Development and Application of Children's Sex- and Age-Specific Fat-Mass and Muscle-Mass Reference Curves using the LMS Methodology

Stephanie Tanasia Saputra

A Thesis

In the Department of

Mathematics and Statistics

Presented in Partial Fulfilment of the Requirements

For the Degree of

Master of Science (Mathematics and Statistics)

at Concordia University

Montréal, Québec, Canada

April 2023

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Stephanie Tanasia Saputra

Entitled: Development and Application of Children's Sex- and Age-Specific Fat-Mass and Muscle-Mass Reference Curves using the LMS Methodology

and submitted in partial fulfilment of the requirements for the degree of

Master of Science (Mathematics and Statistics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____Chair
Dr. Simone Brugiapaglia

_____Examiner
Dr. Andraea van Hulst

_____Examiner
Dr. Simone Brugiapaglia

_____Examiner
Dr. Yogen Chaubey

_____Thesis Supervisor
Dr. Lisa Kakinami

Approved by _____
Dr. Yogen Chaubey, Graduate Program Director

_____
Dr. Pascale Sicotte, Dean of Faculty

Abstract for Masters

Development and Application of Children's Sex- and Age-Specific Fat-Mass and Muscle-Mass Reference Curves using the LMS Methodology

Stephanie Tanasia Saputra

Body mass index cannot distinguish between fat-mass and muscle-mass, which may result in obesity misclassification. A dual-energy x-ray absorptiometry (DXA)-derived phenotype classification based on fat-mass and muscle-mass has been proposed for adults (≥18 yo). We extend this research by developing children's fat-mass and muscle-mass reference curves and determining their utility in identifying cardiometabolic risk. Children's (≤17 yo) DXA data in NHANES, a US national health survey (n=6,120) were used to generate sex- and age-specific deciles of appendicular skeletal muscle index and fat mass index (kg/m^2) with the Lambda Mu Sigma (LMS) method. The final curves were selected through goodness of fit (AIC, Q-tests, detrended Q-Q plot). Four phenotypes (high [H] or low [L], adiposity [A] and muscle mass [M]: HA-HM, HA-LM, LA-HM, LA-LM) were identified using the literature's guidelines above/below the median compared to same-sex and same-age peers. The curves and their corresponding phenotypes were applied to QUALITY data, a longitudinal cohort (n=630, 8-10 yo in 2005) to assess whether the phenotypes correctly identified cardiometabolic risk using multiple linear regression at baseline, follow-up one (2008-2010), or follow-up two (2015-2017). Models were adjusted for age, sex, and Tanner's stage. Chained equation was used to impute missing values in QUALITY. Compared to LA-HM, LA-LM was associated with lower glucose at baseline; HA-HM was associated with lower HDL-c and higher LDL-c, triglycerides, and HOMA-IR; HA-LM was associated with elevated triglycerides and HOMA-IR at all timepoints (all $p<0.05$). These phenotypes allowed for discrimination of cross-sectional cardiometabolic risks, but further longitudinal exploration is recommended.

Acknowledgements

This completion of thesis would not be possible without my supervisor, Dr. Lisa Kakinami. Throughout my master's degree journey, she has given me countless insightful guidance and wisdom. Her systematic approach to research has inspired me to strive for knowledge. I am grateful for the time and effort that she dedicated to my development as a scholar; from team meetings to abstract reviews- I am fortunate to have had the opportunity to learn from her.

My sincere gratitude also goes to Dr. Simone Brugiapaglia and Dr. Andraea van Hulst for their invaluable guidance and support throughout my thesis. Their expertise, insights, and support have been instrumental in shaping my research.

I would also like to thank my family; especially Mama for her encouragement and Cece who's always been there for me, either as a fellow Statistician or a big sister when I need one.

Lastly, I would like to thank my partner, Jeremy Schlitt, whose unwavering support has accompanied me through late-night thesis work. His motivation for research has made me strive for the better.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike Information Criterion
ASMI	Appendicular Skeletal Muscle Index
BMI	Body Mass Index
CV	Coefficient of Variation
DXA	Dual Energy X-ray Absorptiometry
EDF	Equivalent Degrees of Freedom
FMI	Fat Mass Index
HA-HM	High Adiposity High Muscle
HA-LM	High Adiposity Low Muscle
HDL	High Density Lipoprotein
HOMA-IR	Homeostatic Model Assessment of Insulin Resistance
LA-HM	Low Adiposity High Muscle
LA-LM	Low Adiposity Low Muscle
LDL	Low Density Lipoprotein
LMS	Lamba Mu Sigma
MICE	Multiple Imputation by Chained Equation
MLR	Multiple Linear Regression
TC	Total Cholesterol
QUALITY	**Qu**ebec **A**dipose and **L**ifestyle **I**nvestigation in **Y**outh
VF	Visceral Fat
V1	Visit 1/Baseline at QUALITY
V2	Visit 2/First follow-up in QUALITY
V3	Visit 3/Second follow-up in QUALITY
WHO	World Health Organization

1 Introduction

Worldwide, obesity has nearly tripled since 1975. According to the World Health Organization, over 340 million children and adolescents aged 5-19 have overweight or obesity in 2016 [1]. Childhood obesity is associated with a higher chance of severe diseases in adulthood such as stroke, heart disease, and diabetes [2]. Although obesity is preventable and there have been a lot of research and public programs to reduce its increase, the prevalence of obesity and overweightness has risen dramatically from 4% to 18% among children aged 5-19 in the past 40 years [1]. Childhood obesity remains a public health concern in Canada, with an estimated 30% of children having overweight or obesity [2]. One of the emerging types is sarcopenic obesity, often characterized by excessive fat mass in the presence of reduced muscle mass [3]. There is evidence that people with sarcopenic obesity have higher rates of illness, disability, and mortality compared to people with either low muscle mass or obesity [4].

It is pivotal to have obesity measures due to the risk that it imposes. Body mass index (BMI) is commonly used as a proxy for adiposity [5]. However, it does not distinguish between fat and muscle mass as it is calculated by dividing weight by height squared. This can lead to obesity misclassification due to over/underestimation of fat mass. To better assess this, more precise methods such as dual-energy x-ray absorptiometry (DXA) need to be used. DXA can calculate the fat, muscle, and bone mass in a body. Thus, recent studies have been using DXA as an adiposity measure since they allow distinctions between the fat and muscle mass. For instance, in 2019, a study was conducted to compare the ability of DXA to other anthropometric measures in predicting visceral fat (VF) of 9 to 13 years old girls. The authors concluded that adding DXA to their model improved the predictability of VF [6].

Figure 1: Dual-energy x-ray absorptiometry (DXA) machine [7]

Furthermore, DXA results can be used to calculate derivative values such as fat mass index (FMI: fat mass/height^2) [8] and appendicular skeletal muscle index (ASMI: appendicular lean mass/height^2) [9]. These indices are commonly used as they account for height. Baumgartner used

the percentiles of ASMI and FMI to create four mutually exclusive phenotypes to better explain body composition [10]. He created cut-offs for low (<50th percentile) or high (≥50th percentile) adiposity and muscle mass and four mutually-exclusive phenotypes were: low adiposity with high muscle mass (LA-HM), high adiposity with high muscle mass (HA-HM), low adiposity with low muscle mass (LA-LM), and HA-LM. Using these phenotypes, HA-LM was associated with a drastic reduction of insulin sensitivity [11]. In 2014, Prado et al., extended this literature to derive sex- and age-adjusted FMI and ASMI reference curves by incorporating the Lambda Mu Sigma (LMS) method while also considering Baumgartner's phenotypes in the process [10], [12], [13].

Reference curves are known to be a quick and reliable tool to assess a growth/development of a measure along with age. These are particularly crucial for assessing development in young populations, as they are still growing. It is used by doctors, paediatricians, and parents to quickly assess children's development according to the reference population of same-age, same-sex peers [14]. Reference curves are presented as several centiles to see where the measure of interest falls between them. Figure 2 shows the Centers for Disease Control and Prevention BMI reference curves for girls aged 2-20 years old [15]. For instance, if a 12-year-old girl has a BMI of 20, from the reference curves, we see that her BMI falls between the 95th and 98th percentile of the reference population. Thus, nurses/paediatricians could accurately observe that her BMI is high just by observing where it falls in the reference curves; her BMI is greater than 95% of the reference population's BMI from her same-age, same-sex peers. Then, along with other health diagnostics, they can prescribe a proper dietary measure for the patients.

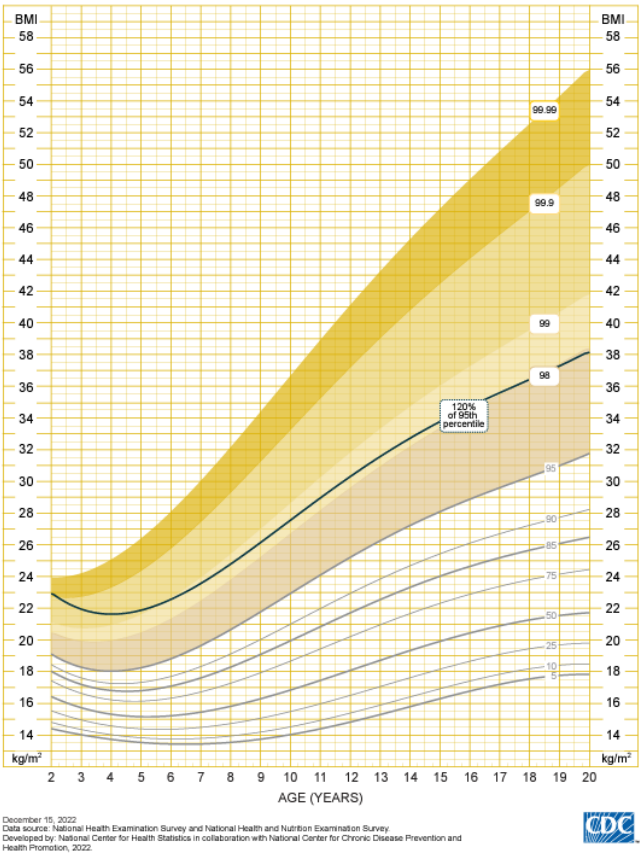


Figure 2: BMI reference curves from CDC, for girls aged 2-20 years old (2022)

The LMS method is commonly used to construct these reference curves [13]. The LMS method presents a way of obtaining normalized growth centiles after accounting for the skewness that is often present in the distribution of data [16]. First, the median (μ) of the variable of interest is calculated as well as the coefficient of variation, σ. Next, the Box-Cox power transformation is applied to the data, under the parameter λ. After the transformation, z-score and centiles can be calculated through the standard normal distribution assumption which will then be plotted into reference curves.

The LMS method was introduced by Cole in 1989 [16] and has been widely applied in practice to develop reference curves among children and youth. For example, the World Health Organization (WHO) uses this method to develop their 2007 centiles and z-scores for height, weight, and BMI [17]. Moreover, in 2001, Lun et al. developed a fetal reference chart and they found the LMS method to have a better result compared to the regular "mean and SD method" [18].

In their study, Prado et al. [12] developed adiposity and muscle-mass reference curves and their corresponding phenotypes using NHANES, a nationally representative adult population (> 17 years). Although DXA is not collected in routine clinical care, improving our understanding of the methodological and substantive aspects of DXA phenotypes is a necessary preliminary step in order to identify appropriate proxies in the future. As a result, developing and improving DXA phenotypes is one viable way to increase the usability of the data. However, for youth populations, there are currently no adiposity and muscle mass reference curves and phenotypes analogous to Prado's work. In addition, empirical research applying the LMS-derived phenotypes of fat- and muscle-mass are lacking.

Our study attempts to address the aforementioned limitations. Thus, the objective of this study was to:

1. Develop reference curves and body composition phenotypes for children based on their FMI and ASMI data using the LMS method.
2. Determine whether our DXA-derived muscle- and fat-mass phenotypes in children are associated with cross-sectional cardiometabolic risks.

2 Methodology

2.1 Study sample

Prado et al. utilized the 1999-2004 data from the National Health and Nutrition Examination Survey (NHANES) to develop the reference curves in adults [12]. NHANES is a program of studies intended to evaluate the health and nutritional status of adults and children in the United States. Cross-sectional samples were collected every two years since 1999 [19].

Individuals were invited to participate in NHANES using a complex multistage probability sampling technique, which included oversampling of African Americans, Mexican Americans, low-income whites, and older people (≥60 years) in order to provide a representative sample of the general US population [19]. All participants provided informed consent and ethics approval was obtained from the National Center for Health Statistics Ethics Review Board [20].

To develop the reference curves in youth for this study, the cross-sectional data from 1999-2006 NHANES were used ($n = 41{,}474$). Respondents were excluded if they were younger than eight years of age (the minimum age required for the dual-energy x-ray absorptiometry [DXA] scans in NHANES; $n = 8{,}655$), older than 18 years ($n = 22{,}624$), or if they had no DXA data due to height/weight restrictions (weight limit of 136 kg, height limit of 196 cm) or pregnancy ($n = 4{,}075$) [21]. The final analytical sample was 6,120 (3,708 male and 2,412 female). Missing DXA data were imputed using multiple-imputation methodology into five data sets [21]. This was incorporated into the data analysis, as described in a later section.

2.2 Anthropometric and body composition

All body measures were obtained by trained health technicians. A stadiometer was used to measure height, and respondents were asked to take a deep inhalation. Weight was measured by an electronic digital scale calibrated in kilograms.

Whole-body DXA scans were taken with a Hologic QDR-4500A fan-beam densitometer (Hologic, Inc., Bedford, Massachusetts) with a software version of 8.26 [21]. The DXA scans provide bone and soft tissue measurements for the total body, both arms, and both legs. A high level of quality control was maintained throughout the DXA data collection and scan analysis, including a rigorous phantom scanning schedule. A detailed description of data acquisition and precision can be found in the documentation [22], [23].

Appendicular skeletal mass index (ASMI, representing muscle mass) was calculated using the DXA lean mass for arms and legs, divided by height squared. Fat mass index (FMI, representing fat mass) was calculated by dividing total fat mass with height squared. These values then informed the development of the reference curves, as described in the subsequent section.

2.3 LMS method

It is a common practice to express child growth in the form of reference centile curves as it shows the distribution of a measurement as it changes accordingly to some covariate, often age [24]. In this study, a curve fitting procedure called the Lambda Mu Sigma (LMS) method was used to

generate the reference curves for children under the age of 18 years. The lmsChartMaker Pro version 2.54 was used for this procedure [25].

2.3.1 Calculating lambda, mu, sigma

The FMI and ASMI data are first divided into distinct age groups. Ideally there should be a one-year gap or less between each group. In this study, the age group was from 8 years to 17 years, with a one-year gap. Thus, we have 10 age groups in total.

The variable of interest, (in this case FMI and ASMI) is denoted by $y \geq 0$. Suppose that y has a median of μ_y and that y^λ is normally distributed. Based on the technique proposed by Box and Cox [26], it is then defined that

$$x = \begin{cases} \dfrac{(y/\mu)^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln\left(\dfrac{y}{\mu}\right), & \lambda = 0. \end{cases} \tag{1}$$

This transformation ensures the mapping of μ_y to $x = 0$ and is continuous at $\lambda = 0$. Since λ is a free parameter, the value that minimizes the standard deviation of x is the optimal solution for it. A power transformation is applied to the age-stratified FMI and ASMI data to eliminate the skewness of the dependent variable. This is because in general, anthropometry data tends to be skewed to the right [13]. After x is calculated, the z-score for which the curves will depend on can be calculated through the standard normal distribution assumption (mean of the transformed variable is assumed to be zero). The z-score for x is defined as

$$z = \frac{x}{\sigma} = \begin{cases} \dfrac{(y/\mu)^\lambda - 1}{\lambda\sigma}, & \lambda \neq 0 \\ \dfrac{\ln\left(\dfrac{y}{\mu}\right)}{\sigma}, & \lambda = 0, \end{cases} \tag{2}$$

where σ is the standard deviation of the transformed variable x (or the CV of y).

