cdot}10^{-7}$ ($1.9{\cdot}10^{-6}$)	$< 2{\cdot}10^{-16}$
Age Spline 1	-0.1 (2.0)	-0.3 (0.7)	-0.6 (0.6)	-0.3 (0.9)	-0.7 (0.6)	$<2{\cdot}10^{-16}$
Age Spline 2	$1.2{\cdot}10^{-2}$ ($2.2{\cdot}10^{-2}$)	$3.2{\cdot}10^{-3}$ ($8.4{\cdot}10^{-3}$)	$4.9{\cdot}10^{-3}$ ($5.1{\cdot}10^{-3}$)	$1.3{\cdot}10^{-3}$ ($9.0{\cdot}10^{-3}$)	$6.4{\cdot}10^{-3}$ ($5.8{\cdot}10^{-3}$)	$<2{\cdot}10^{-16}$
Age Spline 3	$-3.1{\cdot}10^{-5}$ ($6.3{\cdot}10^{-5}$)	$-8.2{\cdot}10^{-6}$ ($2.4{\cdot}10^{-5}$)	$-1.2{\cdot}10^{-5}$ ($1.3{\cdot}10^{-5}$)	$-1.7{\cdot}10^{-6}$ ($2.5{\cdot}10^{-5}$)	$-1.5{\cdot}10^{-5}$ ($1.6{\cdot}10^{-5}$)	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	0.0	0.0	0.0	0.0	

Table 16: Summary of estimates of β coefficients for Non-Hispanic White Males. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Inter-cept	130 (73)	23 (21)	15 (49)	28 (59)	42 (35)	$1.2{\cdot}10^{-8}$
Height Spline 1	-1.2 (0.8)	-0.3 (0.2)	-0.2 (0.6)	-0.4 (0.7)	-0.5 (0.4)	$1.1{\cdot}10^{-13}$
Height Spline 2	$2.8{\cdot}10^{-3}$ ($2.1{\cdot}10^{-3}$)	$6.2{\cdot}10^{-4}$ ($6.2{\cdot}10^{-3}$)	$4.9{\cdot}10^{-4}$ ($1.4{\cdot}10^{-3}$)	$1.0{\cdot}10^{-3}$ ($1.7{\cdot}10^{-3}$)	$1.3{\cdot}10^{-3}$ ($1.0{\cdot}10^{-3}$)	$1.1{\cdot}10^{-13}$
Height Spline 3	$-1.9{\cdot}10^{-6}$ (0.3)	$-4.3{\cdot}10^{-7}$ ($9.0{\cdot}10^{-2}$)	$-3.8{\cdot}10^{-7}$ (0.2)	$-7.9{\cdot}10^{-7}$ (0.3)	$-9.4{\cdot}10^{-7}$ ($7.6{\cdot}10^{-7}$)	$1.1{\cdot}10^{-13}$
Age Spline 1	0.3 (0.3)	$5.3{\cdot}10^{-2}$ ($9.0{\cdot}10^{-2}$)	-0.1 (0.2)	-0.3 (0.3)	$-8.3{\cdot}10^{-2}$ (0.2)	$<2{\cdot}10^{-16}$
Age Spline 2	$-2.7{\cdot}10^{-3}$ ($2.3{\cdot}10^{-3}$)	$-6.2{\cdot}10^{-4}$ ($7.8{\cdot}10^{-4}$)	$4.6{\cdot}10^{-4}$ ($1.6{\cdot}10^{-3}$)	$2.2{\cdot}10^{-3}$ ($2.3{\cdot}10^{-3}$)	$9.2{\cdot}10^{-5}$ ($1.5{\cdot}10^{-3}$)	$<2{\cdot}10^{-16}$
Age Spline 3	$7.0{\cdot}10^{-6}$ ($6.7{\cdot}10^{-6}$)	$1.9{\cdot}10^{-6}$ ($1.8{\cdot}10^{-6}$)	$5.7{\cdot}10^{-7}$ ($3.8{\cdot}10^{-6}$)	$-3.8{\cdot}10^{-6}$ ($5.4{\cdot}10^{-6}$)	$1.1{\cdot}10^{-6}$ ($3.5{\cdot}10^{-6}$)	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	$6.3{\cdot}10^{-16}$	$<2{\cdot}10^{-16}$	0.5	0.5	

Table 17: Summary of estimates of β coefficients for Non-Hispanic Black Males. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Intercept	240 (400)	57 (120)	150 (750)	170 (690)	130 (320)	$3.7{\cdot}10^{-8}$
Height Spline 1	-2.5 (4.2)	-0.6 (1.2)	-1.6 (7.7)	-1.9 (7.1)	-1.4 (3.3)	$7.8{\cdot}10^{-15}$
Height Spline 2	$5.9{\cdot}10^{-3}$ $(9.1{\cdot}10^{-3})$	$1.5{\cdot}10^{-3}$ $(2.6{\cdot}10^{-3})$	$3.9{\cdot}10^{-3}$ $(1.7{\cdot}10^{-2})$	$4.6{\cdot}10^{-3}$ $(1.6{\cdot}10^{-2})$	$3.5{\cdot}10^{-3}$ $(7.3{\cdot}10^{-3})$	$7.8{\cdot}10^{-15}$
Height Spline 3	$-4.0{\cdot}10^{-6}$ $(6.0{\cdot}10^{-6})$	$-1.0{\cdot}10^{-6}$ $(1.7{\cdot}10^{-6})$	$-2.8{\cdot}10^{-6}$ $(1.2{\cdot}10^{-5})$	$-3.3{\cdot}10^{-6}$ $(1.1{\cdot}10^{-5})$	$-2.4{\cdot}10^{-6}$ $(4.9{\cdot}10^{-6})$	$7.8{\cdot}10^{-15}$
Age Spline 1	-0.4 (0.6)	-0.1 (0.2)	$-4.0{\cdot}10^{-2}$ (0.6)	$1.6{\cdot}10^{-2}$ (0.6)	-0.3 (0.4)	$<2{\cdot}10^{-16}$
Age Spline 2	$3.3{\cdot}10^{-3}$ $(5.8{\cdot}10^{-3})$	$7.7{\cdot}10^{-4}$ $(1.6{\cdot}10^{-3})$	$-2.3{\cdot}10^{-4}$ $(5.6{\cdot}10^{-3})$	$-4.0{\cdot}10^{-4}$ $(5.4{\cdot}10^{-3})$	$2.1{\cdot}10^{-3}$ $(3.7{\cdot}10^{-3})$	$<2{\cdot}10^{-16}$
Age Spline 3	$-7.6{\cdot}10^{-6}$ $(1.5{\cdot}10^{-5})$	$-1.4{\cdot}10^{-6}$ $(4.{\cdot}10^{-6})$	$2.1{\cdot}10^{-6}$ $(1.5{\cdot}10^{-5})$	$1.8{\cdot}10^{-6}$ $(1.4{\cdot}10^{-5})$	$-3.9{\cdot}10^{-6}$ $(9.6{\cdot}10^{-6})$	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	0.0	$3.8\ {\cdot}10^{-11}$	0.6	0.6	

Table 18: Summary of estimates of β coefficients for Other Race including Multi-Racial Males. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Intercept	130 (100)	30 (27)	-2.6 (68)	-85 (90)	-8.4 (44)	0.1
Height Spline 1	-1.4 (1.1)	-0.4 (0.3)	$-5.5{\cdot}10^{-2}$ (0.8)	0.9 (1.0)	$-9.9{\cdot}10^{-3}$ (0.5)	$<2{\cdot}10^{-16}$
Height Spline 2	$3.7{\cdot}10^{-3}$ $(2.9{\cdot}10^{-3})$	$1.1{\cdot}10^{-3}$ $(8.0{\cdot}10^{-4})$	$3.2{\cdot}10^{-4}$ $(2.0{\cdot}10^{-3})$	$-2.2{\cdot}10^{-6}$ $(2.7{\cdot}10^{-3})$	$3.1{\cdot}10^{-4}$ $(1.4{\cdot}10^{-3})$	$<2{\cdot}10^{-16}$
Height Spline 3	$-2.3{\cdot}10^{-6}$ $(2.1{\cdot}10^{-6})$	$-8.4{\cdot}10^{-7}$ $(5.9{\cdot}10^{-7})$	$-3.8{\cdot}10^{-7}$ $(1.5{\cdot}10^{-6})$	$1.5{\cdot}10^{-6}$ $(2.0{\cdot}10^{-6})$	$-4.0{\cdot}10^{-7}$ $(1.1{\cdot}10^{-6})$	$<2{\cdot}10^{-16}$
Age Spline 1	-0.2 (0.6)	$-7.6{\cdot}10^{-2}$ (0.2)	$-6.3{\cdot}10^{-2}$ (0.3)	$1.2{\cdot}10^{-2}$ (0.4)	-0.2 (0.3)	$<2{\cdot}10^{-16}$
Age Spline 2	$2.7{\cdot}10^{-3}$ $(5.3{\cdot}10^{-3})$	$8.4{\cdot}10^{-4}$ $(1.6{\cdot}10^{-3})$	$5.7{\cdot}10^{-4}$ $(2.3{\cdot}10^{-3})$	$-1.4{\cdot}10^{-3}$ $(3.0{\cdot}10^{-3})$	$1.5{\cdot}10^{-3}$ $(2.3{\cdot}10^{-3})$	$<2{\cdot}10^{-16}$
Age Spline 3	$-7.8{\cdot}10^{-6}$ $(1.3{\cdot}10^{-5})$	$-2.2{\cdot}10^{-6}$ $(4.0{\cdot}10^{-6})$	$-1.0{\cdot}10^{-6}$ $(5.3{\cdot}10^{-6})$	$4.2{\cdot}10^{-6}$ $(6.9{\cdot}10^{-6})$	$-3.0{\cdot}10^{-6}$ $(5.5{\cdot}10^{-6})$	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	0.0	$1.1{\cdot}10^{-7}$	0.6	0.6	