The method is called LMS since the parameters are: lambda (λ, for the Box-Cox power transformation), median (μ, the median of the variable), and coefficient of variation (σ). With the assumption that the residuals follow a normal distribution, the desired percentile can be calculated. L, M, and S are the estimation of λ, μ, and σ respectively. Cole et al. derived a step by step account of how to calculate L, M, and S for each of the age groups [13]:

1) The mean and standard deviation (SD) of the natural logarithms of the measurements (FMI or ASMI) is calculated. Let the antilog of the mean be called the ***geometric*** mean, denoted by M_g. Calculate the coefficient of variation (CV): S_g by dividing SD with M_g.
2) Repeat step (1) for the original measurements; the mean is the ***arithmetic*** mean of the measurements, denoted by M_a. The arithmetic CV; S_a is calculated by dividing SD with M_a.

3) Repeat step (1) for the reciprocal measurements; the mean is the ***harmonic*** mean of the measurements, denoted by M_h. The arithmetic CV; S_h is calculated by multiplying SD with M_g.
4) S_g, S_h, S_a should be very similar. Let

$$A = \log\left(\frac{S_a}{S_h}\right), B = \log\left(\frac{S_a S_h}{S_g^2}\right) \tag{3}$$

The estimate of the Box-Cox power L is then given by

$$L = -\frac{A}{2B}\ . \tag{4}$$

5) The generalized coefficient of variation *S* at the value of *L* is obtained by interpolating the three coefficients of variations to find the minimum value thus making sure that *S* is slightly smaller than the three. It is formulated as:

$$S = S_g \exp\left(\frac{AL}{4}\right). \tag{5}$$

6) Similarly, the generalized mean *M* is obtained by interpolating the three means:

$$M = M_g + \frac{(M_a - M_h)}{2} L + \frac{(M_a - 2M_g + M_h)}{2} L^2\ . \tag{6}$$

For generalization, assume now that the distribution of y varies with covariate *t* (in this study, *t* is age) and that $\mu, \lambda,$ and σ at *t* are M(*t*), L(*t*), and S(*t*). Then, we can rewrite Equation (2) as:

$$z = \begin{cases} \dfrac{(y/M(t))^{L(t)} - 1}{L(t)\, S(t)}\ , & L(t) \neq 0 \\ \dfrac{\ln\left(\frac{y}{L(t)}\right)}{S(t)}, & L(t) = 0. \end{cases} \tag{7}$$

With a simple rearranging, the centile 100α of y at t is

$$C_{100\alpha}(t) = \begin{cases} M(t)\,[1 + L(t)\, S(t)\, Z_\alpha]^{\frac{1}{L(t)}}\ , & L(t) \neq 0 \\ M(\mathrm{t}) \exp[S(t)\, Z_\alpha]\ , & L(t) = 0, \end{cases} \tag{8}$$

where Z_α is the z-score value under normal distribution with a significance level of α. After calculating all of the desired percentiles, the curves can be plotted. As an example, Figure 3 shows a centile curve of a measure where the bottom curve is the 10th percentile, then the 20th percentile, and so on until the top curve which represents the 90th percentile.

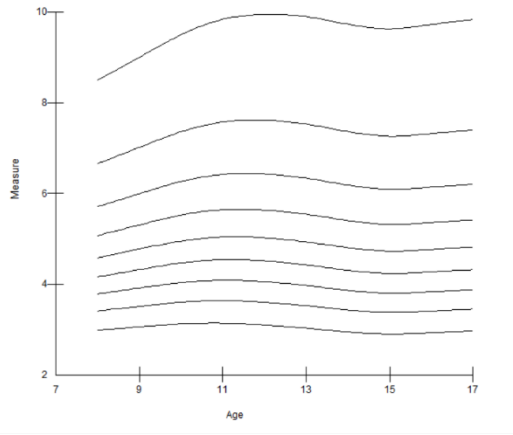


Figure 3: Centile curves of a measure where smoothness is maximized

2.3.2 Fitting and smoothing the curves

Based on the formulations in the previous section, the values of L, M, and S are obtained for each age group: L(*t*), M(*t*), and S(*t*) and plotted separately against their group mean ages. To achieve a set of smooth curves, the curves can be fitted by several methods, for example with a linear equation or even a cubic spline.

With the same software to calculate the values for L, M, and S, this study used lmsChartMaker Pro [25] to fit the curve using maximum penalized likelihood. Let the observations at t_i be y_i where $i = 1,2, \dots, n$. The log-likelihood function l can be derived from Equation (7):

$$l = \sum_{i=1}^{n} L(t_i) \log \frac{y_i}{M(t_i)} - \log S(t_i) - \frac{{z_i}^2}{2}, \tag{9}$$

where z_i are the standard deviation scores corresponding to y_i. Then, the curves are estimated by maximizing the penalized likelihood through the following objective function [24]:

$$O(t) = l - \frac{1}{2}\alpha_\lambda \int \{L''(t)\}^2 dt - \frac{1}{2}\alpha_\mu \int \{M''(t)\}^2 dt - \frac{1}{2}\alpha_\sigma \int \{S''(t)\}^2 dt, \tag{10}$$

where $\alpha_\lambda, \alpha_\mu$, and α_σ are smoothing parameters. Equation (10) has three integrals for each of the values and they serve as penalties according to the squared derivatives of L, M, and S curves. It can be shown that these penalties lead to natural cubic splines with knots at each distinct value of *t*.

Since the objective is to maximize the function, there is a clear tradeoff between the likelihood and the smoothness. Increasing the smoothness (by the $\alpha_\lambda, \alpha_\mu$, and α_σ) yields higher values for the integral side of the function (Figure 3), which lowers the objective. In contrast, compared to Figure 1, lowering the smoothness yields a higher value for the objective function but the curves might be under-smoothed (Figure 4).

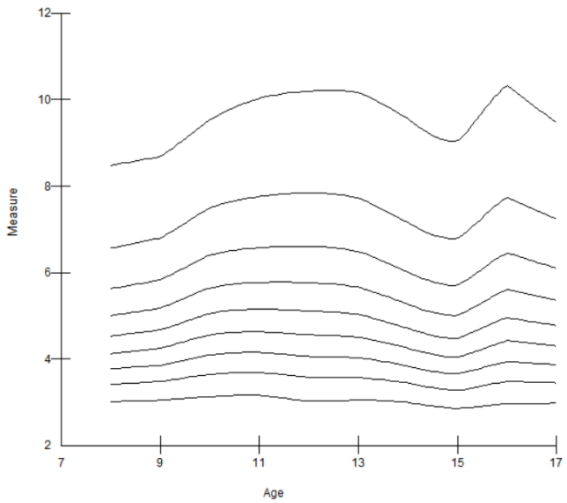


Figure 4: Centile curves of a measure where the objective function is maximized

The complexity of the curves is not measured directly by the smoothing parameters, but with its equivalent degrees of freedom (edf). For a parameter p, the edf is defined as:

$$edf(p) = trace\left(W_p + \alpha_p K\right)^{-1} W_p \,, \tag{11}$$

where W_p is the second derivative of l with respect to p and K is an $m \times m$ matrix. A detailed explanation about the smoothing algorithms and proofs of the equations are available in Cole and Green's work [24].

Thus, changing the edf changes the smoothness of the curves as well as its deviance. The lower the edf, the smoother the curves will be. For instance, Figure 3 has edf of 3,5,3, and is significantly smoother than Figure 4 with an edf of 10,10,10 of the same data. However, the deviance of 10,10,10's curves (deviance = 16,816.1) is lower than the deviance of 3,5,3 edf (deviance = 16,839.3) by 23.2 even though the curves are jagged. One other important thing to note is over-smoothing. Figure 3 shows an over-smoothed curve since we're missing the underlying curve shape: the dip around age 15 in the curves that are visible in Figure 4 have been lost. Therefore, goodness-of-fit tests are needed to help optimize the tradeoff.

2.3.3 Goodness of fit tests

In order to get the best set of curves, comparisons of goodness-of-fit tests are necessary. The simplest ones are the deviance measures: Penalized Deviance, Schwarz Bayesian Criterion, Generalized Akaike Information Criterion (GAIC), Akaike Information Criterion (AIC), and Global Deviance (D_g). In this study, Penalized Deviance (D_p) and AIC were selected as they are commonly used in the pertinent literature [7], [8]. The penalized deviance is defined as:

$$D_p = -2 \times O(t)$$

and AIC is defined as:

$$AIC = -2 \times D_g + 2 \times \left(edf_\lambda + edf_\mu + edf_\sigma\right).$$

For penalized deviance, the lower the value the better the fit since we would like *O(t)* in the function to be maximized. Similar to penalized deviance, the model's AIC should be low since it penalizes the model that uses more parameters; the simpler the model, the lower the AIC. Therefore, observing both parameters will give us the parsimonious balance between less splines in the curves and a closer fit to the actual data.

In addition to these quantitative measures, there are some qualitative tools available to check the goodness of fit. In this study, we looked at two of them: detrended Q-Q plot and Q-tests. For both tests, the data is converted into its z-scores. Detrended Q-Q plot is a detrended form of a Q-Q plot which is often used to assess whether data follow a particular distribution. First, the z-scores (observed) are sorted into increasing order and then each is plotted against its expected z-scores based on its centile assuming a normal distribution. The points appear as a worm shape where the extreme points are more variable than those in the center. This plot assesses the age-conditioned normality of the transformed data under the LMS method.

van Buuren and Fredriks found that the LMS method generally models the age-conditional mean and skewness better than the age-related deviation and kurtosis [27]. The offset, slope, and curvature of the worm plot are directly related to the fit of the LMS parameters. Cole, van Buuren, and Fredriks noted that there are various possible patterns in the detrended Q-Q plot [27], [28]:

- The slope indicates the fitted variance condition, if it's positive then the fitted variance is too small; negative values indicate that the fitted variance is too big.
- The offset is related to the fitted mean's condition; a positive offset indicates that the fitted mean is too small while negative offset means that it's too big.
- The skewness and kurtosis of the fit can be observed by the worm plot's curvature. U-shape indicates skewness to the left and an inverted one indicates right skewness. A facing down S-shape indicates platykurtosis while a facing up S-shape indicates leptokurtosis in the distribution.

These key points served as guidelines to assess how changing the edf values affected the model fit. Figure 5 shows an example of a worm plot where the worm is fairly central with a positive slope, indicating that the fitted variance is too small. The skewness can be reduced by increasing the edf but the kurtosis cannot be reduced in the LMS method since it only models the first three moments to bring them close to normality [28]. The kurtosis is not explicitly modelled by the method based on the assumption that its offset from normality is minor. In general, when the plot is out of the 95% confidence interval, the model is poorly fitted.

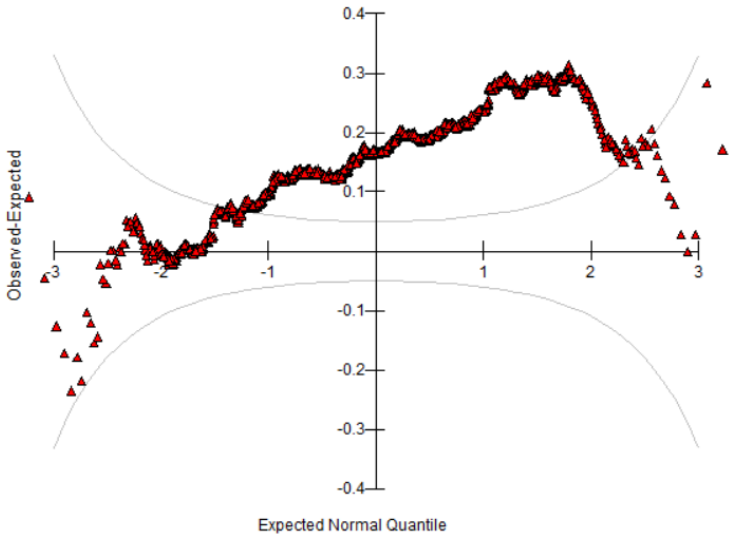


Figure 5: Detrended Q-Q plot

Lastly, a Q-test can be performed to see whether the Z-scores are normally distributed independent of age by looking at its standardized residuals. Royston and Wright conducted a study to compare the goodness-of-fit statistics for age-specific reference intervals, where they described the Q-test to be simple yet inferential, having more power compared to the other tests that they performed [29]. The age range is first divided into G groups with n_g as the size of g^{th} group. Let n be the total sample size, $\bar{z}_g$ and d_g^2 as the group mean and variance, and u_g and v_g the group normal equivalent deviates for skewness and kurtosis. To calculate the Q-statistics for the fitted curves *M, L, S,* and also for the kurtosis, we use the following equations [28]:

$$Q_M = \sum_{g=1}^{G} n_g \, \bar{z}_g{}^2 \tag{12}$$

$$Q_S = \sum_{g=1}^{G} \frac{\left[d_g^{\frac{2}{3}} - \left(1 - \frac{2}{9(n_g - 1)} \right) \right]}{\left[\frac{2}{9(n_g - 1)} \right]} \tag{13}$$

$$Q_L = \sum_{g=1}^{G} u_g{}^2 \tag{14}$$

$$Q_K = \sum_{g=1}^{G} v_g{}^2 \tag{15}$$

Then, the standardized residuals of the Q statistics are plotted against the corresponding degrees of freedom. The standardized residual is denoted as Z and the Q-statistic as Q:

$$Z = \frac{\left(\frac{Q}{v}\right)^{\frac{1}{3}} - \left(1 - \frac{2}{9v}\right)}{\sqrt{\frac{2}{9v}}} \tag{16}$$

Similar to the worm plot, if the curve is adequately fitted, the lines should be centered around zero, roughly within the 95% confidence interval.

Finally, there is still a matter of determining if the curves are undersmoothed/oversmoothed. As seen in Figure 3 and Figure 4, there is a dip at age 15 in the high edf setting (Figure 4) that is not as visible in the low edf setting (Figure 3) due to smoothing. Thus, Cole and Green established two ways to indicate whether the bumps/dips that appears in the curves is a genuine feature of the data or an overfit indication [24]. As the centile curves are derived from the L, M, and S values, they can be plotted separately. If the dip/bump is apparent in the median curve as well as the centile curves, then it is a genuine feature of the data [24]. Another way to check this is to assess the curves' smoothness progression from the lowest to the highest edf setting, and to see if the bump/dip is visible in the low to medium edf settings. As seen in Figure 6, with the edf of 3,5,3; the dip at age 15 is still visible even though it's not as extreme as the dip in the bigger edf settings, with a much more extreme edf setting of 10,10,10. Therefore, this dip is a genuine feature of the data set.

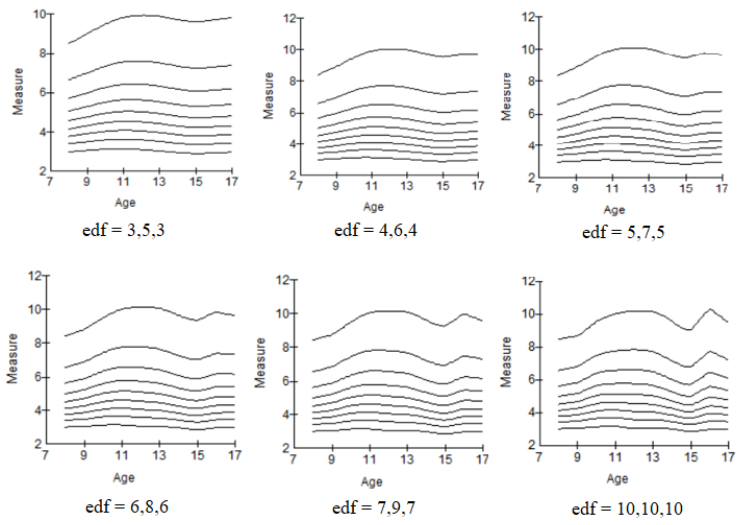


Figure 6: Centile curves smoothness progression using different edf settings

2.4 General framework

After fitting the curves and selecting the optimal edf values based on the qualitative and quantitative assessments as described previously, the decile values (stratified by sex) for each age group were obtained based on Equation (8). Under the same cutoffs (above/below the median) that were used by Prado et al. that was originally proposed by Baumgartner, the ASMI and FMI were used to define four mutually-exclusive body-composition phenotypes: high adiposity and high muscle (HA-HM), high adiposity and low muscle (HA-LM), low adiposity and high muscle (LA-

HM), and low adiposity and low muscle (LA-LM) [12] [10]. The cutoffs were defined according to participants' sex and age accordingly to the following deciles in Table 1. For example, if participant A has an FMI less than the 50th calculated decile for their age and sex and an ASMI higher than the 50th calculated decile for their age and sex, participant A has a phenotype of LA-HM.