Table 19: Summary of estimates of β coefficients for Mexican-American Females. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Intercept	190 (1200)	34 (360)	400 (3100)	440 (2900)	170 (1100)	0.0
Height Spline 1	-1.9 (14)	-0.4 (4.1)	-4.6 (34)	-5.0 (32)	-1.9 (12)	$5.7{\cdot}10^{-14}$
Height Spline 2	$4.6{\cdot}10^{-3}$ ($3.6{\cdot}10^{-2}$)	$9.2{\cdot}10^{-4}$ ($1.1{\cdot}10^{-2}$)	$1.1{\cdot}10^{-2}$ ($8.3{\cdot}10^{-2}$)	$1.2{\cdot}10^{-2}$ ($7.8{\cdot}10^{-2}$)	$4.6{\cdot}10^{-3}$ ($3.1{\cdot}10^{-2}$)	$5.7{\cdot}10^{-14}$
Height Spline 3	$-3.4{\cdot}10^{-6}$ ($2.7{\cdot}10^{-5}$)	$-7.2{\cdot}10^{-7}$ ($7.9{\cdot}10^{-6}$)	$-8.6{\cdot}10^{-6}$ ($6.0{\cdot}10^{-5}$)	$-9.1{\cdot}10^{-6}$ (5.610^{-5})	$-3.4{\cdot}10^{-6}$ ($2.2{\cdot}10^{-5}$)	$\cdot10^{-6}5.7\cdot 10^{-14}$
Age Spline 1	1.1 (0.7)	0.3 (0.2)	-0.3 (1.1)	-0.6 (1.2)	$7.6{\cdot}10^{-2}$ (0.5)	$<2{\cdot}10^{-16}$
Age Spline 2	$-8.2{\cdot}10^{-3}$ ($6.2{\cdot}10^{-3}$)	$-2.4{\cdot}10^{-3}$ ($1.9{\cdot}10^{-3}$)	$1.9{\cdot}10^{-3}$ ($9.1{\cdot}10^{-3}$)	$4.9{\cdot}10^{-3}$ ($9.0{\cdot}10^{-3}$)	$-5.8{\cdot}10^{-4}$ ($4.3{\cdot}10^{-3}$)	$<2{\cdot}10^{-16}$
Age Spline 3	$1.7{\cdot}10^{-5}$ ($1.5{\cdot}10^{-5}$)	$5.2{\cdot}10^{-6}$ ($4.7{\cdot}10^{-6}$)	$-4.0{\cdot}10^{-6}$ ($2.0{\cdot}10^{-5}$)	$-1.1{\cdot}10^{-5}$ ($2.2{\cdot}10^{-5}$)	$1.3{\cdot}10^{-6}$ ($9.7{\cdot}10^{-6}$)	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	0.0	$1.4{\cdot}10^{-8}$	0.9	0.9	

Table 20: Summary of estimates of β coefficients for Other Hispanic Females. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Intercept	86 (190)	6.2 (52)	60 (160)	-5.5 (190)	3.7 (110)	0.2
Height Spline 1	-0.9 (2.4)	$-8.7{\cdot}10^{-2}$ (0.7)	-0.7 (1.9)	0.1 (2.3)	$-1.1{\cdot}10^{-2}$ (1.4)	0.0
Height Spline 2	$2.1{\cdot}10^{-3}$ ($6.8{\cdot}10^{-3}$)	$2.5{\cdot}10^{-4}$ ($1.9{\cdot}10^{-3}$)	$1.8{\cdot}10^{-3}$ ($5.2{\cdot}10^{-3}$)	$-4.1{\cdot}10^{-4}$ ($6.2{\cdot}10^{-3}$)	$-1.0{\cdot}10^{-4}$ ($3.8{\cdot}10^{-3}$)	0.0
Height Spline 3	$-1.7{\cdot}10^{-6}$ ($5.5{\cdot}10^{-6}$)	$-2.4{\cdot}10^{-7}$ ($1.5{\cdot}10^{-6}$)	$-1.4{\cdot}10^{-6}$ ($4.1{\cdot}10^{-6}$)	$3.8{\cdot}10^{-7}$ ($4.9{\cdot}10^{-6}$)	$1.1{\cdot}10^{-7}$ ($3.1{\cdot}10^{-6}$)	0.0
Age Spline 1	0.8 (0.5)	0.2 (0.2)	$5.3{\cdot}10^{-2}$ (0.3)	0.3 (0.5)	0.4 (0.3)	$<2{\cdot}10^{-16}$
Age Spline 2	$-7.3{\cdot}10^{-3}$ ($4.6{\cdot}10^{-3}$)	$-1.9{\cdot}10^{-3}$ ($1.3{\cdot}10^{-3}$)	$-1.5{\cdot}10^{-3}$ ($2.8{\cdot}10^{-3}$)	$-2.6{\cdot}10^{-3}$ ($4.2{\cdot}10^{-3}$)	$-4.5{\cdot}10^{-3}$ ($2.4{\cdot}10^{-3}$)	$<2{\cdot}10^{-16}$
Age Spline 3	$2.0{\cdot}10^{-5}$ ($1.1{\cdot}10^{-5}$)	$5.4{\cdot}10^{-6}$ ($3.2{\cdot}10^{-6}$)	$5.3{\cdot}10^{-5}$ ($6.6{\cdot}10^{-6}$)	$5.6{\cdot}10^{-6}$ ($9.9{\cdot}10^{-6}$)	$1.2{\cdot}10^{-5}$ ($5.5{\cdot}10^{-6}$)	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	0.0	$1.1{\cdot}10^{-5}$	0.2	0.2	

Table 21: Summary of estimates of β coefficients for Non-Hispanic White Females. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Inter-cept	170 (86)	39 (26)	-49 (58)	-61 (81)	15 (53)	$1.0{\cdot}10^{-10}$
Height Spline 1	-2.0 (1.0)	-0.5 (0.3)	0.6 (0.7)	-0.8 (1.0)	-0.2 (0.7)	$3.0{\cdot}10^{-16}$
Height Spline 2	$5.2{\cdot}10^{-3}$ ($2.8{\cdot}10^{-3}$)	$1.2{\cdot}10^{-3}$ ($8.1{\cdot}10^{-4}$)	$-1.8{\cdot}10^{-3}$ ($1.9{\cdot}10^{-3}$)	$-2.3{\cdot}10^{-3}$ ($2.7{\cdot}10^{-3}$)	$5.0{\cdot}10^{-4}$ ($1.8{\cdot}10^{-3}$)	$3.0{\cdot}10^{-16}$
Height Spline 3	$-3.8{\cdot}10^{-6}$ ($2.4{\cdot}10^{-6}$)	$-9.5{\cdot}10^{-7}$ ($6.3{\cdot}10^{-7}$)	$1.5{\cdot}10^{-6}$ ($1.5{\cdot}10^{-6}$)	$1.9{\cdot}10^{-6}$ ($2.1{\cdot}10^{-6}$)	$-3.5{\cdot}10^{-7}$ ($1.4{\cdot}10^{-16}$)	$3.0{\cdot}10^{-6}$
Age Spline 1	$-1.9{\cdot}10^{-2}$ (0.4)	$-1.0{\cdot}10^{-2}$ (0.1)	-0.1 (0.2)	0.1 (0.3)	$-8.7{\cdot}10^{-2}$ (0.2)	$<2{\cdot}10^{-16}$
Age Spline 2	$7.2{\cdot}10^{-4}$ ($3.3{\cdot}10^{-3}$)	$1.7{\cdot}10^{-4}$ ($9.0{\cdot}10^{-4}$)	$6.9{\cdot}10^{-4}$ ($1.8{\cdot}10^{-3}$)	$-9.4{\cdot}10^{-4}$ ($2.5{\cdot}10^{-3}$)	$7.6{\cdot}10^{-4}$ ($1.7{\cdot}10^{-3}$)	$<2{\cdot}10^{-16}$
Age Spline 3	$-3.4{\cdot}10^{-6}$ ($7.2{\cdot}10^{-6}$)	$-6.9{\cdot}10^{-7}$ ($2.1{\cdot}10^{-6}$)	$-1.4{\cdot}10^{-6}$ ($4.3{\cdot}10^{-6}$)	$1.4{\cdot}10^{-6}$ ($6.0{\cdot}10^{-6}$)	$-2.3{\cdot}10^{-6}$ ($3.9{\cdot}10^{-6}$)	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	0.0	$5.0{\cdot}10^{-12}$	0.0	0.0	

Table 22: Summary of estimates of β coefficients for Non-Hispanic Black Females. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Inter-cept	110 (130)	1.6 (35)	-39 (87)	5.8 (110)	5.2 (57)	$2.5{\cdot}10^{-5}$
Height Spline 1	-1.2 (1.6)	$-4.5{\cdot}10^{-2}$ (0.4)	0.5 (1.1)	$7.7{\cdot}10^{-5}$ (1.3)	-0.7 (0.7)	$8.8{\cdot}10^{-13}$
Height Spline 2	$3.3{\cdot}10^{-3}$ ($4.2{\cdot}10^{-3}$)	$2.0{\cdot}10^{-4}$ ($1.2{\cdot}10^{-3}$)	$-1.4{\cdot}10^{-3}$ ($2.8{\cdot}10^{-3}$)	$-2.4{\cdot}10^{-4}$ ($3.4{\cdot}10^{-3}$)	$1.8{\cdot}10^{-3}$ ($1.9{\cdot}10^{-3}$)	$8.8{\cdot}10^{-13}$
Height Spline 3	$-2.6{\cdot}10^{-6}$ ($3.3{\cdot}10^{-6}$)	$-2.2{\cdot}10^{-7}$ ($9.2{\cdot}10^{-7}$)	$1.1{\cdot}10^{-6}$ ($2.1{\cdot}10^{-6}$)	$2.7{\cdot}10^{-7}$ ($2.7{\cdot}10^{-6}$)	$-1.5{\cdot}10^{-6}$ ($1.5{\cdot}10^{-6}$)	$8.8{\cdot}10^{-13}$
Age Spline 1	0.3 (0.5)	$1.1{\cdot}10^{-2}$ (0.2)	0.2 (0.3)	0.6 (0.3)	0.3 (0.3	$<2{\cdot}10^{-16}$
Age Spline 2	$-2.8{\cdot}10^{-3}$ ($5.2{\cdot}10^{-3}$)	$-3.1{\cdot}10^{-4}$ ($1.5{\cdot}10^{-3}$)	$-2.2{\cdot}10^{-3}$ ($2.2{\cdot}10^{-3}$)	$-4.1{\cdot}10^{-3}$ ($2.9{\cdot}10^{-3}$)	$-2.6{\cdot}10^{-3}$ ($2.4{\cdot}10^{-3}$)	$<2{\cdot}10^{-16}$
Age Spline 3	$8.7{\cdot}10^{-6}$ ($1.4{\cdot}10^{-5}$)	$1.5{\cdot}10^{-6}$ ($4.1{\cdot}10^{-6}$)	$5.3{\cdot}10^{-6}$ ($5.2{\cdot}10^{-6}$)	$6.9{\cdot}10^{-6}$ ($6.7{\cdot}10^{-6}$)	$6.5{\cdot}10^{-6}$ ($6.3{\cdot}10^{-6}$)	$<2{\cdot}10^{-16}$
p-value	$<2{\cdot}10^{-16}$	0.0	$3.6{\cdot}10^{-11}$	0.0	0.0	

Table 23: Summary of estimates of β coefficients for Other Race including Multi-Racial Females. The standard errors are within the parenthesis. The p-values represent the significance of the null hypothesis that the value of θ in each column, is 0.