Phenotype	FMI Deciles	ASMI Deciles
HA-HM	50-100	50-100
HA-LM	50-100	0-49.99
LA-HM	0-49.99	50-100
LA-LM	0-49.99	0-49.99

Table 1: Classification of body-composition phenotype according to Prado and Baumgartner [10], [12]

2.4.1 Application of methodology

Our first study objective was to create reference curves for children to complement Prado et al.'s work in adults as mentioned in Section 1. To do so, we first replicated Prado et al.'s work in developing the reference curves, and categorizing the adult participants into their phenotype classification according to Table 1. Similar to Prado et al.'s work, the analysis in this study pooled the five DXA imputations available from NHANES while also incorporating the sample weight for each observation [12]. After verifying that our LMS values and corresponding deciles were consistent with the original Prado et al.'s work in the same sample of adults, the methodology was extended to the NHANES children's data.

Finally, to demonstrate its application and fulfill our second objective, cross-sectional data from the Quebec Adipose and Lifestyle InvesTigation in Youth (QUALITY) cohort as outlined in the subsequent section were used. A more detailed framework can be seen in the figure below.

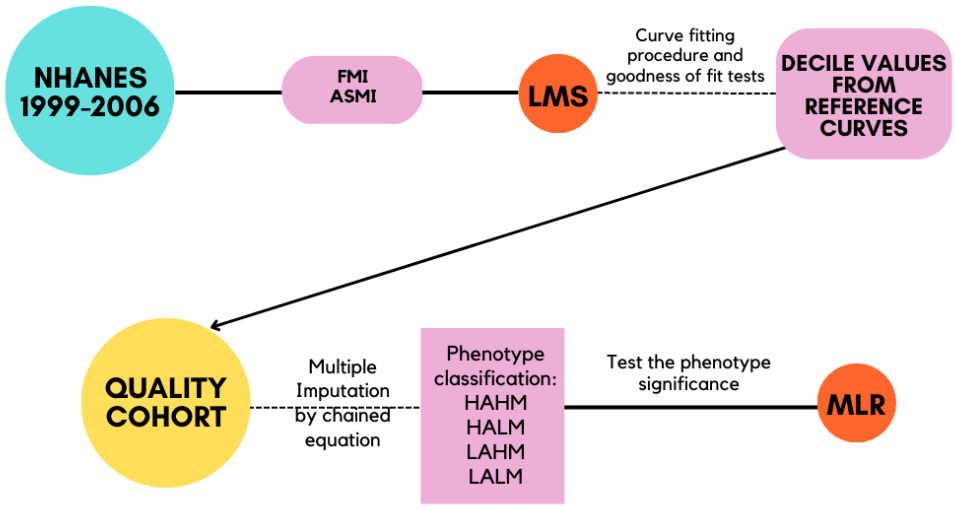


Figure 7: Framework of the study

2.5 QUALITY Cohort

Although the nature of QUALITY is longitudinal, the primary objectives of this thesis were to develop the ASMI and FMI reference curves in NHANES, and to then validate it in a youth sample. Thus, validation analyses were focused on QUALITY's cross-sectional data. Longitudinal analysis is beyond the scope of this thesis but should be conducted in future research.

The QUALITY cohort is an ongoing longitudinal study that aims to investigate the natural history of obesity in children. It used a school-based sampling strategy to identify participants; eligible participants consisted of Caucasian children aged 8-10 years with at least one biological parent with obesity at baseline [30]. Informed consent to participate were provided by parents and participating children provided verbal assent. QUALITY has ethics approval from the Ethics Review Boards of the CHU Sainte-Justine and the Quebec Heart and Lung Institute; this secondary data analysis was approved by the Concordia ethics board (#0000840540).

A total of 634 families participated at the baseline visit (V1) which was completed between 2005-2008. However, four families were removed from the cohort following V1 because the child and/or parents did not complete most or all of the baseline data collection despite initially providing consent (n=630). The first follow-up (V2), when the youth were 10-12 years old was completed between 2008-2011 and conducted for 89% of the original cohort (564 families). Out of the original cohort, 377 families stayed for the second follow up (V3), which was conducted between 2012-2015, when the youth were 15-19 years old. A fourth follow up is currently underway.

To maintain a good sample size, we decided to include 18 and 19 years old in V3 even though our focus of the study is on children. As our curves are only appropriate for children aged 8-17, participants who were 18 and 19 years old in V3 (n=110) were categorized based on Prado et al.'s decile values cutoffs [12]. The compatibility between our decile values and Prado et al.'s decile values [12] is explored in a later section.

2.6 Measures

Data collection included interviewer-administered questionnaires for children, self-administered questionnaires for parents, along with the children's biological and physiological measurements. Anthropometric measurements such as height and weight were taken according to standardized protocols [31]. In children, fat mass and bone mineral density were determined with DXA (Prodigy Bone Densitometer System, DFþ14664, GE Lunar Corporation, Madison, WI, USA). Children's sexual stage of maturity was also measured at each visit by trained nurses according to Tanner stages [32], [33]. Trained nurses determined whether children were pre-pubertal or post-pubertal. Relevant questionnaire data for this research included children's time spent watching TV, video, and/or the computer (hours/day), previous year's annual household income, and the highest education obtained by either parent. A more detailed description of the study design and methods is available [30], [34].

2.6.1 Cardiometabolic risk factors

During each visit, blood was drawn by venipuncture after an overnight fast. Samples were immediately stored on ice and then centrifuged, aliquoted, and stored at -80°C until analysis. All biochemistry analyses were conducted at the Department of Clinical Biochemistry of the CHU

Sainte-Justine, which regularly participates in provincial and international quality control programs and is accredited by the International Federation of Clinical Chemistry [30].

Fasting cardiometabolic risk measures included: total cholesterol (TC), high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), triglycerides, and glucose. Homoeostatic model assessment of insulin resistance (HOMA-IR) was calculated as a derivative variable: (fasting insulin (pmol/L))/7.175× (fasting glucose (mmol/L))/22.5 [35].

2.7 Analytic considerations

2.7.1 Multiple Imputation by Chained Equation (MICE)

From V1 to V2, 66 families were lost to follow-up, and from V2 to V3, another 187 families were lost to follow-up [30]. In addition to lost participants, there were also some missing values in biochemical analyses data, likely due to the invasive nature of some tests and procedures (blood draws). To avoid biased results due to listwise deletion, this study utilized multiple imputation by chained equation (MICE) to handle missing/incomplete data in the QUALITY cohort.

Multiple imputation is often preferred than single imputation due to the reduced "noise" that is increased when the latter method is used. Rubin developed a method to average the outcome across the imputation results by pooling them [36]. First, m plausible complete versions of the data are created by drawing several values for each missing data while also including a random error component. Next, the set of *m* new data sets are analyzed using statistical procedures before they're merged (pooled) into a comprehensive analysis where the uncertainty/variance of the missing data is incorporated.

van Buuren and Groothuis-Oudshoorn developed MICE as an algorithm that preserves the relations in the data as well as the uncertainty about those relations [37]. First, let Y_j with $j = 1, \dots, p$ be one of p incomplete variables so that $Y = (Y_1, \dots, Y_P)$. Denote the observed and missing parts of Y_j as Y_j^{obs} and Y_j^{mis}. Let $Y_{-j} = (Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_P)$ be the collection of $p - 1$ variables except Y_j. The number of imputations is denoted by $m \geq 1$ and let the h^{th}imputation be $Y^{(h)}$, $h = 1, \dots, m$. The quantity of interest (regression coefficient) is denoted by Q and often as a multivariate vector.

The name of the method comes from how the imputed values are obtained. Assume that the multivariate distribution of Y; a hypothetically complete data partially observed from p-variate distribution $P(Y|\theta)$, is completely specified by θ, a vector of unknown parameters. MICE obtains the posterior distribution of θ by iterative sampling from the conditional distributions of the form: $P(Y_1|Y_{-1}, \theta_1), \dots, P(Y_p|Y_{-p}, \theta_p)$.

$\theta_1, \dots, \theta_p$ are specific to each conditional density and thus, with a simple draw from observed marginal distributions, the t-th iteration is a Gibbs sampler that successively draws

$$\theta_1^{*(t)} \sim P(\theta_1|Y_1^{obs}, Y_2^{(t-1)}, \dots, Y_p^{(t-1)})$$

$$Y_1^{*(t)} \sim P(Y_1|Y_1^{obs}, Y_2^{(t-1)}, \dots, Y_p^{(t-1)}, \theta_1^{*(t)})$$

$$\vdots$$

$$\theta_p^{*(t)} \sim P(\theta_p|Y_p^{obs}, Y_1^{(t)}, \dots, Y_{p-1}^{(t)})$$

$$Y_p^{*(t)} \sim P\left(Y_p \middle| Y_p^{obs}, Y_1^{(t)}, \dots, Y_p^{(t)}, \theta_p^{*(t)}\right)$$

where $Y_j^{(t)} = (Y_j^{obs}, Y_j^{*(t)})$ is the jth imputed variable at t-th imputation. From the illustration above, observe that the imputations at $(t-1)$ only enters the imputation at (t) through its relation with other variables and not directly. This results in a quicker convergence unlike many other multiple imputation methods. The name chained equations refers to the fact that MICE can be easily implemented as a series of univariate procedures to fill the incomplete data [37].

Adapting Rubin's method by first creating m imputed datasets, we obtained $Y^{(1)}, \dots, Y^{(m)}$ through pre-determined techniques where they were identical for the non-missing data entries but differed in the imputed values. Then, Q was estimated for each of the imputed data to obtain $\widehat{Q^{(k)}}$ where $k = 1,2, \dots, m$. Finally, for quantities Q that are approximately normally distributed, the mean over $\widehat{Q^{(k)}}$ can be calculated, as suggested by Rubin [36], [37]:

$$Q = \frac{1}{m}\sum_{k=1}^{m} \widehat{Q^{(k)}}.$$

There are several univariate imputation techniques available in ***mice***, but this study used predictive mean matching for numeric variables. As the name suggests, predictive mean matching uses regression to build a model to "match" respondents. For each missing entry, the method forms a small set of candidates from all complete cases that have predicted values closest to the predicted value for the missing entry [38]. Then, a value from the candidates is randomly selected for the imputation.

After the imputation, assessment for the imputation and the distribution of the imputed values is needed to make sure that the values are plausible. No clear-cut method is available to determine whether the algorithm has converged but what is often done is to plot the imputed variables' convergence, density function, and residuals. The plotting of convergence can be done using the **`plot.mids()`** object inside the ***mice*** package in R.

Then, the density of both the observed and imputed values of all variables can be plotted to see whether the imputations are reasonable; differences in densities may suggest a problem. Another diagnostic tool is to observe the distribution of observed and imputed data conditional on the propensity score as suggested by Raghunathan and Bondarenko [39]. The propensity score is the average over the imputations, and the conditional distributions should be similar if the model is fitted well. Lastly, it is also recommended by van Buuren to display the distribution of the residuals in the observed and imputed data since if the amount of overlap is large, then the spread of imputations is appropriate [37].

2.7.2 Statistical analyses

As previously described, for the selection of the final reference curves, we considered the penalized deviance, AIC, the worm plot shape, Q-statistics plot, and we observed the dip/bump in the centile curves closely to avoid under-/over-smoothing to identify the best fitting edf. More specifically, in order to choose the edf, we compared the deviances and AICs for various edf settings to optimize the tradeoff in the smoothness and deviance. This was conducted in lmsChartMaker Pro. After deciding the best edf setting for both FMI and ASMI in females and males, the L, M, S, and decile

cutoffs were imported into R. All statistical analyses were conducted with R version 4.2.0 (The R Foundation for Statistical Computing, Vienna, Austria) and performed separately for boys and girls.

Statistical comparisons between boys and girls (in complete case data, before MICE) in V1, V2, and V3 were conducted using t-test for continuous variables and chi-square tests for categorical variables. Bivariate analyses (ANOVA for continuous and chi-square for categorical variables), were performed to check the relationship between the phenotype classifications and cardiometabolic risk factors (TC, HDL, LDL, triglycerides, glucose, and HOMA-IR) at all visits.

Multiple imputation was conducted with the ***mice*** package (version 3.15.0) [37]. After MICE, the statistical comparison and bivariate analyses were conducted again to determine whether relationships differed between complete-case and MI data. To perform these bivariate analyses, we picked one dataset randomly out of 10 as the algorithm of MICE suggested. Statistical comparisons between complete-case and those lost to follow-up were also conducted.

Lastly, multiple linear regression (MLR) was performed to observe the relationship of each of the cardiometabolic risk and the phenotype classification after adjusting for covariates, cross-sectionally at each timepoint. The modelling was done through pooling of the estimates of the regression coefficient. Covariates previously described (pubertal status, children's screen time, household's previous year's annual income, and the highest education obtained by either parent) were entered into the model based on significant differences detected in the descriptive statistical tests or the literature's recommendations [12], [16].

3 Results

3.1 NHANES study sample

In the NHANES dataset, we calculated the ASMI and FMI in children and found the range to be $(3.37, 18.41)$ for ASMI and $(1.83, 30.20)$ for FMI. In comparison, the range for adults aged 18 years and up were $(3.48, 17.62)$ and $(2.06, 36.43)$ for ASMI and FMI, respectively. The density plot of ASMI and FMI for children and adults in NHANES can be seen in Figure 8. A common feature of both the children's and adults' data is the right-skewness of ASMI and FMI. From Figure 8 Panel A, we observed that the ASMI in female adults and children looked similar. In contrast, although both children and adult measurements were right-skewed, the mean of the distribution of male ASMI and FMI seemed to be quite disparate between children and adults. However, the range of the distributions are similar for both ASMI and FMI.