	Intercept	$log(p)$	$log(1-p)$	slp1	slp2	p-value
Inter-cept	340 (1200)	76 (370)	160 (2100)	74 (200)	140 (960)	$4.6 \cdot 10^{-5}$
Height Spline 1	-3.8 (14)	-0.9 (4.2)	-1.8 (23)	-0.7 (21)	-1.6 ($2.6 \cdot 10^{-2}$)	$5.5 \cdot 10^{-12}$
Height Spline 2	$9.7 \cdot 10^{-3}$ ($3.5 \cdot 10^{-2}$)	$2.3 \cdot 10^{-3}$ ($1.1 \cdot 10^{-2}$)	$4.5 \cdot 10^{-3}$ ($5.7 \cdot 10^{-2}$)	$1.4 \cdot 10^{-3}$ ($5.2 \cdot 10^{-2}$)	$4.0 \cdot 10^{-3}$ ($2.6 \cdot 10^{-2}$)	$5.5 \cdot 10^{-12}$
Height Spline 3	$-7.2 \cdot 10^{-6}$ ($2.5 \cdot 10^{-5}$)	$-1.7 \cdot 10^{-6}$ ($7.7 \cdot 10^{-6}$)	$-3.4 \cdot 10^{-6}$ ($4.1 \cdot 10^{-5}$)	$-8.4 \cdot 10^{-7}$ ($3.7 \cdot 10^{-5}$)	$-3.0 \cdot 10^{-6}$ ($1.9 \cdot 10^{-5}$)	$5.5 \cdot 10^{-12}$
Age Spline 1	0.5 (0.8)	$7.7 \cdot 10^{-2}$ (0.3)	0.2 (0.5)	0.2 (0.5)	0.1 (0.4)	$<2 \cdot 10^{-16}$
Age Spline 2	$- 5.2 \cdot 10^{-3}$ ($6.5 \cdot 10^{-3}$)	$-1.1 \cdot 10^{-3}$ ($2.1 \cdot 10^{-3}$)	$-1.8 \cdot 10^{-3}$ ($3.9 \cdot 10^{-3}$)	$-1.4 \cdot 10^{-3}$ ($4.2 \cdot 10^{-3}$)	$-1.3 \cdot 10^{-3}$ ($3.1 \cdot 10^{-3}$)	$<2 \cdot 10^{-16}$
Age Spline 3	$1.5 \cdot 10^{-5}$ ($1.4 \cdot 10^{-5}$)	$3.6 \cdot 10^{-6}$ ($4.8 \cdot 10^{-6}$)	$4.7 \cdot 10^{-6}$ ($8.3 \cdot 10^{-6}$)	$2.7 \cdot 10^{-6}$ ($9.7 \cdot 10^{-6}$)	$3.6 \cdot 10^{-6}$ ($6.8 \cdot 10^{-6}$)	$<2 \cdot 10^{-16}$
p-value	$<2 \cdot 10^{-16}$	0.1	0.0	0.9	0.9	

A2 Estimates from quantile parametric modelling

In tables 24-33 the estimates from the quantile parametric modelling is presented for race and gender separately with the corresponding graphs in figures 16-25.

Table 24: QRCM estimates of the 25th, 50th and 75th quantiles for Mexican-American Males.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	81.8	20.6	41.5	122.1
	0.50	84.3	18.6	47.9	120.6
	0.75	98.0	22.8	53.3	142.6
Height Spline 1	0.25	-0.8	0.2	-1.2	-0.3
	0.50	-0.8	0.2	-1.2	-0.4
	0.75	-1.0	0.3	-1.5	-0.4
Height Spline 2	0.25	$1.8{\cdot}10^{-3}$	$0.7{\cdot}10^{-3}$	$0.5{\cdot}10^{-3}$	$3.1{\cdot}10^{-3}$
	0.50	$2.0{\cdot}10^{-3}$	$0.6{\cdot}10^{-3}$	$0.8{\cdot}10^{-3}$	$3.2{\cdot}10^{-3}$
	0.75	$2.6{\cdot}10^{-3}$	$0.8{\cdot}10^{-3}$	$1.1{\cdot}10^{-3}$	$4.1{\cdot}10^{-3}$
Height Spline 3	0.25	$-1.2{\cdot}10^{-6}$	$5.1{\cdot}10^{-7}$	$-2.2{\cdot}10^{-6}$	$-2.1{\cdot}10^{-7}$
	0.50	$-1.3{\cdot}10^{-6}$	$4.9{\cdot}10^{-7}$	$-2.3{\cdot}10^{-6}$	$-3.8{\cdot}10^{-7}$
	0.75	$-1.8{\cdot}10^{-6}$	$6.0{\cdot}10^{-7}$	$-3.0{\cdot}10^{-6}$	$-6.3{\cdot}10^{-7}$
Age Spline 1	0.25	$4.7{\cdot}10^{-2}$	0.1	-0.1	0.2
	0.50	0.1	0.1	-0.1	0.2
	0.75	0.0	0.1	-0.2	0.3
Age Spline 2	0.25	$0.7{\cdot}10^{-3}$	$0.8{\cdot}10^{-3}$	$-0.9{\cdot}10^{-3}$	$2.2{\cdot}10^{-3}$
	0.50	$0.7{\cdot}10^{-3}$	$0.8{\cdot}10^{-3}$	$-0.9{\cdot}10^{-3}$	$2.3{\cdot}10^{-3}$
	0.75	$0.6{\cdot}10^{-3}$	$1.1{\cdot}10^{-3}$	$-1.4{\cdot}10^{-3}$	$2.3{\cdot}10^{-3}$
Age Spline 3	0.25	$-2.8{\cdot}10^{-6}$	$1.7{\cdot}10^{-6}$	$-6.2{\cdot}10^{-6}$	$6.6{\cdot}10^{-7}$
	0.50	$-3.1{\cdot}10^{-6}$	$1.9{\cdot}10^{-6}$	$-6.7{\cdot}10^{-6}$	$6.0{\cdot}10^{-7}$
	0.75	$-2.9{\cdot}10^{-6}$	$2.4{\cdot}10^{-6}$	$-7.6{\cdot}10^{-6}$	$1.7{\cdot}10^{-6}$

Figure 16: Plotted estimates of splines and intercept for Mexican-American Males.