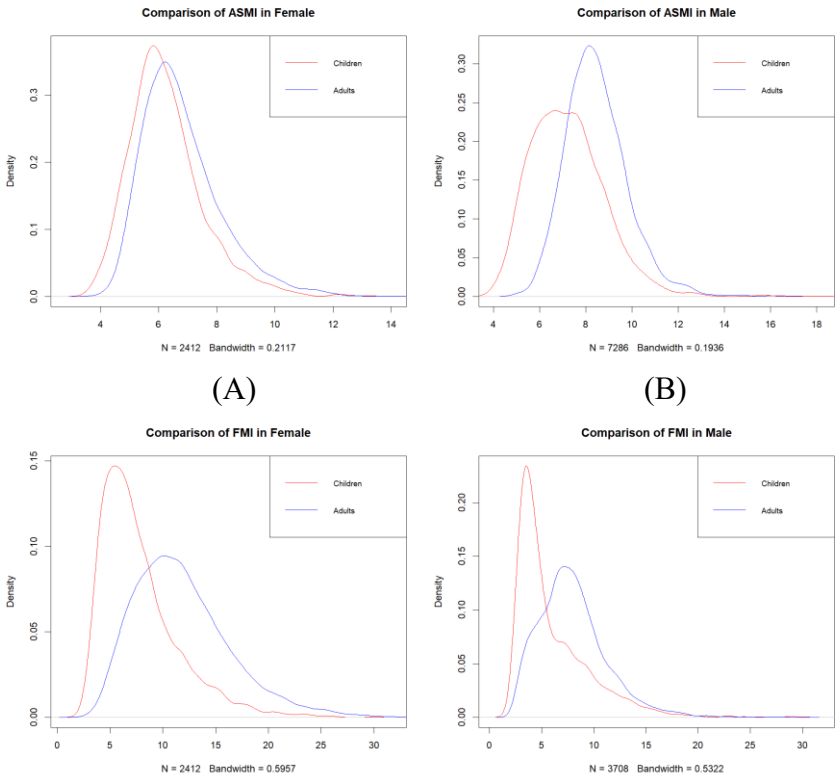


Figure 8: Comparison of ASMI and FMI in children (8-17 years) and adults (≥18 years), separately for females (A, C) and males (B, D)

3.2 Development of reference curves

Table 2 provides the deviance and AIC based on edf settings in female's ASMI data. As can be seen, the deviance decreases as the edf increases since we're fitting more closely to the actual data. The smoothest curve (edf: 3,5,3) has the largest deviance but the second smallest AIC. This is

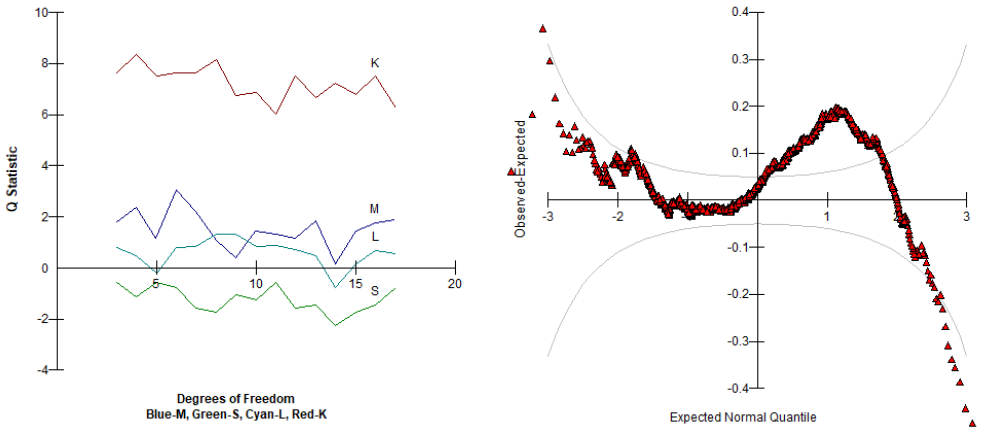
because with AIC, we're penalizing models with more splines/parameters. Thus, looking at both deviance and AIC will help us to find the "best" edf setting for the fit. We determined that the setting of 6,5,3 for female's ASMI yields the best of both: low deviance and low AIC. This same process was applied to female's FMI, as well as the male's ASMI and FMI data.

EDF	P. DEVIANCE	AIC
3,5,3	6,703.1	6,719.8
4,6,4	6,697.0	6,720.6
5,7,5	6,691.8	6,720.9
6,8,6	6,687.0	6,722.9
7,9,7	6,682.9	6,724.6
6,5,3	6,694.1	6,715.8

Table 2: Penalized deviances and AICs for a selection of edf settings in female ASMI

The detrended Q-Q plot and the Q-tests were also assessed as they show the fit of our model by observing the normality of the Z-scores and standardized residuals, respectively. For instance, Figure 9 shows the Q tests (Panel A) and detrended Q-Q plot (Panel B) for female's FMI under the edf of 3,5,3. The Q-tests plot included 4 curves for *L, M, S,* and *K* (kurtosis) where all of them had Q-statistics (as standardized residuals) around 0 except for *K*. This means that the method failed to model the data's kurtosis, as predicted in Section 2.3.3 since it does not model for kurtosis. The worm plot in Figure 9, Panel B has a facing up S-shape indicating a leptokurtosis in the fitted distribution, which aligns with the Q-test results. Increasing the curves' edf did not yield any obvious changes to the shape of the worm plot as the LMS method does not control for kurtosis.

The LMS method performed well for females' FMI as shown in Figure 9, where the Q-statistics of L, M, and S showed normality (centered around zero) except for kurtosis. It also showed that females' FMI had excess kurtosis around 8, also known as leptokurtosis (heavy-tailed).



(A) (B)

Figure 9: Q-test (A) and detrended Q-Q plot (B) for females' FMI with edf 3,5,3

Thus, the edf for our reference curves were chosen by looking at the smoothness of the centile curves and also the fluctuations of penalized deviance and AIC. This was done manually by looping

the algorithm in all of the possible settings for each curve, with the most extreme setting at 10,10,10. For FMI, we found that smaller edf were better than larger and thus we had 3,5,3 for females and 4,6,4 for males. It is worse for males' FMI (Figure 10) than females' FMI since we observed a more prominent facing-up S-shape. The Q-statistics for male FMI are similar to female FMI thus showing that there is a heavy kurtosis in the fitted distribution.

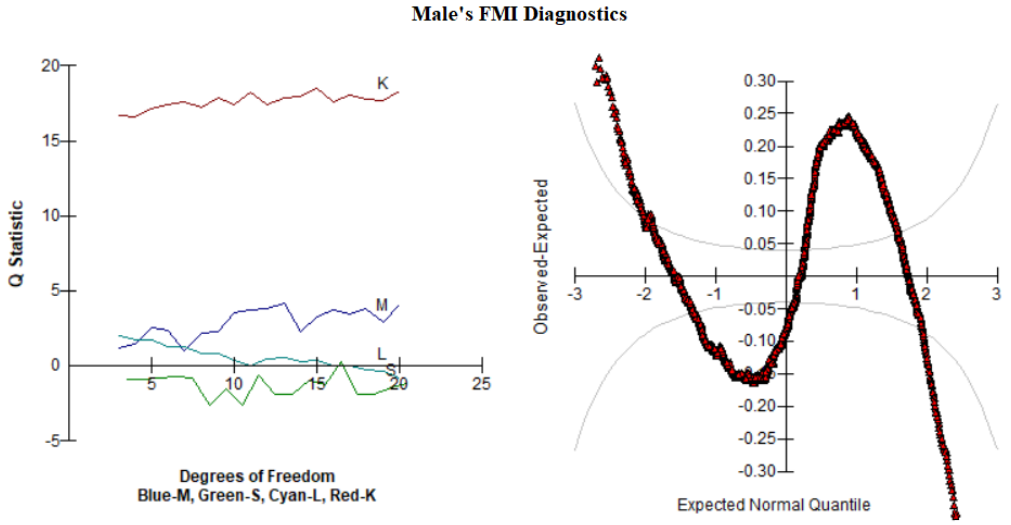


Figure 10: Q-test (left) and detrended Q-Q plot (right) for males' FMI with edf 3,5,3

The worm plot for females' ASMI behaved more appropriately than FMI since we observed a weaker S-shape. The Q-statistics also indicated that the kurtosis is lower than FMI's but it's still there since the curve doesn't revolve around zero. However, the Q-statistics for L, M, S, and K in females' ASMI were very volatile. However, since the lines hovered around the interval of $(-2, 2)$, we have an adequate fit because they are still roughly around zero and within the 95% confidence interval.

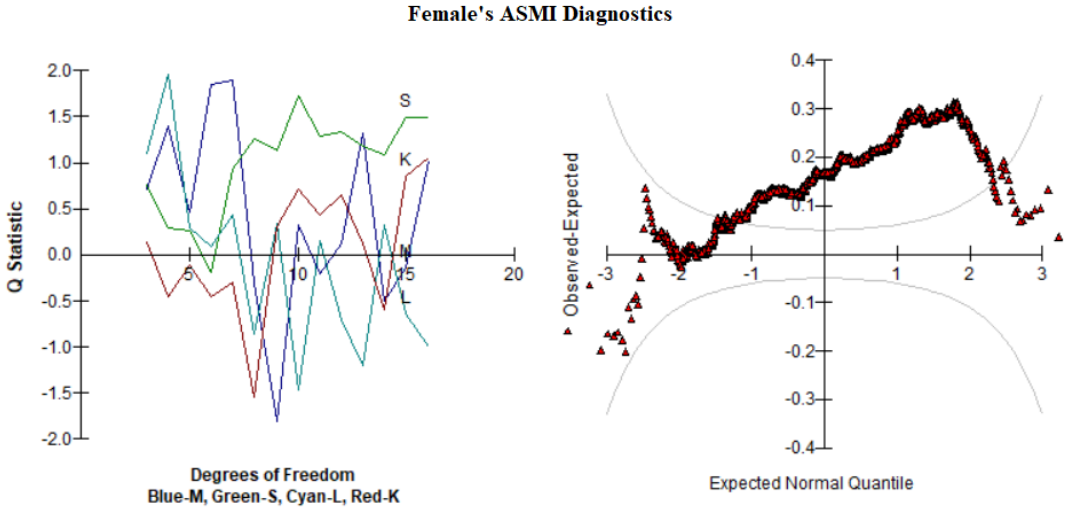


Figure 11: Q-statistics (left) and detrended Q-Q plot (right) for females' ASMI

The Q-statistics for males' ASMI showed different behavior. The S curve is around zero but M and

L are greater than zero, ranging around the interval of $(0.5, 3.5)$. This means that for male ASMI, the Z-scores are not normally distributed independent of age and we also have a kurtosis. Changing the edf setting did not yield a different result since it controlled for the smoothness of the curves and not for the kurtosis and normality. In contrast to the FMI data, ASMI had better deviance and AIC with larger edf settings; we chose 6,5,3 for females and 6,6,8 for males

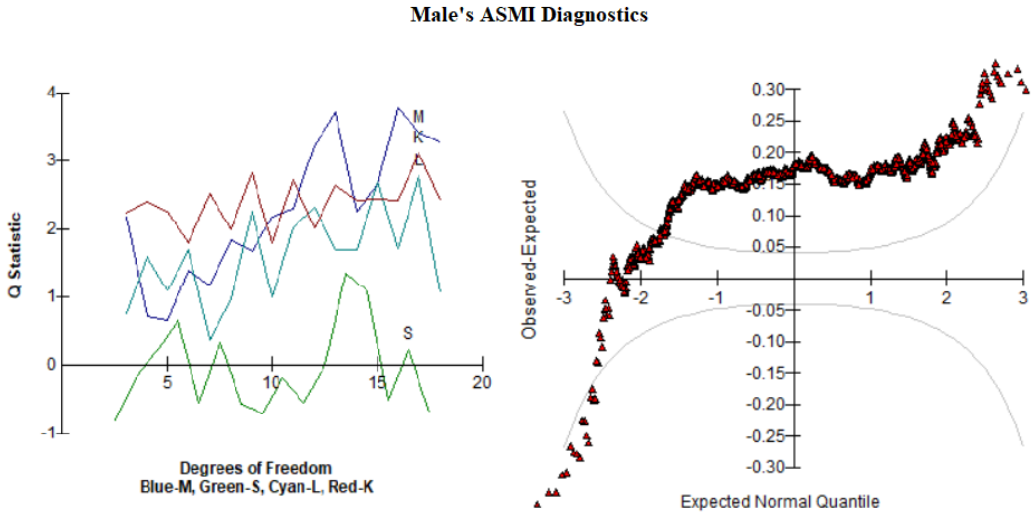


Figure 12: Q-statistics (left) and detrended Q-Q plot (right) for males' ASMI

Therefore, following the considerations of the penalized deviance, AIC, detrended Q-Q plot, and Q-tests, we were able to choose the edf for FMI and ASMI in female and male as follows in Table 3.

Measures	Female	Male
ASMI		
edf	6,5,3	6,6,8
Penalized deviance	6,694.1	10,747.1
AIC	6,715.8	10,772.3
FMI		
edf	3,5,3	4,6,4
Penalized deviance	11,728.6	16,832.8
AIC	11,745.0	16,856.1

Table 3: Final edf settings for female and male reference curves

It's not recommended to choose a much more extreme setting (e.g., 10,10,10) as the final edf because the curves will be extremely jagged. But, to further strengthen our confidence in the results, we tested the decile values of our chosen setting and compared it with the 10,10,10 results. To test this, we imported the decile values, from the 10th percentile to the 90th percentile into R, where we categorized the NHANES sample into the phenotypes (HA-HM, HA-LM, LA-HM, LA-LM) based on Table 1. Next, we calculated the classification difference between our chosen edf and the 10,10,10. For females, the difference was 0.038% and it was even lower for male, 0.028%. This is a good sign since this means that our model has a high similarity with the 10,10,10 but it's smoothed while still keeping its important features.

3.2.1 L, M, S curves

Before we plot the centile curves, it is important to also observe the L, M, and S curves. The L curve will help identify the skewness of the distribution and the S curve will show the coefficient of variation over time. Finally, as mentioned in Section 2.3.3, the M curves are used to detect whether a dip/bump in the centile is a true feature or an overfit indication.

As seen in Figure 13 and Figure 14, there are no bumps/dips noticeable in the M curves for females. For both measurements, the L curve shows the box-cox power used to normalize the observation. Overall, the (absolute) box-cox power is larger for ASMI, indicating that its distribution is more skewed than FMI. The S curves explains the coefficient of variation (CV) in the distribution of ASMI/FMI. In ASMI, we observed a small CV at pubertal age (8-11 years) while it's the opposite for FMI. Then, the curve began to increase for ASMI with age 14 having the biggest variation. In contrast, age 14 has the lowest CV in the S curve of FMI.

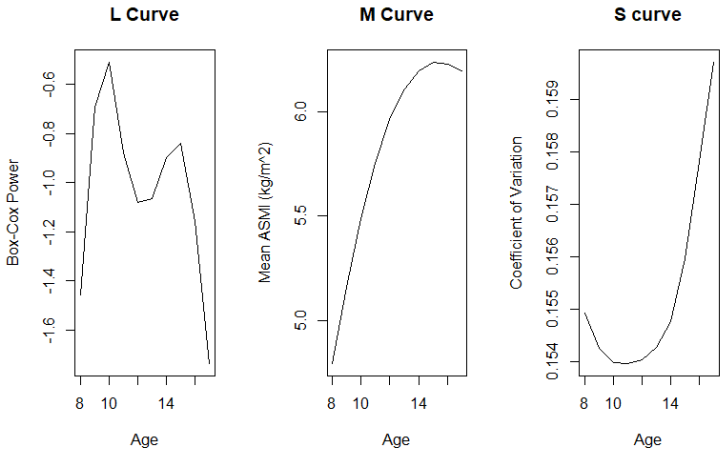


Figure 13: L, M, and S curves for females' ASMI

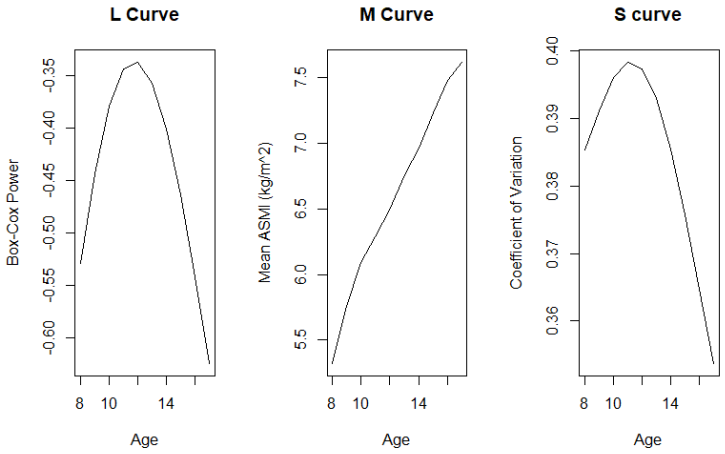


Figure 14: L, M, and S curves for females' FMI

For males, the M curve has no bump/dip in ASMI at all ages, as seen in Figure 13. However, there is a steep decline in the M curve of FMI at age 15. Similar to the female data, male's ASMI has a

larger (absolute) value of box-cox power than male's FMI. There is a different trend in the S curve in males compared to females which is due to the later pubertal age in males. The peak of the S curve for both ASMI and FMI in males appears to be in age 13-14 years old. The *L, M*, and *S* values as well as the deciles can be found in Appendix A.

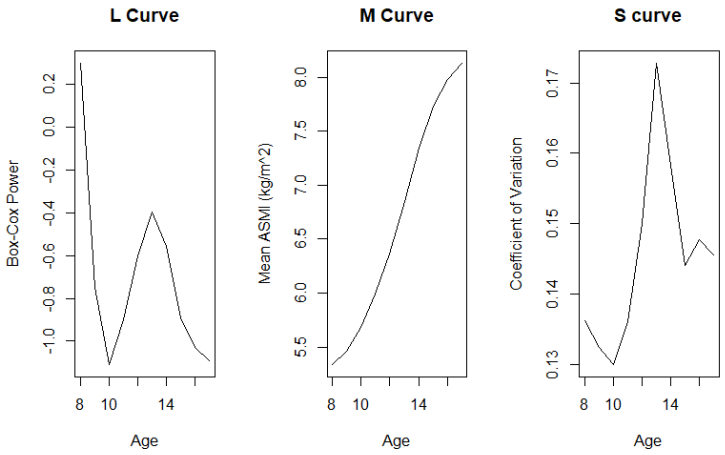


Figure 15: L, M, and S curves for males' ASMI

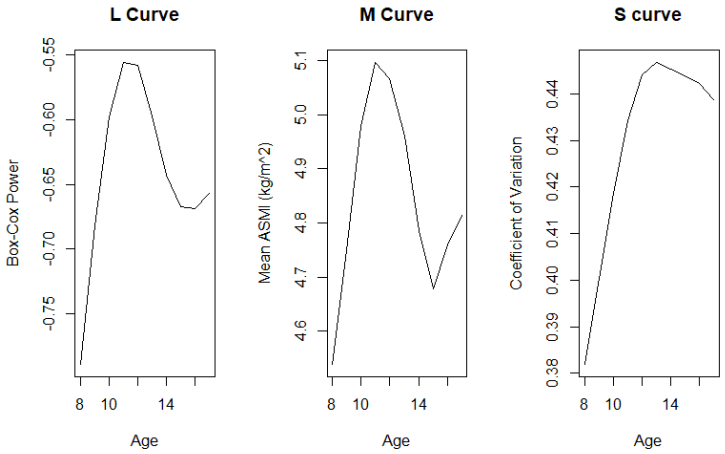


Figure 16: L, M, and S curves for males' FMI

3.2.2 Centile curves

An increasing trend was observed in the ASMI curves (Figure 17 Panel A) for female aged 8-17 years old, where the slope is steepest from age 8-13 years, and begins to flatten thereafter. The female FMI curves in Figure 17 Panel B also have a similar pattern but with a less steep slope compared to ASMI. In addition, there is a big jump from the 80th to 90th percentile, which is not as visible in the ASMI curves. Compared to females, Figure 17 Panel C shows that males have an ASMI curve with a steeper slope. A significant increase was observed from age 12 to 13 years, and after 13 years old the curve appears to be more stable. The FMI curves for males (Figure 17 Panel D) showed a large increase from age 8-12 years, followed by a decrease from age 12-15 years, where we observed a significant dip in the age of 15 years. This dip is a feature of the dataset since it also appeared in the M curve in Figure 9. After age 15, the curve began to increase slightly. This

differs greatly compared to the FMI curves for females seen in Figure 17 Panel B as we did not see any dips and the slopes were more consistent throughout the ages.

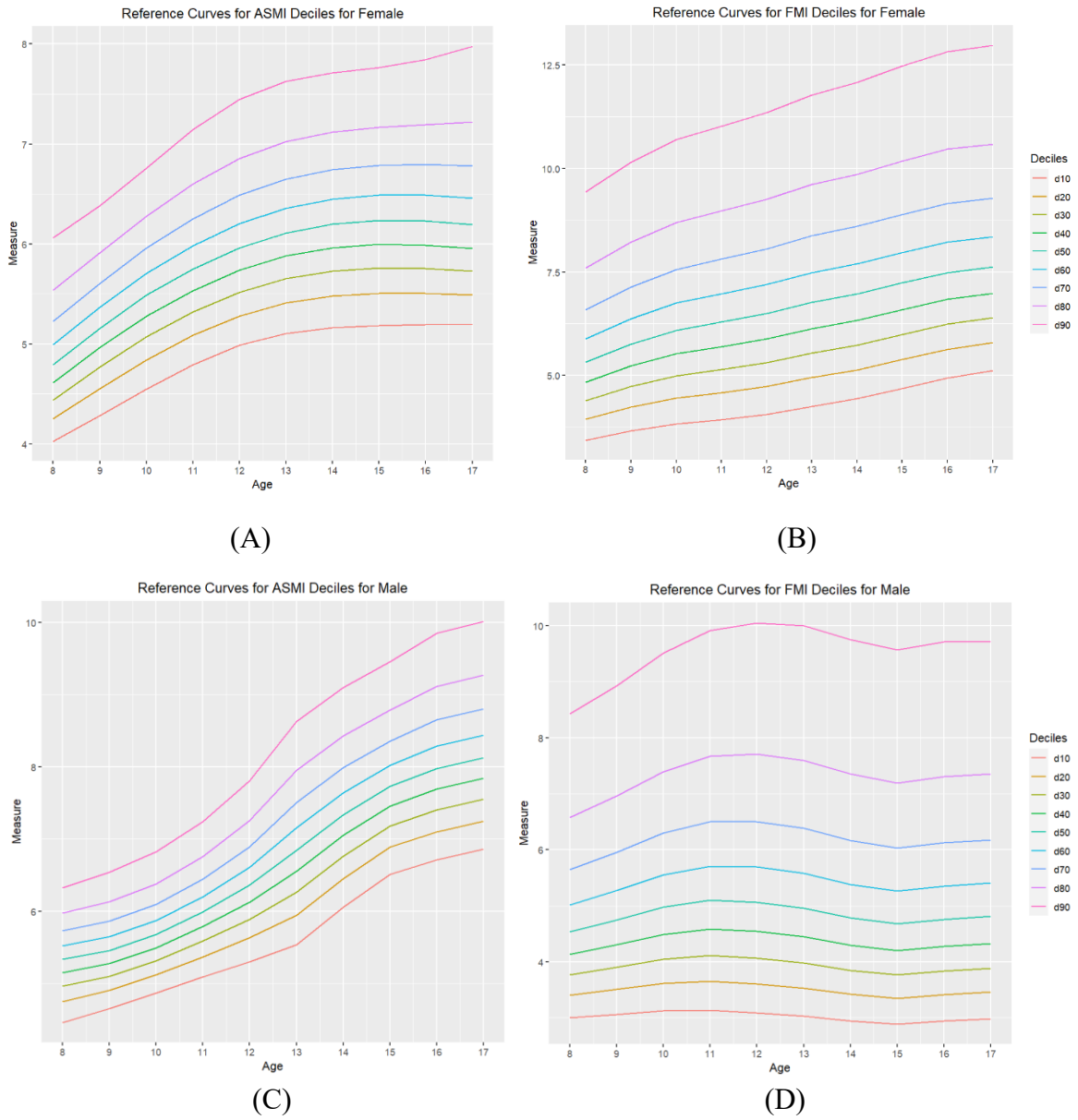


Figure 17: ASMI and FMI reference curves for female (A, B) and male (C, D) age 8-17 years

As previously mentioned, in order to maximize our sample size at visit three, we incorporated the reference curves developed in adults by Prado et al [12]. Thus, similarity between our curves for 8-17 year olds from Prado et al.'s curves for 18+ year olds must be assessed. We observed a large jump between our 50th percentile at 17 years old to Prado's 50th percentile at 18 years old for FMI among females (=7.618 and 8.658, respectively) as seen in Figure 17 (Panel B) and Figure 18 (Panel B) [12]. A smaller jump was also visible in females' ASMI. In contrast for males' ASMI, there was no substantial jump noticed from age 17 to age 18 and the lines looked smooth. Furthermore, the increase was 1.04 for females FMI, 0.72 for males FMI, and 0.293 for females ASMI.

In other words, the median of FMI for age 18-19 years is substantially higher than the median for 17 years old. Thus, the cutoff value for participants who turned 18 or 19 is higher. We found that out of 201 children aged 18-19 years at V3, 60 of them switched to a different phenotype, after being stable in one phenotype group throughout V1 to V2. Out of the 60 who changed phenotype, 35 were in LA-LM at V3. The implications of this on our study results are described in the discussion. The complete L, M, S, and decile values of our reference curves for children aged 8-17 years, as well as Prado's results for age 18-19 years can be found in Appendix A.

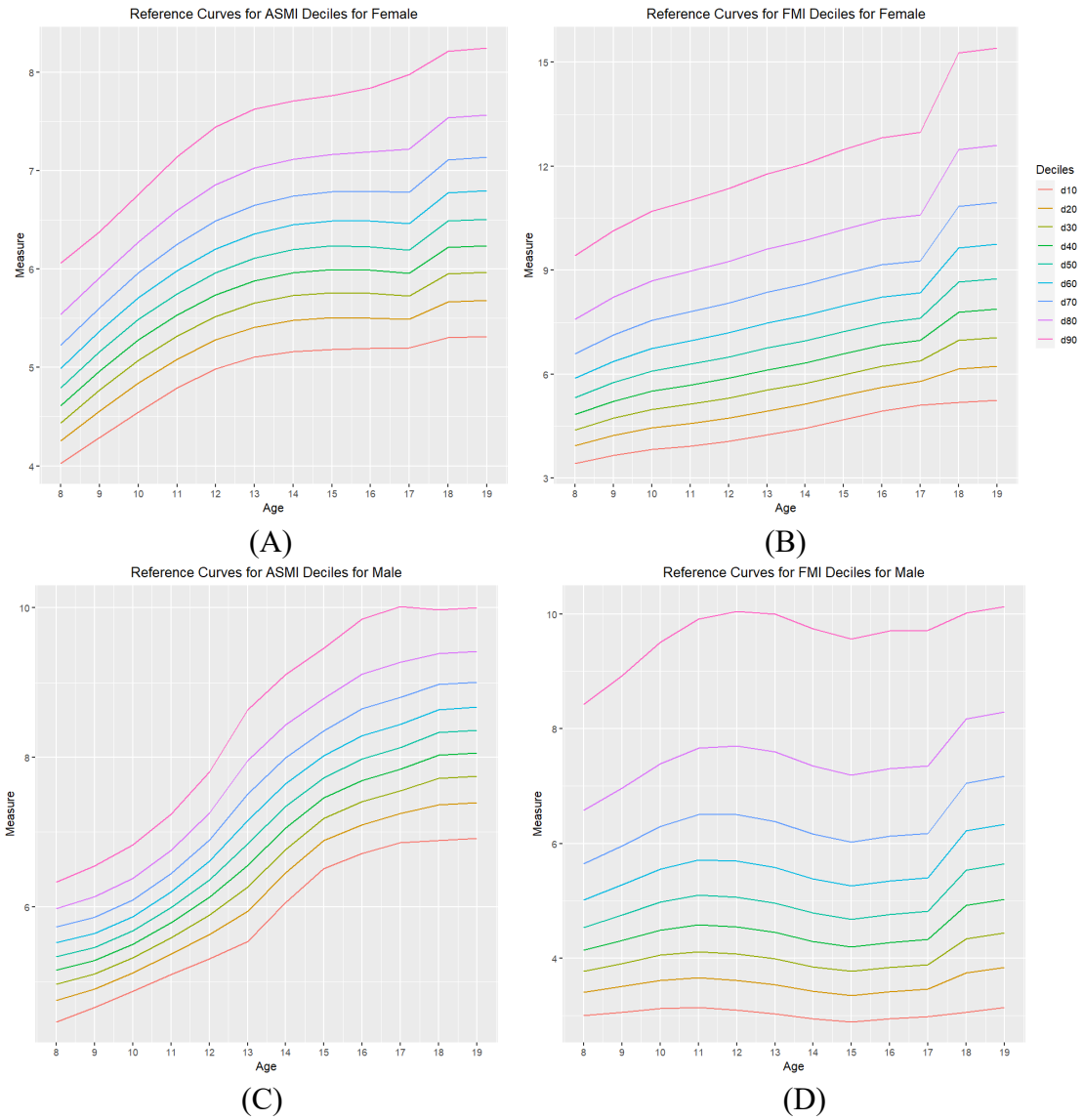


Figure 18: ASMI and FMI reference curves for female (A, B) and male (C, D) age 8-19 years from Prado et al. (2014)

3.3 QUALITY

The QUALITY cohort data set had missing values in all waves, the largest occurred in the third wave with 352 missing values. The decile cutoffs found in Appendix A was used to categorize the participants based on their sex, age, FMI, and ASMI. We assessed the descriptive statistics of sex-stratified complete case observations for V1, V2, and V3, presented in Table 4. Almost all characteristics were different between females and males in V1 and V3, except for their demographics such as age, household income, and parent's education. We observed a different pattern in V2, where some of the cardiometabolic risks that were significant at V1 were not significant at V2 such as: total cholesterol, HDL, LDL, triglycerides, and FMI. Those cardiometabolic risks began to differ significantly again in V3. The phenotype between male and female were not significantly different. Overall, the phenotype classification was distributed similarly in both male and female at all three timepoints.

Table 4: Characteristics of the complete-case observations in the QUALITY cohort study, by sex

Characteristics	**Male**			**Female**		
	V1 ($n = 331$)	**V2 ($n = 179$)**	**V3 ($n = 144$)**	**V1 ($n = 280$)**	**V2 ($n = 155$)**	**V3 ($n = 134$)**
Age (years), mean (SD)	9.62 (0.90)	11.59 (0.91)	16.81 (0.94)	9.56 (0.95)	11.67 (0.99)	16.80 (1.01)
Tanner's stage (%)[1]						
Not in pubertal stage	**91.24**	**44.70**	**0.00**	**64.29**	**21.23**	**0.75**
Parent's education (%)						
1 or both parents with a university degree	45.02	50.84	N/A	46.79	47.74	N/A
Household income, mean (SD)	42,946 (18.8K)	49,076 (20.5K)	56,995 (23.3K)	41,910 (18.4K)	49,833 (22.2K)	59,968 (24.8K)
Screen use (hours/week), mean (SD)	**6.96 (4.33)**	**10.19 (4.84)**	**10.08 (5.04)**	**5.26 (4.01)**	**7.99 (4.20)**	**7.65 (3.95)**
ASMI (kg/m^2), mean (SD)	**5.65 (0.65)**	**6.09 (0.78)**	**8.11 (0.97)**	**5.31 (0.67)**	**5.70 (0.76)**	**6.33 (0.84)**
FMI (kg/m^2), mean (SD)	**5.02 (3.45)**	5.74 (3.46)	**5.76 (3.89)**	**6.05 (3.35)**	6.21 (3.11)	**8.40 (4.06)**
Cardiometabolic risks, mean (SD)						
Total cholesterol(mmol/L)	**3.86 (0.68)**	3.75 (0.61)	**3.50 (0.59)**	**4.01 (0.72)**	3.73 (0.62)	**3.85 (0.77)**
HDL (mmol/L)	**1.21 (0.26)**	1.18 (0.26)	**1.05 (0.20)**	**1.16 (0.24)**	1.19 (0.24)	**1.24 (0.25)**
LDL (mmol/L)	**2.30 (0.57)**	2.25 (0.51)	**2.05 (0.49)**	**2.45 (0.60)**	2.19 (0.53)	**2.24 (0.62)**
Triglycerides (mmol/L)	**0.76 (0.37)**	0.71 (0.37)	0.91 (0.45)	**0.89 (0.44)**	0.78 (0.44)	0.88 (0.39)
Glucose (mmol/L)	**5.00 (0.36)**	**5.08 (0.30)**	**5.19 (0.36)**	**4.89 (0.36)**	**5.00 (0.36)**	**4.95 (0.38)**
HOMA-IR (mmol/L)	**0.93 (0.62)**	**1.25 (0.77)**	**1.60 (1.15)**	**1.16 (0.81)**	**1.49 (0.97)**	**1.36 (0.77)**
Phenotype (%)						
HA-HM	25.98	24.58	31.25	32.14	22.58	27.61
HA-LM	14.20	23.46	13.89	12.50	16.13	15.67
LA-HM	17.52	10.06	18.75	13.57	14.19	15.67
LA-LM	42.30	41.90	36.11	41.79	47.10	41.05

[1]Chi-square test was used only for V1 and V2 since all male were in pubertal age at V3.