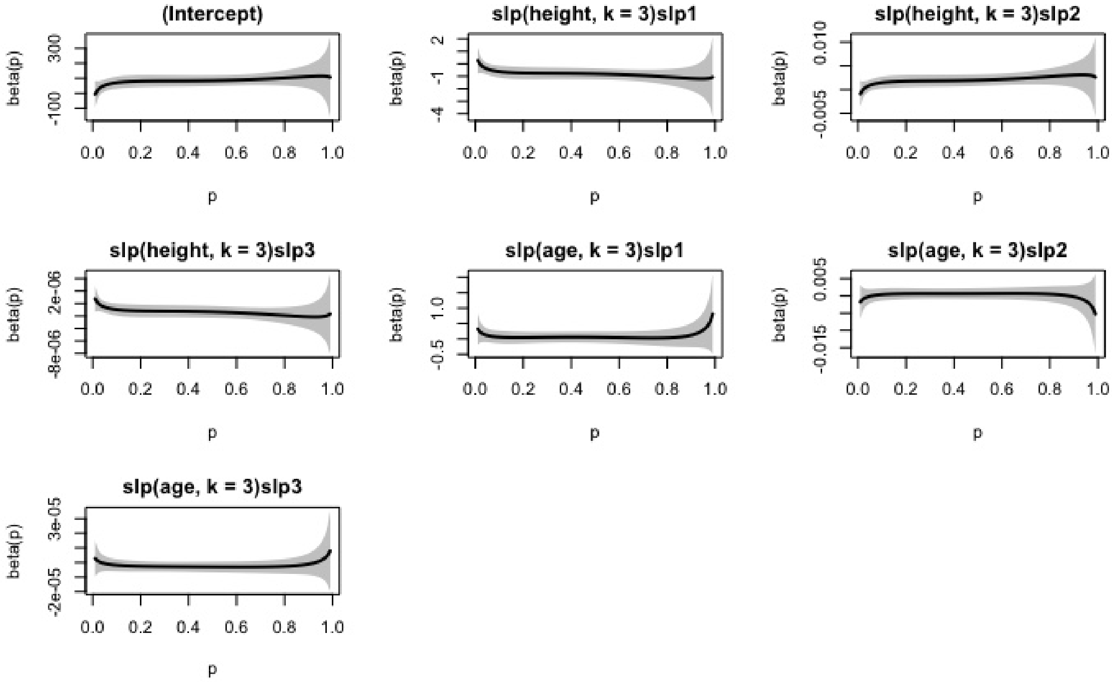


Table 25: QRCM estimates of the 25th, 50th and 75th quantiles for Other Hispanic Males.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	78.0	17.0	44.6	111.4
	0.50	103.5	19.4	19.4	141.4
	0.75	134.9	32.6	71.0	198.6
Height Spline 1	0.25	-0.7	0.2	-1.1	-0.3
	0.50	-1.0	0.2	-1.4	-0.5
	0.75	-1.4	0.4	-2.1	-0.6
Height Spline 2	0.25	$1.5 \cdot 10^{-3}$	$0.6 \cdot 10^{-3}$	$0.4 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$
	0.50	$2.2 \cdot 10^{-3}$	$0.6 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$
	0.75	$3.5 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$5.4 \cdot 10^{-3}$
Height Spline 3	0.25	$-9.2 \cdot 10^{-6}$	$4.3 \cdot 10^{-7}$	$-1.8 \cdot 10^{-6}$	$-7.5 \cdot 10^{-8}$
	0.50	$-1.5 \cdot 10^{-6}$	$4.4 \cdot 10^{-7}$	$-2.3 \cdot 10^{-6}$	$-6.0 \cdot 10^{-7}$
	0.75	$-2.4 \cdot 10^{-6}$	$7.1 \cdot 10^{-7}$	$-3.8 \cdot 10^{-6}$	$-1.1 \cdot 10^{-6}$
Age Spline 1	0.25	0.2	0.1	$3.2 \cdot 10^{-2}$	0.4
	0.50	0.4	0.1	0.2	0.6
	0.75	0.2	0.2	-0.1	0.6
Age Spline 2	0.25	$-0.7 \cdot 10^{-3}$	$0.9 \cdot 10^{-3}$	$-2.6 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$
	0.50	$-2.3 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$-4.2 \cdot 10^{-3}$	$-0.2 \cdot 10^{-3}$
	0.75	$-1.4 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$-4.7 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$
Age Spline 3	0.25	$-1.3 \cdot 10^{-7}$	$2.3 \cdot 10^{-6}$	$-4.6 \cdot 10^{-6}$	$4.3 \cdot 10^{-6}$
	0.50	$3.6 \cdot 10^{-6}$	$2.5 \cdot 10^{-6}$	$-1.2 \cdot 10^{-6}$	$8.5 \cdot 10^{-6}$
	0.75	$1.8 \cdot 10^{-6}$	$4.1 \cdot 10^{-6}$	$-6.2 \cdot 10^{-6}$	$9.9 \cdot 10^{-6}$

Figure 17: Plotted estimates of splines and intercept for Other Hispanic Males.

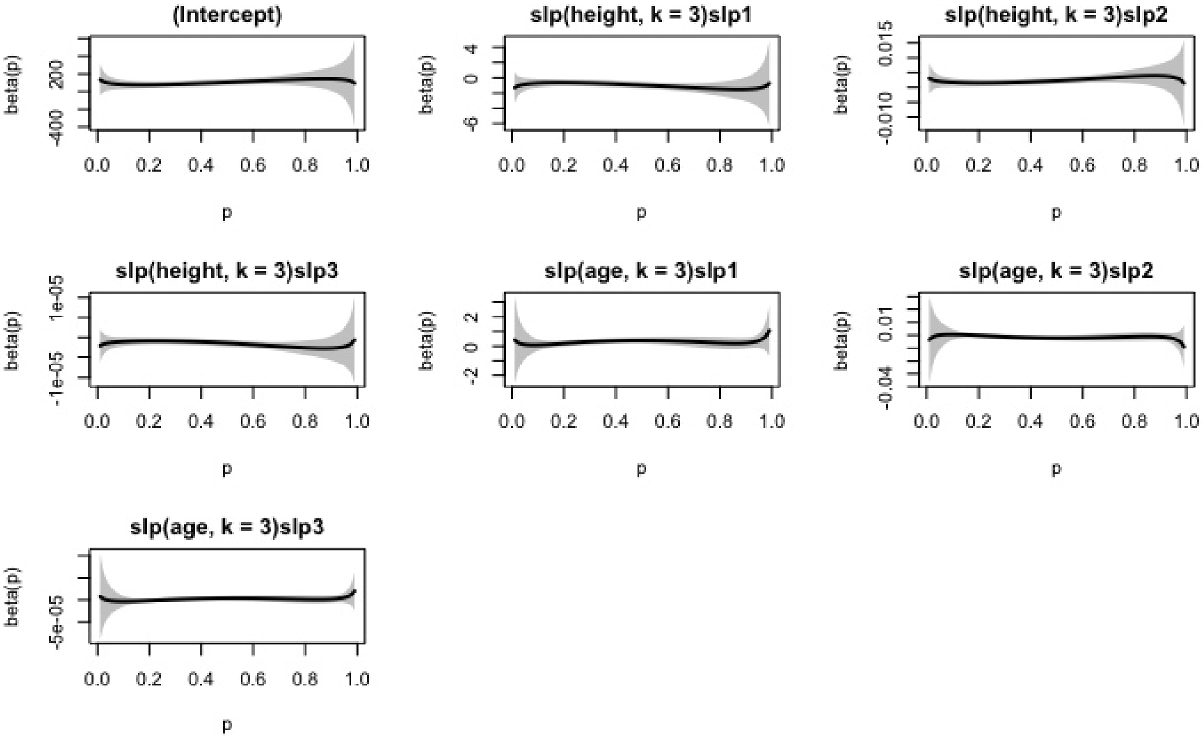


Table 26: QRCM estimates of the 25th, 50th and 75th quantiles for Non-Hispanic White Males.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	55.8	10.3	35.6	76.0
	0.50	63.8	12.7	38.9	88.6
	0.75	92.5	19.8	53.7	131.2
Height Spline 1	0.25	-0.4	-0.4	-0.7	-0.2
	0.50	-0.5	0.1	-0.8	-0.2
	0.75	-0.9	0.2	-1.3	-0.4
Height Spline 2	0.25	$0.9{\cdot}10^{-3}$	$0.3{\cdot}10^{-3}$	$0.3{\cdot}10^{-3}$	$1.5{\cdot}10^{-3}$
	0.50	$1.5{\cdot}10^{-3}$	$0.4{\cdot}10^{-3}$	$0.4{\cdot}10^{-3}$	$2.0{\cdot}10^{-3}$
	0.75	$2.1{\cdot}10^{-3}$	$0.6{\cdot}10^{-3}$	$0.9{\cdot}10^{-3}$	$3.2{\cdot}10^{-3}$
Height Spline 3	0.25	$-5.3{\cdot}10^{-7}$	$2.3{\cdot}10^{-7}$	$-9.7{\cdot}10^{-7}$	$-8.4{\cdot}10^{-8}$
	0.50	$-7.1{\cdot}10^{-7}$	$2.8{\cdot}10^{-7}$	$-1.3{\cdot}10^{-6}$	$-1.6{\cdot}10^{-7}$
	0.75	$-1.4{\cdot}10^{-6}$	$4.4{\cdot}10^{-7}$	$-2.2{\cdot}10^{-6}$	$-5.1{\cdot}10^{-7}$
Age Spline 1	0.25	0.2	0.1	0.1	0.4
	0.50	0.2	$8.0{\cdot}10^{-2}$	$4.7{\cdot}10^{-2}$	0.4
	0.75	0.1	0.1	-0.1	0.3
Age Spline 2	0.25	$-1.0{\cdot}10^{-3}$	$0.6{\cdot}10^{-3}$	$-2.2{\cdot}10^{-3}$	$0.2{\cdot}10^{-3}$
	0.50	$-5.5{\cdot}10^{-4}$	$0.7{\cdot}10^{-3}$	$-1.9{\cdot}10^{-3}$	$0.7{\cdot}10^{-3}$
	0.75	$1.5{\cdot}10^{-5}$	$0.9{\cdot}10^{-3}$	$-1.8{\cdot}10^{-3}$	$1.9{\cdot}10^{-3}$
Age Spline 3	0.25	$1.1{\cdot}10^{-6}$	$1.4{\cdot}10^{-6}$	$-1.5{\cdot}10^{-6}$	$3.8{\cdot}10^{-6}$
	0.50	$-8.4{\cdot}10^{-8}$	$1.5{\cdot}10^{-6}$	$-3.0{\cdot}10^{-6}$	$-3.0{\cdot}10^{-6}$
	0.75	$-1.2{\cdot}10^{-6}$	$2.1{\cdot}10^{-6}$	$-5.4{\cdot}10^{-6}$	$3.0{\cdot}10^{-6}$

Figure 18: Plotted estimates of splines and intercept for Non-Hispanic White Males.

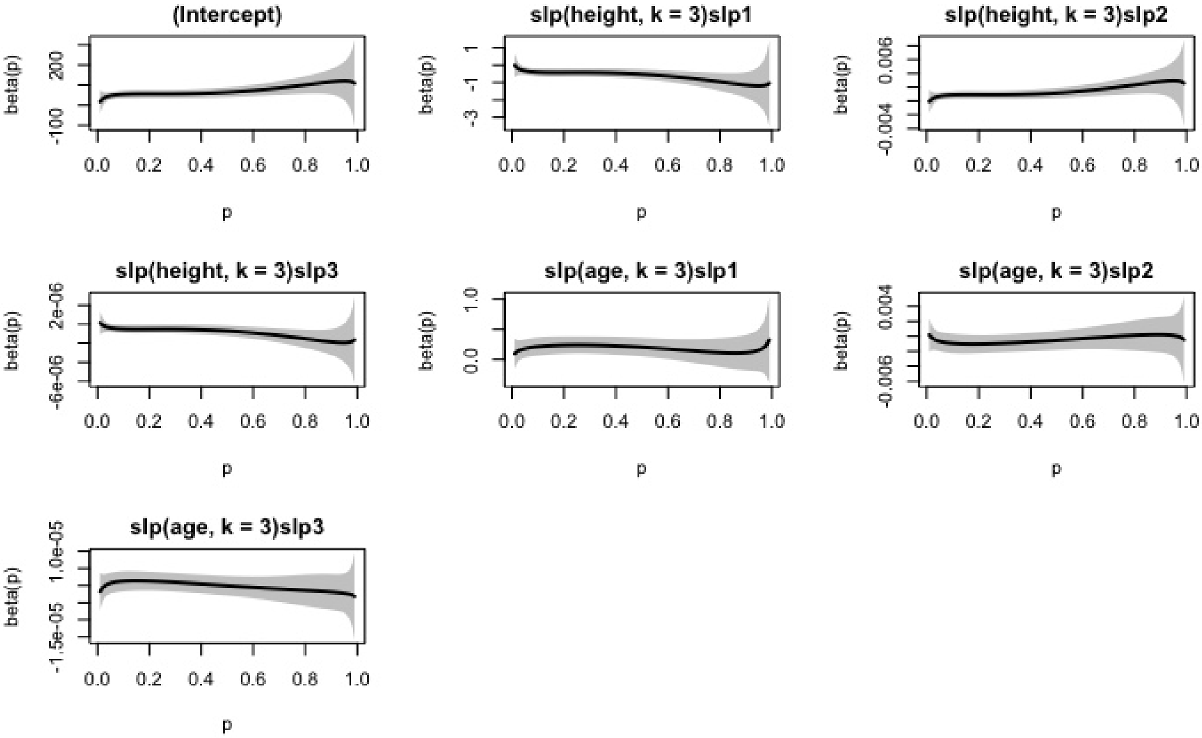


Table 27: QRCM estimates of the 25th, 50th and 75th quantiles for Non-Hispanic Black Males.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	60.9	12.5	36.4	85.4
	0.50	78.3	13.2	52.5	104.1
	0.75	133.6	35.1	64.9	202.3
Height Spline 1	0.25	-0.5	0.1	-0.8	-0.2
	0.50	-0.7	0.2	-1.0	-0.4
	0.75	-1.3	0.4	-2.1	-0.5
Height Spline 2	0.25	$1.0{\cdot}10^{-3}$	$0.4{\cdot}10^{-3}$	$0.3{\cdot}10^{-3}$	$1.8{\cdot}10^{-3}$
	0.50	$1.5{\cdot}10^{-3}$	$4.1{\cdot}10^{-3}$	$0.7{\cdot}10^{-3}$	$2.3{\cdot}10^{-3}$
	0.75	$3.1{\cdot}10^{-3}$	$1.1{\cdot}10^{-3}$	$1.0{\cdot}10^{-3}$	$5.1{\cdot}10^{-3}$
Height Spline 3	0.25	$-6.4{\cdot}10^{-7}$	$2.8{\cdot}10^{-7}$	$-1.2{\cdot}10^{-6}$	$-1.0{\cdot}10^{-7}$
	0.50	$-9.7{\cdot}10^{-7}$	$3.1{\cdot}10^{-7}$	$-1.6{\cdot}10^{-6}$	$-3.6{\cdot}10^{-7}$
	0.75	$-2.0{\cdot}10^{-6}$	$7.7{\cdot}10^{-7}$	$-3.5{\cdot}10^{-6}$	$-5.3{\cdot}10^{-7}$
Age Spline 1	0.25	0.2	0.1	$2.2 \cdot10^{-2}$	0.3
	0.50	0.3	0.1	0.1	0.5
	0.75	0.1	0.1	-0.2	0.4
Age Spline 2	0.25	$-0.3{\cdot}10^{-3}$	$0.7{\cdot}10^{-3}$	$-1.8{\cdot}10^{-3}$	$1.1{\cdot}10^{-3}$
	0.50	$-0.7{\cdot}10^{-3}$	$1.0{\cdot}10^{-3}$	$-2.7{\cdot}10^{-3}$	$1.2{\cdot}10^{-3}$
	0.75	$0.4{\cdot}10^{-3}$	$1.2{\cdot}10^{-3}$	$-2.0{\cdot}10^{-3}$	$2.7{\cdot}10^{-3}$
Age Spline 3	0.25	$-9.6{\cdot}10^{-6}$	$1.8{\cdot}10^{-6}$	$-4.4{\cdot}10^{-6}$	$2.5{\cdot}10^{-6}$
	0.50	$-4.2{\cdot}10^{-7}$	$2.5{\cdot}10^{-6}$	$-5.3{\cdot}10^{-6}$	$4.5{\cdot}10^{-6}$
	0.75	$-3.0{\cdot}10^{-6}$	$2.5{\cdot}10^{-6}$	$-8.0{\cdot}10^{-6}$	$2.0{\cdot}10^{-6}$

Figure 19: Plotted estimates of splines and intercept for Non-Hispanic Black Males.

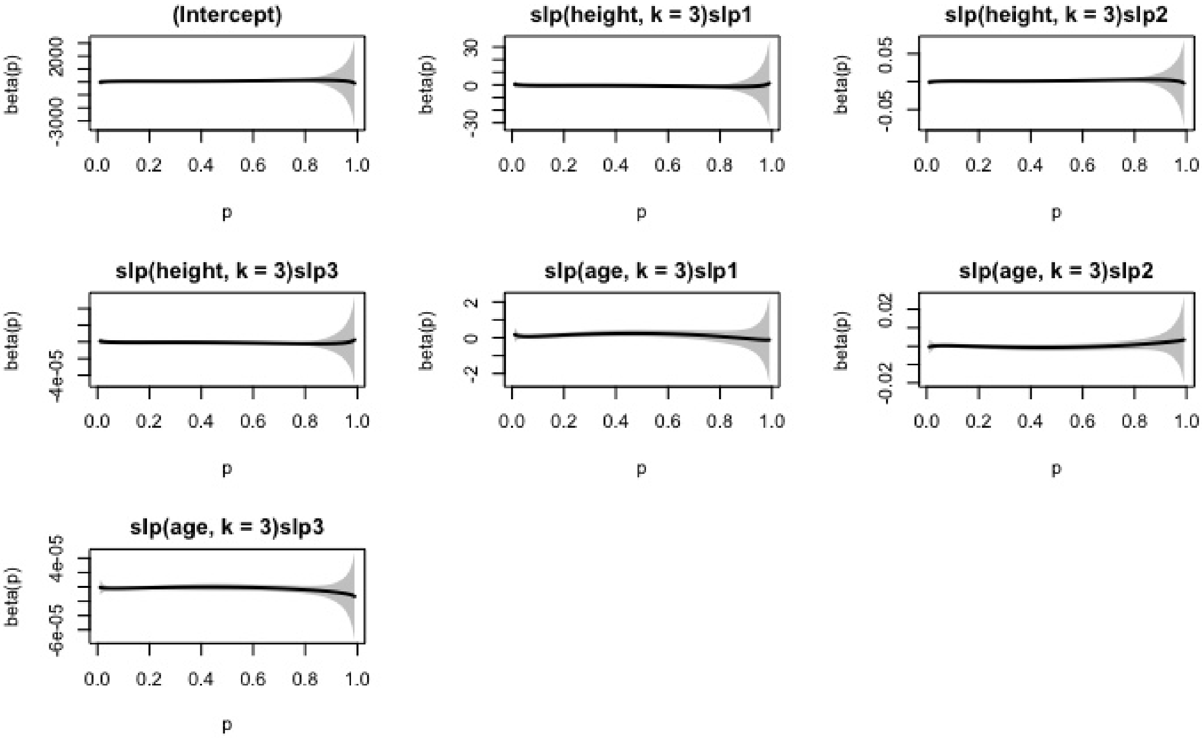


Table 28: QRCM estimates of the 25th, 50th and 75th quantiles for Other Race including Multiracial Males.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	52.3	20.0	13.1	91.6
	0.50	34.7	21.8	-8.0	77.4
	0.75	2.9	26.5	-49.1	55.0
Height Spline 1	0.25	-0.4	0.2	-0.8	0.1
	0.50	-0.1	0.3	-0.7	04
	0.75	0.2	0.3	-0.4	0.8
Height Spline 2	0.25	$6.9{\cdot}10^{-4}$	$0.6{\cdot}10^{-3}$	$-0.5{\cdot}10^{-3}$	$1.2{\cdot}10^{-3}$
	0.50	$7.8{\cdot}10^{-5}$	$7.0{\cdot}10^{-3}$	$-1.3{\cdot}10^{-3}$	$1.5{\cdot}10^{-3}$
	0.75	$-7.0{\cdot}10^{-4}$	$0.8{\cdot}10^{-3}$	$-2.3{\cdot}10^{-3}$	$0.9{\cdot}10^{-3}$
Height Spline 3	0.25	$-3.4{\cdot}10^{-7}$	$4.3{\cdot}10^{-7}$	$-1.2{\cdot}10^{-6}$	$5.0{\cdot}10^{-7}$
	0.50	$1.5{\cdot}10^{-7}$	$5.5{\cdot}10^{-7}$	$-9.3{\cdot}10^{-7}$	$1.2{\cdot}10^{-6}$
	0.75	$6.9{\cdot}10^{-7}$	$6.2{\cdot}10^{-7}$	$-5.3{\cdot}10^{-7}$	$2.0{\cdot}10^{-6}$
Age Spline 1	0.25	0.2	0.1	0.1	0.4
	0.50	0.3	0.1	0.1	0.5
	0.75	0.3	0.1	0.1	0.6
Age Spline 2	0.25	$-1.0{\cdot}10^{-3}$	$0.6{\cdot}10^{-3}$	$-2.2{\cdot}10^{-3}$	$0.2{\cdot}10^{-3}$
	0.50	$-1.2{\cdot}10^{-3}$	$0.9{\cdot}10^{-3}$	$-3.4{\cdot}10^{-3}$	$-0.2{\cdot}10^{-3}$
	0.75	$-2.2{\cdot}10^{-3}$	$1.0{\cdot}10^{-3}$	$-4.1{\cdot}10^{-3}$	$-0.2{\cdot}10^{-3}$
Age Spline 3	0.25	$1.1{\cdot}10^{-6}$	$1.4{\cdot}10^{-6}$	$-1.6{\cdot}10^{-6}$	$3.9{\cdot}10^{-6}$
	0.50	$3.3{\cdot}10^{-6}$	$2.2{\cdot}10^{-6}$	$-1.1{\cdot}10^{-6}$	$7.6{\cdot}10^{-6}$
	0.75	$4.0{\cdot}10^{-6}$	$2.4{\cdot}10^{-6}$	$-6.2{\cdot}10^{-7}$	$8.7{\cdot}10^{-6}$

Figure 20: Plotted estimates of splines and intercept for Other Race including Multi-Racial Males.