[2]Boldface indicates the p-value comparing male with female participants was significant (<0.05) at the same timepoint.

From Table 5, it was observed that there was a significance relationship between phenotype with HDL, triglycerides, and HOMA-IR at all timepoints in the complete cases. Phenotype had a significant relationship with age, Tanner's stage, and LDL only at V1. For glucose, phenotype was significant at V1 and V3. Furthermore, phenotype was associated with parent's education, but not with household income or children's screen time.

Table 5: ANOVA and Chi-Square test for phenotype classifications, complete case

Characteristics	***p-value***		
	V1 $(\boldsymbol{n = 611})$	**V2** $(\boldsymbol{n = 334})$	**V3** $(\boldsymbol{n = 278})$
Age (years)	<0.01	0.016	0.436
Sex	0.279	0.255	0.734
Tanner's stage	<0.01	0.076	0.659
Parent's education	0.012	0.012	N/A[1]
Household income	0.165	0.175	0.109
Screen use (hours/week)	0.228	0.159	0.081
Cardiometabolic risks (mmol/L)			
Total cholesterol	0.0184	0.208	0.388
HDL	<0.01	<0.01	<0.01
LDL	<0.01	0.012	0.386
Triglycerides	<0.01	<0.01	<0.01
Glucose	<0.01	0.019	<0.01
HOMA-IR	<0.01	<0.01	<0.01

[1]There were no parents' education information in V3

Due to the loss to follow-up, statistical tests comparing characteristics between those who were retained, and those who were lost to follow-up were conducted (Table 6). To assess differences at V2, baseline characteristics were compared and similarly, at V3, V2 characteristics were compared. There were no significance differences in age, screen use, HDL, triglycerides, glucose, and HOMA-IR level in participants who were lost to follow-up by V2 compared to those who were retained. However, there was strong evidence that their household income differed. Participants who were lost to follow-up had a lower average household income (3,000 dollars lower). In addition, their FMI, LDL, and total cholesterol differed as well. In contrast, there were no statistically significant differences in V2 characteristics between those retained and those lost to follow-up at V3.

Table 6: Characteristics' differences between participants who were retained and lost to follow up

Characteristics	**Retained at V2** ($\boldsymbol{n = 327}$)	**Lost to follow-up at V2** ($\boldsymbol{n = 284}$)	**p-value[1]**	**Retained at V3** ($\boldsymbol{n = 217}$)	**Lost to follow-up at V3** ($\boldsymbol{n = 117}$)	**p-value**
Age (years), mean (SD)	9.57 (0.92)	9.62 (0.93)	0.46	11.68 (0.93)	11.55 (0.97)	0.23
Sex (%)						
Male	52.94	55.55	0.91	53.05	54.54	0.85
Tanner's stage (%)						
Not in puberty	77.40	80.56	0.38	32.39	36.36	0.15
Parent's education						
1 or both parents with university degree	52.94	55.56	0.28	51.64	48.76	0.54
Household income, mean (SD)	43,996 (18.3K)	40,761 (19K)	0.03	50,137 (21.3K)	48,179 (21.5K)	0.42
Screen use (hours/week), mean (SD)	5.91 (4.38)	6.50 (4.14)	0.09	9.38 (4.71)	8.82 (4.64)	0.28
ASMI (kg/m^2), mean (SD)	5.42 (0.64)	5.58 (0.73)	0.03	5.96 (0.80)	5.83 (0.79)	0.16
FMI (kg/m^2), mean (SD)	5.17 (3.19)	5.86 (3.70)	0.01	5.88 (3.31)	6.10 (3.30)	0.57
Cardiometabolic risks, mean (SD)						
Total cholesterol (mmol/L)	3.88 (0.62)	3.99 (0.78)	0.05	3.72 (0.62)	3.77 (0.61)	0.43
HDL (mmol/L)	1.20 (0.24)	1.18 (0.27)	0.28	1.18 (0.25)	1.19 (0.27)	0.87
LDL (mmol/L)	2.32 (0.52)	2.43 (0.65)	0.02	2.20 (0.54)	2.26 (0.48)	0.26
Triglycerides mmol/L)	0.80 (0.39)	0.84 (0.44)	0.18	0.75 (0.45)	0.72 (0.28)	0.49
Glucose (mmol/L)	4.94 (0.34)	4.96 (0.38)	0.59	5.05 (0.33)	5.03 (0.33)	0.59
HOMA-IR (mmol/L)	0.99 (0.64)	1.08 (0.81)	0.14	1.39 (0.95)	1.32 (0.71)	0.48
Phenotype (%)						
HA-HM	26.32	31.60		23.35	20.66	
HA-LM	13.00	13.89	0.74	14.55	29.75	0.43
LA-HM	14.55	17.01		12.21	11.57	
LA-LM	46.13	37.50		49.89	38.02	

[1]*P* comparing participants who stayed and dropped at V2 or V3.

3.3.1 MICE imputation

Due to the large number of participants lost to follow up and other missing data for specific variables at each visit, we performed imputation using MICE with $m = 10$ for each time point. V1 had the lowest number of missing values (611 out of 630 were complete), 331 out of 630 were complete in V2, and 278 out of 630 were complete in V3. After the MICE algorithm was applied, we checked the range of the numeric imputed values for each of the visits (Table 7). As all imputed values were similar with one another, and the range of the values was small, the imputation did not likely severely alter data distributions.

Table 7: The range of the imputed values across all 10 imputed datasets

Characteristics	**V1**	**V2**	**V3**
Observations imputed	19	299	352
Age (years)	9.597	11.62-11.66	16.760-16.890
Household income	42,251-42,433	48,186-49,479	54,569-57,557
Screen use (hours/week)	6.228-6.259	9.041-9.474	8.767-9.347
ASMI (kg/m^2)	5.498-5.504	5.984-6.009	7.168-7.255
FMI (kg/m^2)	5.490-5.513	6.224-6.360	6.697-7.170
Cardiometabolic risks			
Total cholesterol (mmol/L)	3.926-3.932	3.745-3.771	3.681-3.741
HDL (mmol/L)	1.186-1.188	1.160-1.170	1.133-1.144
LDL (mmol/L)	2.367-2.373	2.232-2.26	2.142-2.190
Triglycerides (mmol/L)	0.817-0.821	0.767-0.778	0.889-0.924
Glucose (mmol/L)	4.941-4.949	5.034-5.056	5.039-5.107
HOMA-IR (mmol/L)	1.039-1.048	1.424-1.472	1.440-1.506

After the imputation, it is necessary to check the density plot of each imputed variable to assess whether they behaved similarly with the true distribution. Figure 19 shows the density plot for all imputations across all timepoints; the red lines are the distribution of the imputed values and the blue line is the observed distribution. For V2 and V3, the imputed (red) and observed (blue) lines were consistent with one another, indicating that all of the imputed values are plausible and we achieved a close estimation from the true observed data.

However, for V1, the density plot showed that the imputed values were different from the observed data and that the distributions had different shapes. This is likely due to the small number of missing values in V1. We have 611 out of 630 complete-case observations in V1, and all variables had less than 6 missing values. The density plot showed only the imputed values distribution and not the entirety of the whole imputed data set ($n = 630$). Therefore, due to the lack of incomplete observations, the shape of the imputed values appeared to be very extreme and thin. Nevertheless, the range of imputed values as shown previously in Table 7 suggest that the complete dataset after imputation for V1 behaved appropriately and cohesively with the observed data.

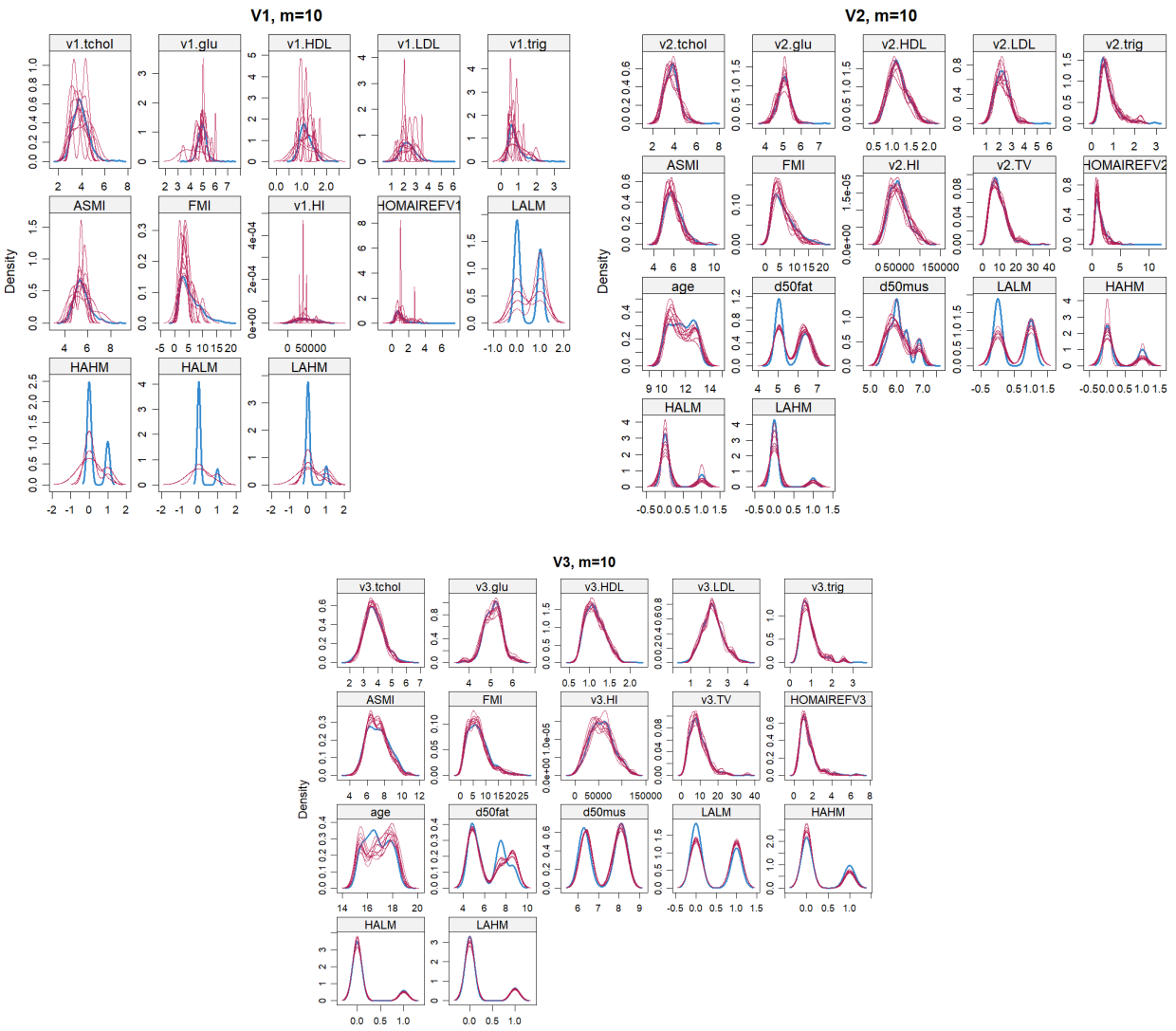


Figure 19: Density estimates for the marginal distributions of the observed data (blue) and the $m = 10$ densities per variable from the imputed data (red) at all timepoints

Next, we compared the imputation results on the variable level. First, we compared the distributions of observed and imputed data based on the propensity score as mentioned in Section 2.7.1. As we found similar results with all other variables, only ASMI and FMI at V2 is presented. Figure 20 shows the ASMI and FMI values at V2 plotted against its probability of missing. The blue points are the score for the observed and red points are the imputed. Notice that there are 11 grids in each plot, with the top left being the actual data without any imputations and the rest (10 plots) are the observed data plotted along with the imputation done with MICE. Overall, the blue and red points conformed well and there seemed to be no red points that behaved like an outlier. This provides evidence that the imputations for ASMI and FMI at V2 are reasonable.

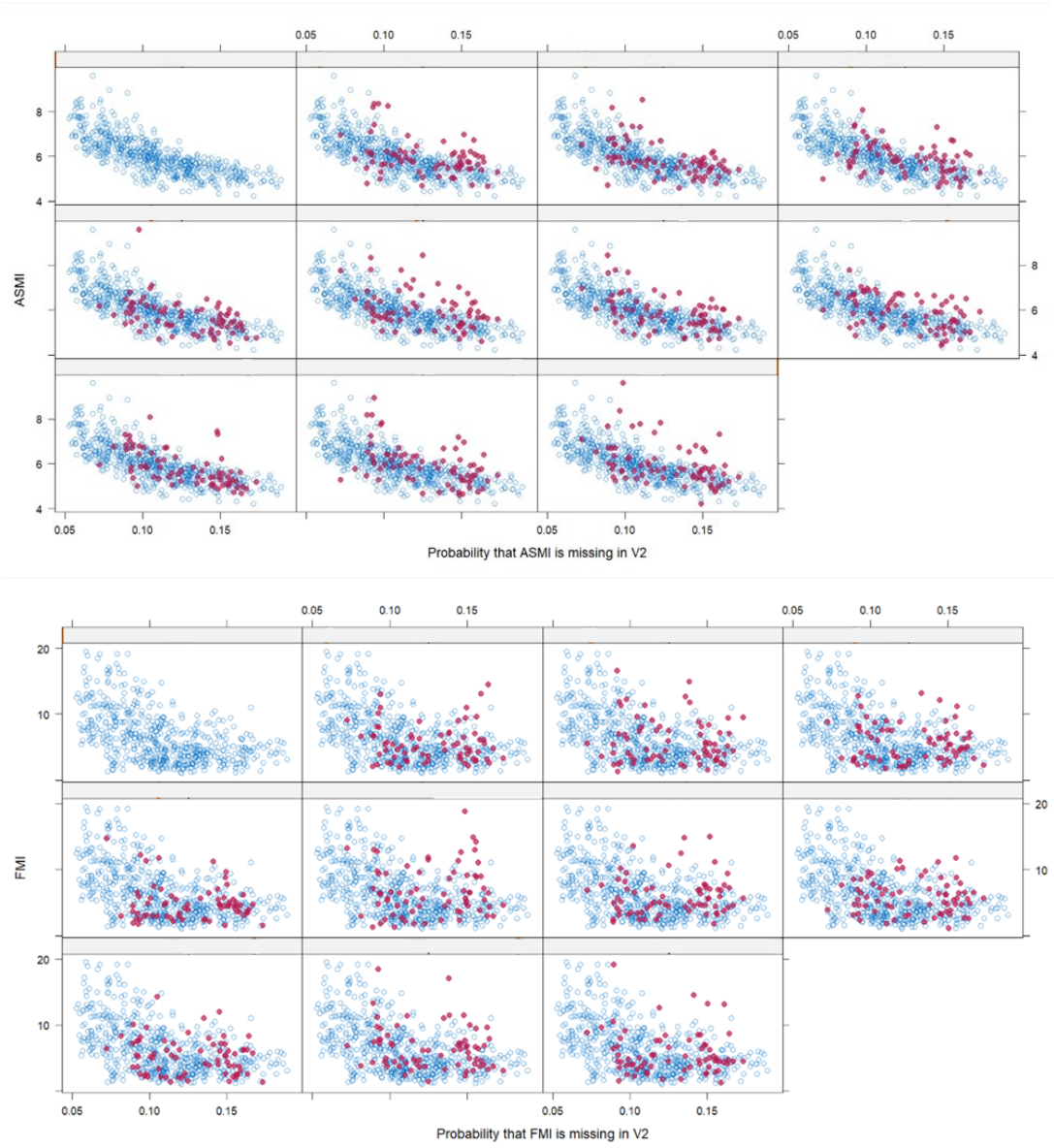


Figure 20: ASMI (top) and FMI (bottom) at V2 against its propensity score (blue for the observed values and red for the imputed values)

We further investigated the imputations' quality by looking at the residual of the missing data for each of the variables. Figure 21 showed that the amount of overlap between the observed and imputed is large, indicating that the spread of imputations is appropriate, as defined by van Buuren and Groothuis-Oudshoorn [37]. Thus, from both the propensity and residual plot, we concluded that the imputation for all variables was appropriate. Based on these analyses of the imputation, it's safe to move forward using our imputed values.

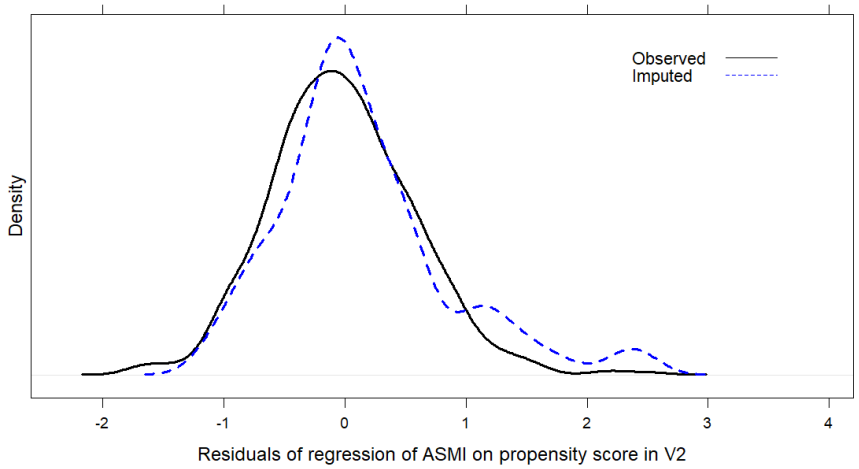


Figure 21: Distributions of the residuals of regression of ASMI at V2

Lastly, we analyzed the bivariate relations between the imputed dataset and original dataset to see if any of the relationship differed. To assess these relationships, we randomly chose one dataset out of 10 since their imputed values from Table 7 did not indicate large variation across the imputed datasets. From Table 8, the phenotype classification for males and females were consistent as before MICE. However, we found that some of the sex-adjusted characteristics had changed compared to the sex comparisons among the complete cases as seen in Table 4. For instance, age at V2 and V3 were now significantly different between males and females, Tanner's stage was no longer significantly different at V3, and FMI at V2 was now significantly different. As for the cardiometabolic risks, similar patterns as before were detected except for HOMA-IR which was now insignificant at V3 (data not shown).

Given that the imputation did not appear to significantly alter data distributions, the rest of the results will focus on the imputed data sets. Figure 22 shows that after imputation, LA-LM had the biggest proportion in the study sample; more detailed classification proportions can be seen in Table 9. Throughout V1 to V3, LA-LM has the largest proportion followed by HA-HM. This means that participants were either (1) lacking in adiposity and muscle or (2) had excess adiposity and high muscle. At V1 and V3, more participants were classified as LA-HM than HA-LM. This is different than the results at V2, since we observed more participants in HA-LM than LA-HM.

Table 8: Characteristics of the MICE-imputed dataset, by sex

Characteristics	**Male ($n = 343$)**			**Female ($n = 287$)**		
	V1	**V2**	**V3**	**V1**	**V2**	**V3**
Age (years), mean (SD)	9.62 (0.89)	**11.72 (0.92)**	**16.70 (0.94)**	9.57 (0.95)	**11.57 (0.97)**	**17.00 (1.10)**
Tanner's stage (%)						
Not in pubertal age	**90.67**	**45.19**	2.33	**63.76**	**17.42**	3.14
Parent's education (%)						
1 or both parents with university degree	54.52	48.69	N/A	54.00	51.92	N/A
Household income, mean (SD)	42,713 (18.8K)	48,781 (21.7K)	55,416 (24.8K)	42,099 (18.3K)	48,122 (22.8K)	56,354 (24K)
Screen use (hours/week), mean (SD)	**6.99 (4.34)**	**10.6 (5.23)**	**9.75 (4.63)**	**5.36 (4.17)**	**8.03 (4.16)**	**8.46 (4.64)**
ASMI (kg/m^2), mean (SD)	**5.66 (0.65)**	**6.20 (0.87)**	**7.95 (0.95)**	**5.33 (0.69)**	**5.78 (0.77)**	**6.31 (0.77)**
FMI (kg/m^2), mean (SD)	**5.01 (3.42)**	**5.94 (3.76)**	**5.55 (3.77)**	**6.10 (3.40)**	**6.84 (3.66)**	**8.57 (4.00)**
Cardiometabolic risks, mean (SD)						
Total cholesterol (mmol/L)	**3.87 (0.68)**	3.76 (0.66)	**3.64 (0.65)**	**3.99 (0.72)**	3.76 (0.67)	**3.84 (0.70)**
HDL (mmol/L)	**1.21 (0.26)**	1.18 (0.26)	**1.08 (0.22)**	**1.16 (0.24)**	1.16 (0.24)	**1.19 (0.24)**
LDL (mmol/L)	**2.31 (0.57)**	2.25 (0.56)	**2.14 (0.53)**	**2.43 (0.60)**	2.23 (0.58)	**2.23 (0.59)**
Triglycerides (mmol/L)	**0.77 (0.37)**	0.74 (0.42)	0.93 (0.51)	**0.89 (0.44)**	0.80 (0.41)	0.91 (0.39)
Glucose (mmol/L)	**4.99 (0.36)**	**5.09 (0.35)**	**5.26 (0.36)**	**4.89 (0.36)**	**5.02 (0.37)**	**4.91 (0.41)**
HOMA-IR (mmol/L)	**0.93 (0.61)**	**1.33 (0.86)**	1.53 (0.99)	**1.18 (0.83)**	**1.64 (1.17)**	1.43 (0.81)
Phenotype (%)						
HA-HM	26.82	27.70	26.82	32.75	30.31	26.83
HA-LM	13.70	22.16	17.20	12.54	14.63	16.03
LA-HM	16.91	11.08	17.49	13.24	11.85	14.63
LA-LM	42.57	39.06	38.49	41.47	43.21	42.51

Note: Boldface indicates the p-value comparing male and female participants was significant (< 0.05) at that time.

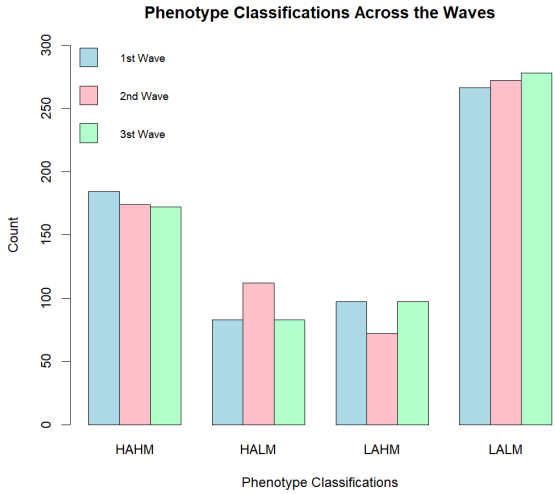


Figure 22: Phenotype classifications based on one randomly selected imputed dataset

Table 9: Phenotype classifications based on one randomly selected imputed dataset, $n = 630$

Phenotype Classification (%)	V1	V2	V3
HA-HM	29.53	28.29	26.20
HA-LM	13.17	18.73	14.29
LA-HM	15.24	11.27	16.51
LA-LM	42.06	41.41	43.01

3.4 Regression results

We performed multiple linear regression to calculate the association between each of the cardiometabolic risks at all time points with the phenotypes, after adjusting for age, sex, Tanner's stage. LA-HM was the reference group for the phenotype classification (Table 10). These results did not differ when stratified by sex; thus, the full sample adjusted for sex are uniformly presented. Compared to the reference group of LA-HM, both HA-HM and HA-LM were significantly associated with worse HDL, worse triglycerides and worse HOMA-IR at V1 (p-value<0.05). Compared to LA-HM, LA-LM at V1 was associated with lower glucose. Compared to LA-HM, HA-HM was the only phenotype that was significant related to higher LDL. We saw similar patterns at V2 and V3, however many findings were no longer statistically significant. The only associations that were significant at all three timepoints were: compared to LA-HM, HA-HM was significantly associated with worse LDL, worse triglycerides, and worse HOMA-IR, and HA-LM was significantly associated with worse HOMA-IR.

Multiple linear regression models were also conducted with the complete cases at each time visit (Table B1 in Appendix B). We found that the significance of the phenotypes' association with cardiometabolic risks remained similar, before and after MICE. Thus, we concluded that the imputed values were plausible as it did not temper the association of the phenotypes and cardiometabolic risks.

Table 10: Pooled MLR results based on imputed dataset ($m = 10, n = 630$)

	Phenotype Classification	Total cholesterol[1,2]	HDL[1,2]	LDL[1,2]	Triglycerides[1,2]	Glucose[1,2]	HOMA-IR[1,2]
V1	HA-HM	0.15 (0.08)	**-0.16 (0.03)**	**0.17 (0.07)**	**0.30 (0.05)**	0.03 (0.04)	**0.76 (0.08)**
	HA-LM	0.11 (0.11)	**-0.07 (0.04)**	0.13 (0.09)	**0.13 (0.06)**	0.03 (0.05)	**0.31 (0.09)**
	LA-LM	-0.06 (0.08)	0.01 (0.03)	-0.05 (0.07)	-0.05 (0.05)	**-0.09 (0.04)**	-0.08 (0.07)
	LA-HM	Reference	Reference	Reference	Reference	Reference	Reference
V2	HA-HM	0.14 (0.09)	**-0.15 (0.03)**	0.15 (0.08)	**0.31 (0.06)**	0.06 (0.05)	**1.06 (0.12)**
	HA-LM	0.10 (0.10)	-0.06 (0.04)	0.10 (0.09)	**0.16 (0.07)**	0.03 (0.05)	**0.45 (0.13)**
	LA-LM	-0.04 (0.09)	0.02 (0.03)	-0.07 (0.08)	0.02 (0.06)	-0.05 (0.05)	-0.14 (0.11)
	LA-HM	Reference	Reference	Reference	Reference	Reference	Reference
V3	HA-HM	0.17 (0.10)	-0.06 (0.04)	**0.16 (0.08)**	**0.17 (0.06)**	0.10 (0.06)	**0.96 (0.14)**
	HA-LM	0.09 (0.11)	0.01 (0.04)	0.05 (0.09)	0.06 (0.07)	0.08 (0.08)	**0.57 (0.14)**
	LA-LM	0.06 (0.10)	0.03 (0.03)	0.03 (0.07)	-0.002 (0.06)	-0.12 (0.06)	-0.05 (0.13)
	LA-HM	Reference	Reference	Reference	Reference	Reference	Reference

[1]Boldface indicates the phenotype had a p-value below 0.05.
[2]Models adjusted for age, sex, and Tanner's stage.

4 Discussion

This study successfully developed ASMI and FMI age- and sex-adjusted reference curves with the LMS method for children aged 8-17 years old using NHANES data from 1999-2006. We extended Prado et al.'s work and identified best edf settings based on balancing between model fit statistics and retaining critical features of children's ASMI and FMI progression. In addition, we also tested the reference curves' significance with cardiometabolic risks available in the QUALITY cohort study. Before testing the utility of our reference curves, we discovered that there were almost 50% missing variables in the study due to loss to follow up at the second and third time points. Multiple imputation was performed using MICE and through the diagnostic tools of this algorithm, we confirmed that the imputed values were appropriate and plausible. After the imputation, we then performed pooled multiple linear regression for cross-sectional cardiometabolic risks after adjusting for age, sex, and Tanner's stage.

In building the reference curves, the identification of appropriate edf for each curve is a critical step. Choosing the edf to smooth a curve has been a subjective exercise or even a black art as mentioned by Cole [24]. In our study, we further strengthened the reasoning behind our chosen edf by showing that the classification between the maximized edf and ours did not alter the distribution. In other words, we have a small misclassification rate compared to a larger edf setting. In a large representative dataset, the small differences in phenotype classification is not likely to be substantial since the misclassification rate is calculated by dividing the number of misclassified participants with the sample size. Thus, this diagnostic that we performed is generalizable towards a larger data set.

Our reference curves all have a positive slope, with some ages showing a steeper slope. Males have a steeper slope in ASMI compared to females, and males' FMI has a significant feature that didn't appear in females: a dip at age 15. For males, puberty begins later and thus it might explain the dip that is a feature of the data. This means that during that age, our results showed that male participants experienced a decrease in fat mass. In addition, they also experienced a significant increase in muscle mass from age 12 to 13 years. Both of these patterns were not observed in females' curves.

In QUALITY, almost all characteristics except age, parents' education, and phenotypes were different for males and females at V1 and V3, but were not different at V2. Similarly, some of the cardiometabolic risks that were significant in V1 were not significant at V2, possibly due to the timing of V2 occurring around the start of puberty (10-14 years). At V3 when the participants had surpassed their pubertal age (15-19 years), we saw that those risks began to differ significantly again. The phenotype classifications were indifferent to males and females where we observed the highest proportion in LA-LM, followed by HA-HM, then either LA-HM or HA-LM, depending on the time point. At V2, we observed that more participants were classified as HA-LM than LA-HM, whereas at V1 and V3, more were in LA-HM. This increase could be explained by the fact that the participants were going through puberty from V1 to V2 and thus created an increase in the HA-LM group.

Due to the loss to follow-up in the study, we investigated the characteristics between participants who were retained and those lost to follow-up. Household income was significantly greater for those who were retained compared to those who exited the study at V2. We found that 39.32% of those who were retained were in HA and those who were lost to follow-up had a larger proportion at 45.49%. Thus, adiposity was higher among participants who were lost to follow-up than among those who stayed. However, with MICE, we were able to impute those who were lost to follow-up, and any other missing variables. The algorithm only had $k < 6$ NAs per variable in V1, which resulted in the density plot looking extreme since it only showed the k imputed values. For V2 and V3, it looked very cohesive and similar since we had more samples to be imputed. Furthermore, by comparing the observed distribution's descriptive statistics to the imputed one, we ensured that the imputed values were plausible and that they behaved similarly.

Nevertheless, some of the characteristics were now significantly different between males and females after the imputation. For instance, age at V2 was statistically different between males and females (mean of 11.72 and 11.57, respectively), rather than in the pre-MICE data (mean of 11.59 and 11.67) which was not statistically significant. However, it should be pointed out that although the age post-MICE is now *statistically* significant, the change in the mean was small; 0.13 for males and 0.10 for females. This translates to a mean age that has shifted by roughly +1.56 months for males and -1.20 for females. Although the numbers appeared to be statistically significant, it likely would not have a real-world significance. Most of the cardiometabolic risks boundaries for paediatric are often categorized based on age in years [40]. Thus, a small change of roughly ± 1 months would not likely largely impact the analysis.

In contrast, the phenotype classification distribution was similar before and after the imputation. QUALITY participants were mostly classified as LA-LM, followed by HA-HM, then LA-HM, and HA-LM at V1 and V3. At V2, we found that HA-LM had a larger proportion than LA-HM. Again, the increase in HA-LM from V1 to V2 can be explained by the fact that the participants were going into their pubertal age and metabolic changes. This means that most of the participants were categorized as having low adiposity and low muscle (more than 40% at all time points).