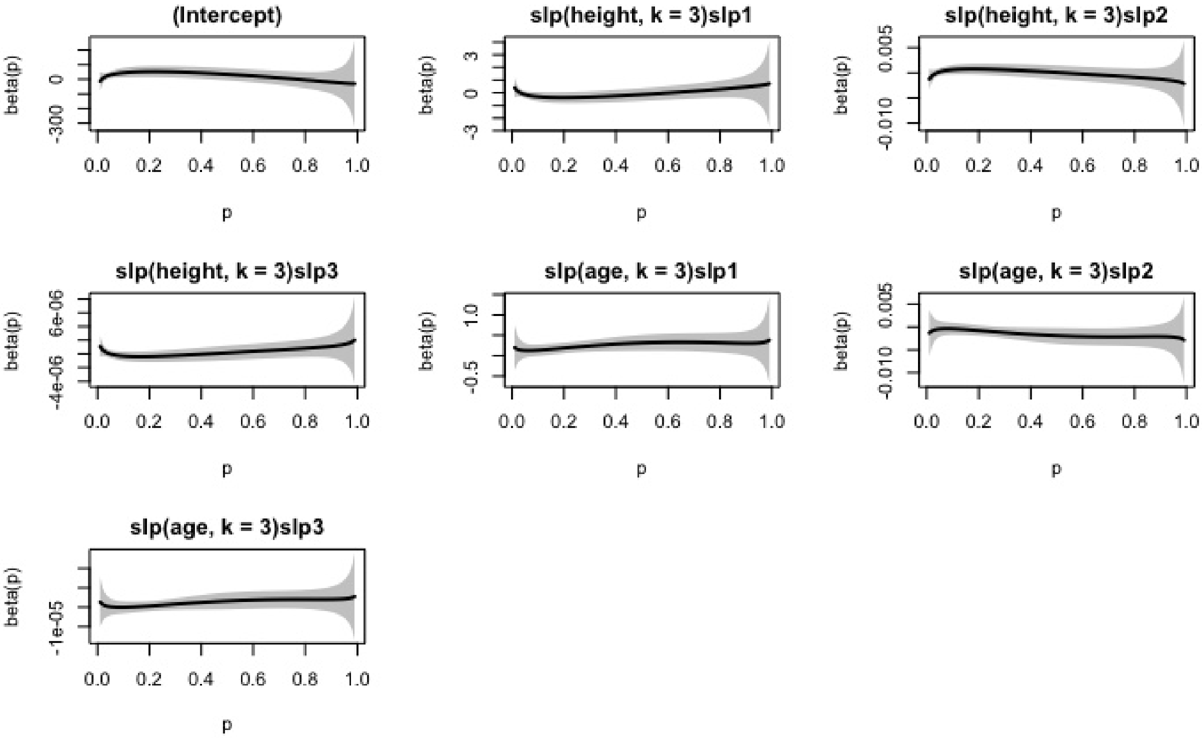


Table 29: QRCM estimates of the 25th, 50th and 75th quantiles for Mexican-American Females.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	60.5	58.5	-54.2	175.2
	0.50	80.0	31.0	19.3	140.7
	0.75	99.8	164.7	-223.0	422.5
Height Spline 1	0.25	-0.5	0.7	-1.8	0.8
	0.50	-0.7	0.4	-1.5	$3.6 \cdot 10^{-2}$
	0.75	-0.9	1.8	-4.5	2.7
Height Spline 2	0.25	$1.2 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$-2.2 \cdot 10^{-3}$	$4.5 \cdot 10^{-3}$
	0.50	$1.2 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$-0.4 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$
	0.75	$2.0 \cdot 10^{-3}$	$4.5 \cdot 10^{-3}$	$-6.8 \cdot 10^{-3}$	$1.1 \cdot 10^{-2}$
Height Spline 3	0.25	$-7.4 \cdot 10^{-7}$	$1.3 \cdot 10^{-6}$	$-3.2 \cdot 10^{-6}$	$1.7 \cdot 10^{-6}$
	0.50	$-1.1 \cdot 10^{-6}$	$8.6 \cdot 10^{-7}$	$-2.7 \cdot 10^{-6}$	$6.3\text{e} \cdot 10^{-7}$
	0.75	$-1.2 \cdot 10^{-6}$	$3.3 \cdot 10^{-6}$	$-7.7 \cdot 10^{-6}$	$5.2 \cdot 10^{-6}$
Age Spline 1	0.25	0.4	0.1	0.2	0.5
	0.50	0.4	0.1	0.2	0.6
	0.75	0.4	0.1	0.1	0.7
Age Spline 2	0.25	$-2.2 \cdot 10^{-3}$	$0.7 \cdot 10^{-3}$	$-3.5 \cdot 10^{-3}$	$-0.9 \cdot 10^{-3}$
	0.50	$-2.0 \cdot 10^{-3}$	$0.9 \cdot 10^{-3}$	$-3.4 \cdot 10^{-3}$	$-0.2 \cdot 10^{-3}$
	0.75	$-2.1 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$-4.5 \cdot 10^{-3}$	$0.4 \cdot 10^{-3}$
Age Spline 3	0.25	$3.8 \cdot 10^{-6}$	$1.6 \cdot 10^{-6}$	$7.6 \cdot 10^{-7}$	$6.9 \cdot 10^{-6}$
	0.50	$3.0 \cdot 10^{-6}$	$2.2 \cdot 10^{-6}$	$-1.3 \cdot 10^{-6}$	$7.4 \cdot 10^{-6}$
	0.75	$2.8 \cdot 10^{-6}$	$2.9 \cdot 10^{-6}$	$-2.8 \cdot 10^{-6}$	$8.5 \cdot 10^{-6}$

Figure 21: Plotted estimates of splines and intercept for Mexican-American Females.

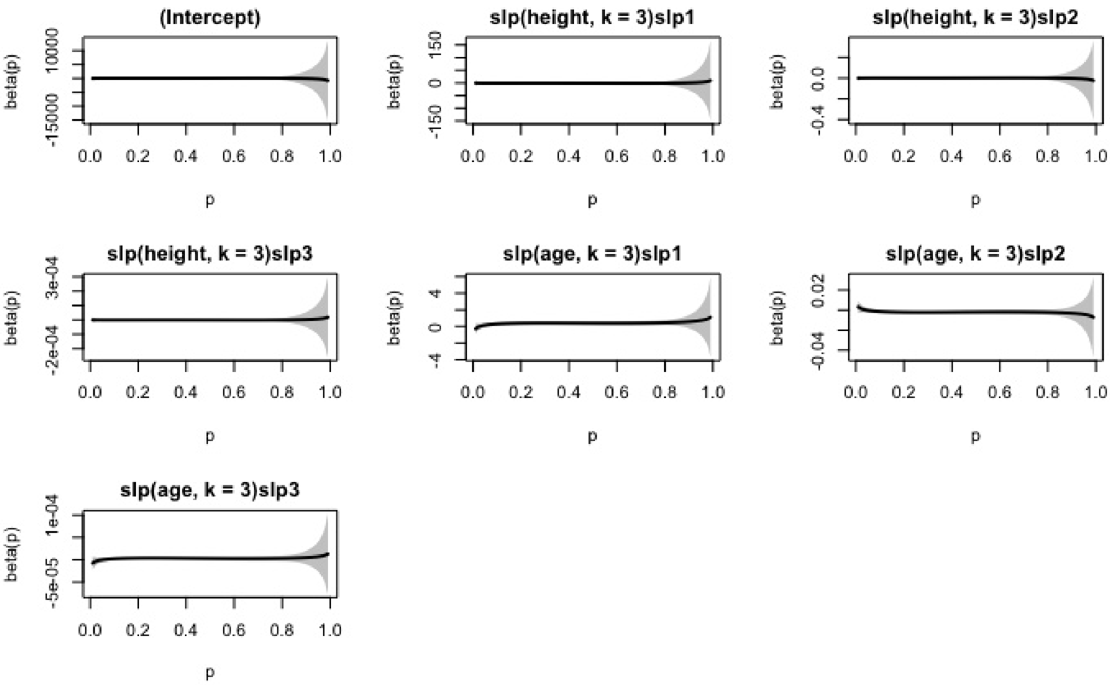


Table 30: QRCM estimates of the 25th, 50th and 75th quantiles for Other Hispanic Females.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	53.3	26.0	2.3	104.3
	0.50	29.2	28.6	-26.9	85.3
	0.75	-11.1	49.2	-107.4	85.3
Height Spline 1	0.25	-0.5	0.3	-1.1	0.2
	0.50	-0.2	0.4	-0.9	0.5
	0.75	0.3	0.6	-0.9	1.5
Height Spline 2	0.25	$1.1{\cdot}10^{-3}$	$1.0{\cdot}10^{-3}$	$-0.7{\cdot}10^{-3}$	$3.0{\cdot}10^{-3}$
	0.50	$0.4{\cdot}10^{-3}$	$1.0{\cdot}10^{-3}$	$-1.5{\cdot}10^{-3}$	$2.4{\cdot}10^{-3}$
	0.75	$-1.0{\cdot}10^{-3}$	$1.7{\cdot}10^{-3}$	$-4.3{\cdot}10^{-3}$	$2.4{\cdot}10^{-3}$
Height Spline 3	0.25	$-8.6{\cdot}10^{-7}$	$7.7{\cdot}10^{-7}$	$-2.4{\cdot}10^{-6}$	$6.6{\cdot}10^{-7}$
	0.50	$-3.1{\cdot}10^{-7}$	$8.0{\cdot}10^{-7}$	$-1.9{\cdot}10^{-6}$	$1.3{\cdot}10^{-6}$
	0.75	$7.9{\cdot}10^{-7}$	$1.4{\cdot}10^{-6}$	$-1.9{\cdot}10^{-6}$	$3.5{\cdot}10^{-6}$
Age Spline 1	0.25	3.5	0.1	$-6.1{\cdot}10^{-3}$	0.4
	0.50	0.3	0.1	0.1	0.5
	0.75	0.6	0.2	0.2	0.9
Age Spline 2	0.25	$-0.4{\cdot}10^{-3}$	$0.8{\cdot}10^{-3}$	$-2.0{\cdot}10^{-3}$	$1.2{\cdot}10^{-3}$
	0.50	$-0.7{\cdot}10^{-3}$	$1.1{\cdot}10^{-3}$	$-2.8{\cdot}10^{-3}$	$1.4{\cdot}10^{-3}$
	0.75	$-3.4{\cdot}10^{-3}$	$1.6{\cdot}10^{-3}$	$-6.4{\cdot}10^{-3}$	$-0.3{\cdot}10^{-3}$
Age Spline 3	0.25	$-3.8{\cdot}10^{-7}$	$1.9{\cdot}10^{-6}$	$-4.0{\cdot}10^{-6}$	$3.3{\cdot}10^{-6}$
	0.50	$-6.0{\cdot}10^{-7}$	$2.5{\cdot}10^{-6}$	$-5.4{\cdot}10^{-6}$	$4.2{\cdot}10^{-6}$
	0.75	$5.3{\cdot}10^{-6}$	$3.7{\cdot}10^{-6}$	$-2.0{\cdot}10^{-6}$	$1.3{\cdot}10^{-5}$

Figure 22: Plotted estimates of splines and intercept for Other Hispanic Females.

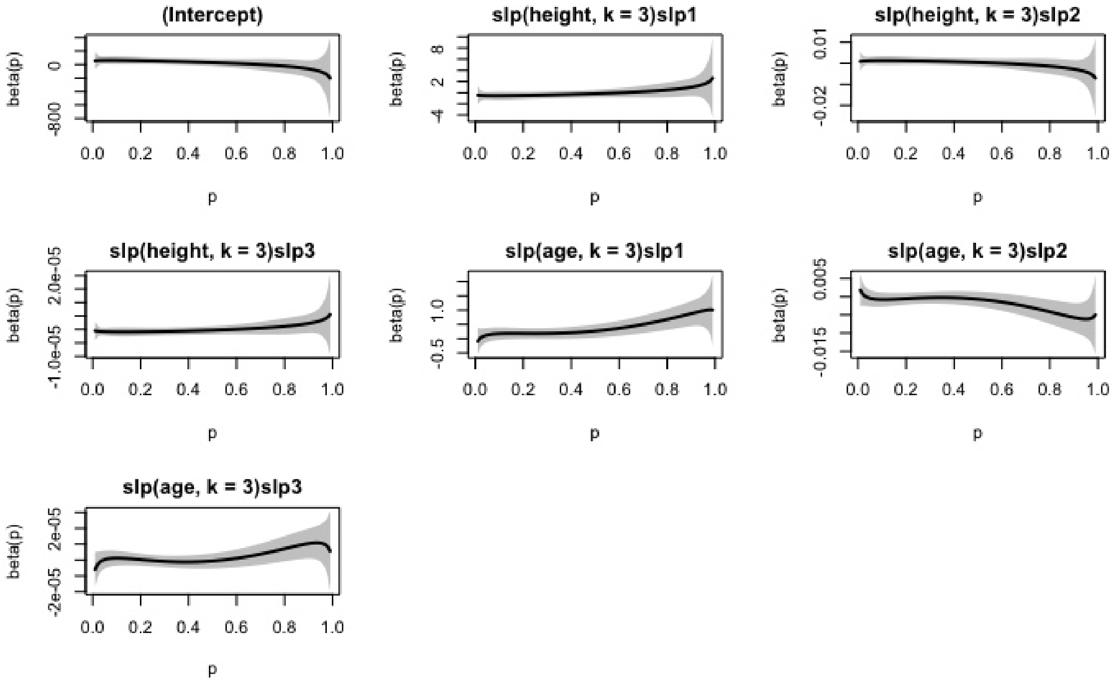


Table 31: QRCM estimates of the 25th, 50th and 75th quantiles for Non-Hispanic White Females.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	97.9	16.5	65.5	130.3
	0.50	108.9	18.6	72.5	145.2
	0.75	134.0	29.8	75.5	192.1
Height Spline 1	0.25	-0.9	0.2	-1.3	-0.5
	0.50	-1.1	0.2	-1.5	-0.6
	0.75	-1.4	0.4	-2.1	-0.7
Height Spline 2	0.25	$2.3 \cdot 10^{-3}$	$0.5 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$
	0.50	$2.6 \cdot 10^{-3}$	$0.6 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$	$3.8 \cdot 10^{-3}$
	0.75	$3.4 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$5.3 \cdot 10^{-3}$
Height Spline 3	0.25	$-1.6 \cdot 10^{-6}$	$4.3 \cdot 10^{-7}$	$-2.5 \cdot 10^{-6}$	$-7.9 \cdot 10^{-7}$
	0.50	$-1.8 \cdot 10^{-6}$	$4.9 \cdot 10^{-7}$	$-2.8 \cdot 10^{-6}$	$-8.9 \cdot 10^{-7}$
	0.75	$-2.5 \cdot 10^{-6}$	$7.7 \cdot 10^{-7}$	$-4.0 \cdot 10^{-6}$	$-9.7 \cdot 10^{-7}$
Age Spline 1	0.25	0.2	0.1	0.1	0.4
	0.50	0.4	0.1	0.3	0.5
	0.75	0.5	0.2	0.3	0.7
Age Spline 2	0.25	$-1.0 \cdot 10^{-3}$	$0.6 \cdot 10^{-3}$	$-2.2 \cdot 10^{-3}$	$8.4 \cdot 10^{-5}$
	0.50	$-2.0 \cdot 10^{-3}$	$0.7 \cdot 10^{-3}$	$-3.4 \cdot 10^{-3}$	$-5.6 \cdot 10^{-4}$
	0.75	$-2.6 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$-4.6 \cdot 10^{-3}$	$-5.4 \cdot 10^{-4}$
Age Spline 3	0.25	$1.2 \cdot 10^{-6}$	$1.4 \cdot 10^{-6}$	$-1.5 \cdot 10^{-6}$	$4.0 \cdot 10^{-6}$
	0.50	$2.9 \cdot 10^{-6}$	$1.7 \cdot 10^{-6}$	$-4.1 \cdot 10^{-7}$	$6.2 \cdot 10^{-6}$
	0.75	$3.4 \cdot 10^{-6}$	$2.4 \cdot 10^{-6}$	$-1.3 \cdot 10^{-6}$	$8.1 \cdot 10^{-6}$

Figure 23: Plotted estimates of splines and intercept for Non-Hispanic White Females.

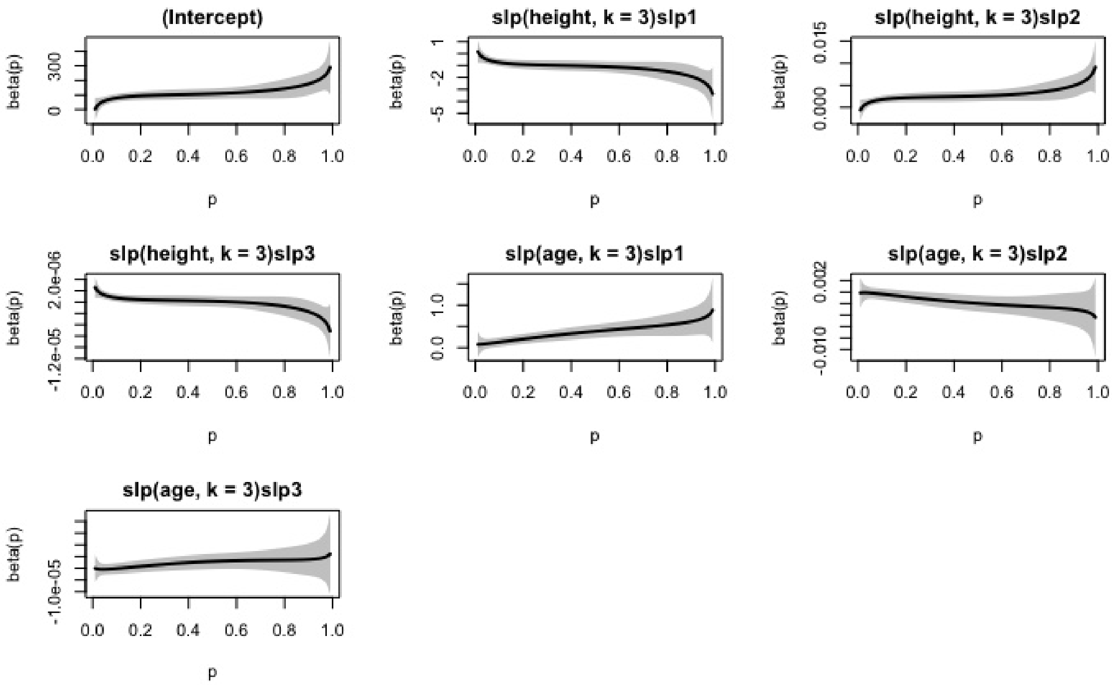


Table 32: QRCM estimates of the 25th, 50th and 75th quantiles for Non-Hispanic White Females.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	65.7	23.8	19.0	112.5
	0.50	66.3	23.6	20.0	112.6
	0.75	116.5	31.5	54.8	178.3
Height Spline 1	0.25	-0.5	0.3	-1.1	$2.8 \cdot 10^{-2}$
	0.50	-0.5	0.3	-1.1	$4.5 \cdot 10^{-2}$
	0.75	-1.1	0.4	-1.9	-0.4
Height Spline 2	0.25	$1.2 \cdot 10^{-3}$	$0.8 \cdot 10^{-3}$	$-0.3 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$
	0.50	$1.1 \cdot 10^{-3}$	$0.8 \cdot 10^{-3}$	$-0.5 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$
	0.75	$2.8 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$0.7 \cdot 10^{-3}$	$0.5 \cdot 10^{-3}$
Height Spline 3	0.25	$-7.4 \cdot 10^{-7}$	$6.3 \cdot 10^{-7}$	$-2.0 \cdot 10^{-6}$	$4.9 \cdot 10^{-7}$
	0.50	$-6.2 \cdot 10^{-7}$	$6.4 \cdot 10^{-7}$	$-1.9 \cdot 10^{-6}$	$6.8 \cdot 10^{-7}$
	0.75	$-1.9 \cdot 10^{-6}$	$8.5 \cdot 10^{-7}$	$-3.5 \cdot 10^{-6}$	$-2.3 \cdot 10^{-7}$
Age Spline 1	0.25	0.2	0.1	0.1	0.4
	0.50	0.3	0.1	0.1	0.5
	0.75	0.5	0.1	0.3	0.8
Age Spline 2	0.25	$-0.9 \cdot 10^{-3}$	$0.8 \cdot 10^{-3}$	$-2.3 \cdot 10^{-3}$	$0.6 \cdot 10^{-3}$
	0.50	$-0.1 \cdot 10^{-3}$	$0.9 \cdot 10^{-3}$	$-3.0 \cdot 10^{-3}$	$0.6 \cdot 10^{-3}$
	0.75	$-2.8 \cdot 10^{-3}$	$0.1 \cdot 10^{-3}$	$-0.5 \cdot 10^{-3}$	$-0.5 \cdot 10^{-3}$
Age Spline 3	0.25	$1.2 \cdot 10^{-6}$	$1.9 \cdot 10^{-6}$	$-2.5 \cdot 10^{-6}$	$4.8 \cdot 10^{-6}$
	0.50	$1.1 \cdot 10^{-6}$	$2.0 \cdot 10^{-6}$	$-2.9 \cdot 10^{-6}$	$5.1 \cdot 10^{-6}$
	0.75	$3.9 \cdot 10^{-6}$	$2.9 \cdot 10^{-6}$	$-1.8 \cdot 10^{-6}$	$9.6 \cdot 10^{-6}$

Figure 24: Plotted estimates of splines and intercept for Non-Hispanic Black Females.

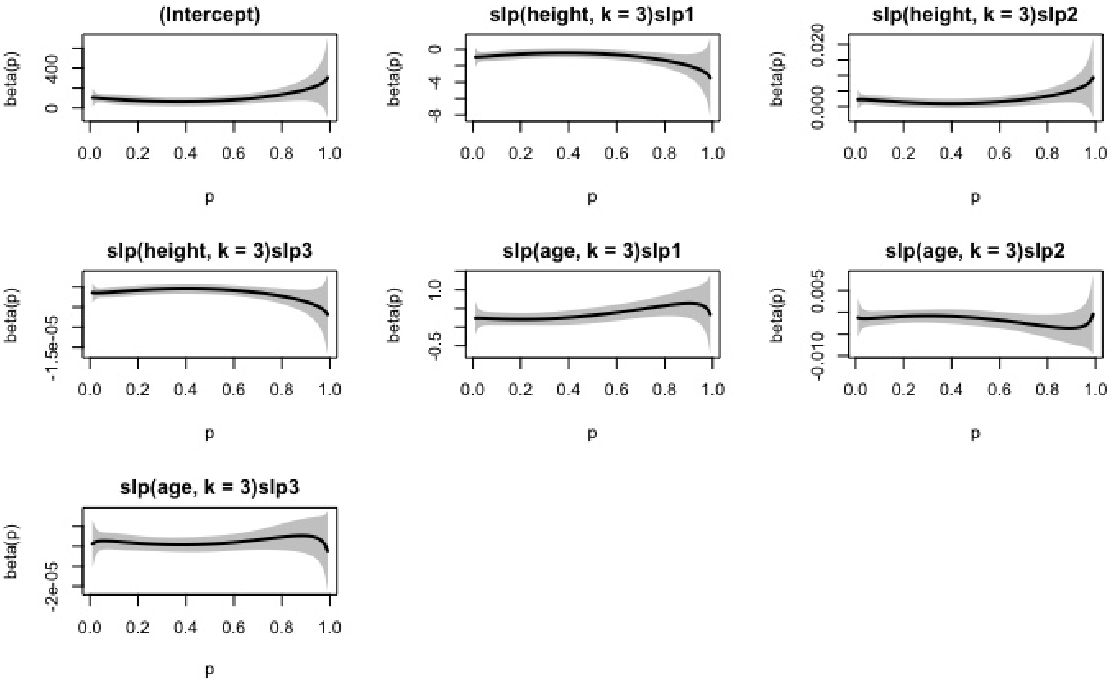
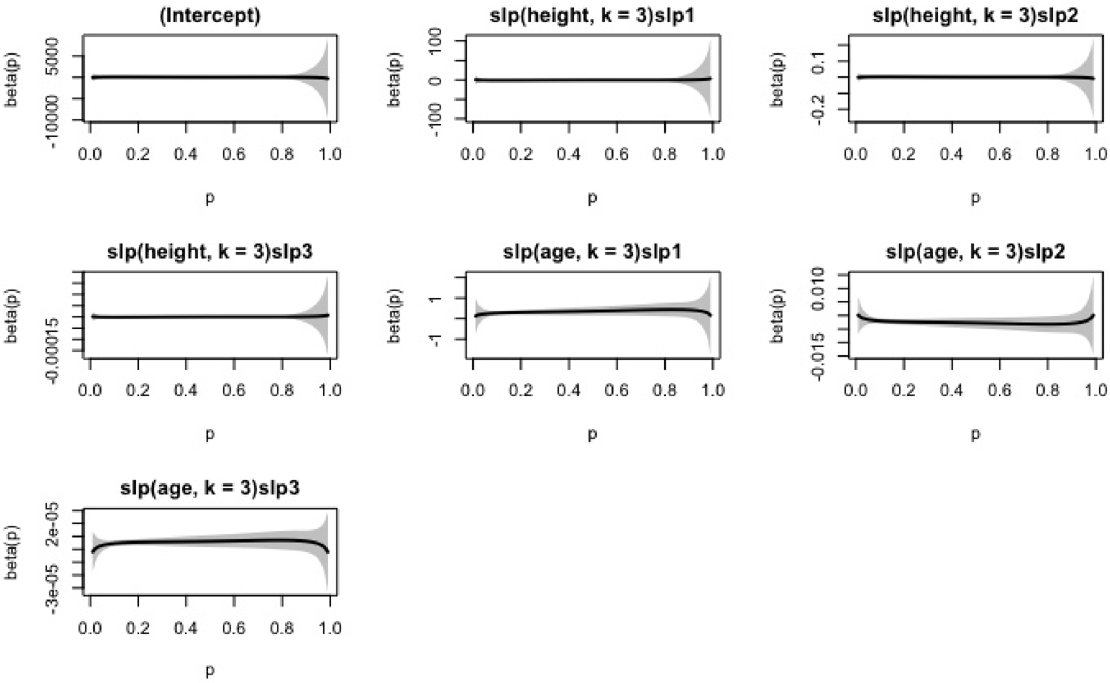


Table 33: QRCM estimates of the 25th, 50th and 75th quantiles for Other Race including Multi-racial Females.

Estimates	Quantiles	Beta, β	Standard Error	Lower confidence level	Upper confidence level
Intercept	0.25	70.0	24.1	22.8	117.3
	0.50	42.4	42.9	-41.7	126.5
	0.75	47.5	42.3	-35.4	130.5
Height Spline 1	0.25	-0.6	0.3	-1.2	-0.1
	0.50	-0.3	0.5	-1.3	0.8
	0.75	-0.3	0.5	-1.3	0.8
Height Spline 2	0.25	$0.1{\cdot}10^{-3}$	0.8	$-8.2{\cdot}10^{-5}$	$2.9{\cdot}10^{-3}$
	0.50	$0.4{\cdot}10^{-3}$	$1.4{\cdot}10^{-3}$	$-2.4{\cdot}10^{-3}$	$3.2{\cdot}10^{-3}$
	0.75	$3.7{\cdot}10^{-3}$	$1.5{\cdot}10^{-3}$	$-2.5{\cdot}10^{-3}$	$3.2{\cdot}10^{-3}$
Height Spline 3	0.25	$-9.2{\cdot}10^{-7}$	$5.9{\cdot}10^{-7}$	$-2.1{\cdot}10^{-6}$	$2.3{\cdot}10^{-7}$
	0.50	$-7.1{\cdot}10^{-8}$	$1.1{\cdot}10^{-6}$	$-2.3{\cdot}10^{-6}$	$2.2{\cdot}10^{-6}$
	0.75	$2.1{\cdot}10^{-8}$	$1.2{\cdot}10^{-6}$	$-2.2{\cdot}10^{-6}$	$2.3{\cdot}10^{-6}$
Age Spline 1	0.25	0.3	0.1	0.2	0.4
	0.50	0.4	0.1	0.2	0.6
	0.75	0.4	0.2	0.1	0.7
Age Spline 2	0.25	$-2.4{\cdot}10^{-3}$	$0.6{\cdot}10^{-3}$	$-3.7{\cdot}10^{-3}$	$-1.2{\cdot}10^{-3}$
	0.50	$-2.8{\cdot}10^{-3}$	$1.0{\cdot}10^{-3}$	$-4.7{\cdot}10^{-3}$	$-0.83{\cdot}10^{-3}$
	0.75	$-3.2{\cdot}10^{-3}$	$1.4{\cdot}10^{-3}$	$-6.0{\cdot}10^{-3}$	$-0.4{\cdot}10^{-3}$
Age Spline 3	0.25	$5.7{\cdot}10^{-6}$	$1.5{\cdot}10^{-6}$	$2.8{\cdot}10^{-6}$	$8.5{\cdot}10^{-6}$
	0.50	$6.2{\cdot}10^{-6}$	$2.4{\cdot}10^{-6}$	$1.5{\cdot}10^{-6}$	$1.1{\cdot}10^{-5}$
	0.75	$7.1{\cdot}10^{-6}$	$3.5{\cdot}10^{-6}$	$2.9{\cdot}10^{-7}$	$1.4{\cdot}10^{-5}$

Figure 25: Plotted estimates of splines and intercept for Other Racial including Multi-Racial Females.



A3 R Code

The following R Code has been used when performing the data analysis.

~~~
“‘{r}
rm(list=ls())
library(qrcm)
library(splines)

# load the data
data <- read.table("Data.txt", fill = TRUE, header = TRUE)

# check the data
head(data)

# exclusion criteria
data <- data[!is.na(data$bmi)&data$age<80&data$gender==2,]

# fit the qrcm with an asymmetric logistic intercept and legandre splines
for the coefficients of age and height
s <- matrix(1,6,5)
s[1,4:5] <- 0
s[2:6,2:3] <- 0
qrcm1 <- iqr(bmi ~ slp(height,k=3) + slp(age,k=3), formula.p = ~ log(p)+log(1-
p)+slp(p,k=2), data=data)

#estimated parameters and standard errors
summary(qrcm1)

#p-values for testing that each parameter is equal to zero
round(1 - pchisq((summary(qrcm1)$coef/summary(qrcm1)$se)^2, 1), 3)

#plot of all the regression coefficients
par(mfrow = c(3,3))
plot(qrcm1, ask=F)

#estimate regression coefficients at specified quantiles
predict(qrcm1, type = "beta", p = c(0.25,0.5,0.75))
#plot the estimated quantiles of BMI for two individuals
#plot(data$age, data$height)
names <- c("Age 5", "Age 31", "Age 75")
par(mfrow = c(1,1))
matplot(1:999/1000,
cbind(
c(predict(qrcm1, type = "QF", p = 1:999/1000, newdata = data.frame(height
~~~

= 114, age = 5))$fit),
c(predict(qrcm1, type = "QF", p = 1:999/1000, newdata = data.frame(height = 159, age = 31))$fit),
c(predict(qrcm1, type = "QF", p = 1:999/1000, newdata = data.frame(height = 152, age = 75))$fit)),
type="l", ylab="Estimated quantile function", xlab="Proportion")

legend("topleft", inset=0.01, legend=names, col=c(1:5), lty=1:3, bg= ("white"), horiz=F)

#estimate the 0.95, 0.99, and 0.999 quantile of BMI for any subject with given age and height
predict(qrcm1, type = "QF", p =c(.10,.50, .95), newdata = data.frame(height = 161, age = 31))