We found that for ASMI and FMI, the distributions were heavily right-skewed and that the distribution for children and adult were similar in range. Thus, our curves were based on those overall high values of ASMI and FMI, which yields high 50^{th} decile values. Since the NHANES-derived cutoffs are high for the participants, most of the QUALITY cohort participants failed to surpass the 50^{th} percentile, leading them to be categorized as LA-LM.

In addition to that, the high proportion in LA-LM can also be explained by the fact that we merged our decile values with Prado's [12]. Although we managed to maximize our sample size by including 18-19 years old at V3 (total sample size of $n = 630$, instead of $n = 429$), it can be seen that the jump in females and males' FMI from our decile values to Prado's was substantial. This creates a slight disconnect between the median of 17-year-olds and 18-year-olds while also creating a higher cutoff for the participants that were 18-19 years in V3. Due to this, 35/201 participants were classified into LA-LM at V3 even though we observed a stable trend for their FMI and ASMI throughout V1 to V2. For future studies, it's best to create a complete reference curve based on the full age range of the study sample to avoid such issues.

To date, this is the first study to develop and test children's DXA-derived reference curves' significance by looking at its association with cardiometabolic risks. We extended Prado's research by (1) creating children's reference curves using a large, representative dataset of the general US population (NHANES) and (2) testing the significance with QUALITY [12]. Pooled regression estimates at all time points showed that most of the phenotypes didn't have a significant association with the cardiometabolic risks. High adiposity class regardless of the muscle level (which includes HA-LM which is used to characterize sarcopenic obesity) was associated with worse cardiometabolic risks (HOMA-IR, HDL, LDL and triglycerides at baseline), and is consistent with the literature [4], [11], [41].

The overall weak association of our phenotypes might be due to the fact that the phenotypes were built based on a simple 50th percentile cutoff method. This method might be oversimplified and we should account for more conditions of classification. For example, consider two females aged 10 years old, A (FMI = 5.60, ASMI = 5.43) and B (FMI = 5.62, ASMI = 5.50). By our decile cutoffs in Appendix A, person A is in the LA-HM group and B is in the LA-LM group. However, the values of their FMI and ASMI are not too far apart. By our phenotype classification, person B is likely to be at greater risk for worse outcomes than A. Further work is needed to classify phenotypes based on clinically meaningful thresholds connected with health outcomes. For example, we could create more levels inside the four phenotypes to describe the extremities of how far the measurements are from the cutoffs. For instance, let the measurement be x. We can create more specific cutoffs by making some classes of severity:

$$\text{Severity Classification} = \begin{cases} \text{Class I}, & d50 < x \leq d60 \\ \text{Class II}, & d60 < x \leq d70 \\ \text{Class III}, & d70 < x \leq d80 \\ \text{Class IV}, & d80 < x \leq d90 \\ \text{Class V}, & x > d90, \end{cases}$$

where d50 stands for the 50th decile, d60 stands for the 60th, and so on. Therefore, while two participants can be in the same group together, they will have a different class of severity based on where their measurements would fall in the deciles range; this creates more intricate and precise phenotypes.

Moreover, we only tested the phenotypes with cardiometabolic risks such as total cholesterol, glucose, HDL, LDL, triglycerides, and HOMA-IR. These cardiometabolic risks were chosen since they are central indicators of health and also biomarkers of obesity [42]. However, since our study sample was coming from a young population aged 8-17 years old, these biomarkers might not be as apparent as they are in adults. Thus, it would be better to assess the phenotypes' significance with other measures that have a more direct significance for children such as fitness assessment and VO2 max (maximal oxygen uptake) [43], [44].

This study has several limitations. Other methods besides the LMS should be considered while creating the reference curves as our data showed a large kurtosis. For instance, since kurtosis can't be modeled with the LMS method, the Box-Cox power exponential (BCPE) transformation as developed by Rigby and Stasinopoulos could be tested [45]. The distribution is defined as Y^{ν} (it was Y for regular LMS) and they introduced a shifted and scaled power exponential transformation,

τ. Therefore, this method which is called LMSP has one extra parameter to control the kurtosis in the fitted distribution.

MICE was performed with only 10 imputations due to limited computing power. However, since our data is not that large ($n = 630$), the literature indicates that it is unnecessary to use more than 10 imputations since MICE doesn't need as much iteration to converge [37]. Similarly, Rubin indicates that 3-10 imputations are typically sufficient for the size of our data [36]. As MICE uses marginal density to impute the values, including variables such as weight, or BMI can further improve the imputation as they are related to the measures of interest. Furthermore, multilevel MICE should be considered to perform longitudinal imputation since we only viewed QUALITY as cross-sectional data.

In addition, other methods of imputations such as Matrix Completion [46] and Markov Chain Monte Carlo (MCMC) can be used for comparison of the imputed values. Matrix completion is a method that estimates missing data based on the low rank matrix structure of the observed [46]. To precisely estimate the missing entries, the approach takes advantage of the underlying matrix's low-rank structure or sparsity assumption [46], [47]. However, when there are many variables that need to be imputed, the low-rank assumption might not hold, making it inappropriate to use conventional matrix completion techniques that depends on low-rank structure. As there were many variables in QUALITY that needed to be imputed, the matrix structure will be large. Thus, even though matrix completion is known to perform well, we made the analytic decision not to use it in this study.

The construction of the reference curves was based on NHANES, a representative sample of the US population. In contrast, we tested the significance of the reference curves with QUALITY, a Canadian study sample where one of the inclusion criteria was that the children had to have at least one obese parent. Hence, our analysis results are not generalizable toward a different, healthier population. In addition, although they share a lot of similarities, the US and Canadian population are notably different in several aspects. For instance, the United States' race/ethnicity has a different composition from Canada, with larger Hispanic, Black, and other visible minorities in comparison to Canada [42],[43]. In addition, healthcare access in the US varies depending on employment and insurance coverage, while in Canada it is provided primarily through a universal healthcare system [50]. Future work should consider adding a representative Canadian population while building the curves so that the results can be more widely applicable towards a Canadian population.

Lastly, since we only viewed QUALITY as cross-sectional data, we did not perform any longitudinal analysis to test the DXA-phenotypes' predictive power. Future research should focus more on the longitudinal aspect of the data in order to test the predictive power of the phenotypes.

5 Conclusion

Currently, there are no ASMI and FMI reference curves for children, there are only BMI guidelines. However, BMI often leads to misclassification as it does not distinguish between fat mass and muscle mass. In our study, we utilized the LMS method to develop ASMI and FMI reference curves for children aged 8-17 years old using NHANES, a large cross-sectional study sample representative of the general US population. Our reference curves possess future clinical significance as they can potentially be used as a reliable tool to assess obesity in children, especially sarcopenic obesity. While testing the utility of the curves with QUALITY data, it was found that overall, the phenotypes had weak association with cardiometabolic risks. This may be further improved by testing the phenotypes with other outcomes that might have more direct significance for a young population, such as fitness level of the children.

From the pooled multiple linear regression results, we found that HA-HM and HA-LM were significantly associated with worse HDL, LDL, triglycerides, and HOMA-IR at baseline compared to the reference group LA-HM. These patterns were consistent at V2 and V3, although some associations were no longer statistically significant.

Although our results indicate that there are statistically significant associations between the phenotypes and cardiometabolic risks, the magnitude of the associations were small. Identification of ASMI and FMI severity would likely better represent health risks than a median split adjusted for age and sex. Further improvements of the reference curves and phenotype classification are needed.

REFERENCES

[1] "Obesity and overweight." https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (accessed Mar. 31, 2023).

[2] Public Health Agency of Canada, "Tackling obesity in Canada: Childhood obesity and excess weight rates in Canada," Jan. 31, 2018. https://www.canada.ca/en/public-health/services/publications/healthy-living/obesity-excess-weight-rates-canadian-children.html (accessed Mar. 31, 2023).

[3] "Study Quality Assessment Tools | NHLBI, NIH." https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools (accessed Mar. 31, 2023).

[4] S. Purcell *et al.*, "Prevalence of Sarcopenic Obesity Using Different Definitions and the Relationship With Strength and Physical Performance in the Canadian Longitudinal Study of Aging," *Front. Physiol.*, vol. 11, Jan. 2021, doi: 10.3389/fphys.2020.583825.

[5] K. K. Panuganti, M. Nguyen, and R. K. Kshirsagar, "Obesity," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2023. Accessed: Mar. 31, 2023. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK459357/

[6] V. Lee *et al.*, "Estimation of visceral fat in 9- to 13-year-old girls using dual-energy X-ray absorptiometry (DXA) and anthropometry," *Obes. Sci. Pract.*, vol. 4, no. 5, pp. 437–447, Sep. 2018, doi: 10.1002/osp4.297.

[7] University of Bristol, "Media resources." http://www.bristol.ac.uk/alspac/media/ (accessed May 11, 2023).

[8] T. B. VanItallie, M. U. Yang, S. B. Heymsfield, R. C. Funk, and R. A. Boileau, "Height-normalized indices of the body's fat-free mass and fat mass: potentially useful indicators of nutritional status," *Am. J. Clin. Nutr.*, vol. 52, no. 6, pp. 953–959, Dec. 1990, doi: 10.1093/ajcn/52.6.953.

[9] I. H. Rosenberg, "Epidemiologic and Methodologic Problems in Determining Nutritional Status of Older Persons: Summary comments," *Am. J. Clin. Nutr.*, vol. 50, no. 5, pp. 1231–1233, Nov. 1989, doi: 10.1093/ajcn/50.5.1231.

[10] R. N. Baumgartner, "Body Composition in Healthy Aging," *Ann. N. Y. Acad. Sci.*, vol. 904, no. 1, pp. 437–448, Jan. 2006, doi: 10.1111/j.1749-6632.2000.tb06498.x.

[11] E. Poggiogalle *et al.*, "Sarcopenic obesity and insulin resistance: Application of novel body composition models," *Nutr. Burbank Los Angel. Cty. Calif*, vol. 75–76, p. 110765, 2020, doi: 10.1016/j.nut.2020.110765.

[12] C. M. Prado *et al.*, "A population-based approach to define body-composition phenotypes," *Am. J. Clin. Nutr.*, vol. 99, no. 6, pp. 1369–1377, Jun. 2014, doi: 10.3945/ajcn.113.078576.

[13] T. J. Cole, "The LMS method for constructing normalized growth standards," *Eur. J. Clin. Nutr.*, vol. 44, no. 1, pp. 45–60, Jan. 1990.

[14] "Growth Charts - Homepage," Apr. 04, 2023. https://www.cdc.gov/growthcharts/index.htm (accessed May 06, 2023).

[15] "Growth Charts - CDC Extended BMI-for-Age Growth Charts - Download," Dec. 21, 2022. https://www.cdc.gov/growthcharts/Extended-BMI-Charts.html (accessed May 06, 2023).

[16] T. J. Cole, "Using the LMS method to measure skewness in the NCHS and Dutch National height standards," *Ann. Hum. Biol.*, vol. 16, no. 5, pp. 407–419, Jan. 1989, doi: 10.1080/03014468900000532.

[17] WHO multicentre Growth Reference Study Group, "WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development." Geneva: World Health Organization, 2006.

[18] M.-X. Lun *et al.*, “Application of the LMS method of constructing fetal reference charts: comparison with the original method,” *J. Matern. Fetal Neonatal Med.*, vol. 34, no. 3, pp. 395–402, Feb. 2021, doi: 10.1080/14767058.2019.1608942.
[19] “NHANES Questionnaires, Datasets, and Related Documentation.” https://wwwn.cdc.gov/nchs/nhanes/default.aspx (accessed Feb. 13, 2023).
[20] “NHANES - NCHS Research Ethics Review Board Approval,” Aug. 25, 2022. https://www.cdc.gov/nchs/nhanes/irba98.htm (accessed Feb. 21, 2023).
[21] “NHANES 1999-2006 DXA Multiple Imputation Data Files.” https://wwwn.cdc.gov/Nchs/Nhanes/Dxa/Dxa.aspx (accessed Feb. 13, 2023).
[22] Centers for Disease Control and Prevention, “Anthropometry Procedures Manual.” National Health and Nutrition Examination Survey, 2007. Accessed: Feb. 20, 2023. [Online]. Available: https://www.cdc.gov/nchs/data/nhanes/nhanes_07_08/manual_an.pdf
[23] Centers for Disease Control and Prevention, “National Health and Nutrition Examination Survey: Technical Documentation for the 1999-2004 Dual Energy X-Ray Absorptiometry (DXA) Multiple Imputation Data files.” National Health and Nutrition Examination Survey. Accessed: Feb. 20, 2023. [Online]. Available: https://wwwn.cdc.gov/nchs/data/nhanes/dxa/dxa_techdoc.pdf
[24] T. J. Cole and P. J. Green, “Smoothing reference centile curves: The lms method and penalized likelihood,” *Stat. Med.*, vol. 11, no. 10, pp. 1305–1319, 1992, doi: 10.1002/sim.4780111005.
[25] “LMSchartmaker Pro | Health for all Children.” https://www.healthforallchildren.com/shop-base/shop/software/lmschartmaker-pro/ (accessed Feb. 13, 2023).
[26] G. E. P. Box and D. R. Cox, “An Analysis of Transformations,” *J. R. Stat. Soc. Ser. B Methodol.*, vol. 26, no. 2, pp. 211–252, 1964.
[27] S. van Buuren and M. Fredriks, “Worm plot: a simple diagnostic device for modelling growth reference curves,” *Stat. Med.*, vol. 20, no. 8, pp. 1259–1277, Apr. 2001, doi: 10.1002/sim.746.
[28] H. Pan and T. J. Cole, “A comparison of goodness of fit tests for age-related reference ranges,” *Stat. Med.*, vol. 23, no. 11, pp. 1749–1765, Jun. 2004, doi: 10.1002/sim.1692.
[29] P. Royston and E. M. Wright, “Goodness-of-fit statistics for age-specific reference intervals,” *Stat. Med.*, vol. 19, no. 21, pp. 2943–2962, Nov. 2000, doi: 10.1002/1097-0258(20001115)19:21<2943::aid-sim559>3.0.co;2-5.
[30] M. Lambert *et al.*, “Cohort Profile: The Quebec Adipose and Lifestyle Investigation in Youth Cohort,” *Int. J. Epidemiol.*, vol. 41, no. 6, pp. 1533–1544, Dec. 2012, doi: 10.1093/ije/dyr111.
[31] T. Lohman, A. Roche, and R. Martorell, *Anthropometric Standardization Reference Manual*. Champaign, IL: Human Kinetics, 1988.
[32] W. A. Marshall and J. M. Tanner, “Variations in pattern of pubertal changes in girls.,” *Arch. Dis. Child.*, vol. 44, no. 235, pp. 291–303, Jun. 1969.
[33] W. A. Marshall and J. M. Tanner, “Variations in the pattern of pubertal changes in boys,” *Arch. Dis. Child.*, vol. 45, pp. 13–23, 1970, doi: 10.1136/adc.45.239.13.
[34] M. Henderson, “The QUebec Adipose and Lifestyle InvesTigation in Youth (QUALITY) Cohort,” clinicaltrials.gov, Clinical trial registration NCT03356262, Nov. 2022. Accessed: Feb. 21, 2023. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT03356262
[35] D. R. Matthews, J. P. Hosker, A. S. Rudenski, B. A. Naylor, D. F. Treacher, and R. C. Turner, “Homeostasis model assessment: insulin resistance and beta-cell function from

fasting plasma glucose and insulin concentrations in man," *Diabetologia*, vol. 28, no. 7, pp. 412–419, Jul. 1985, doi: 10.1007/BF00280883.
[36] D. B. Rubin, *Multiple imputation for nonresponse in surveys*. in Wiley series in probability and mathematical statistics. New York: Wiley, 1987.
[37] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate Imputation by Chained Equations in R," *J. Stat. Softw.*, vol. 45, pp. 1–67, Dec. 2011, doi: 10.18637/jss.v045.i03.
[38] S. van Buuren, "Flexible Imputation of Missing Data".
[39] T. Raghunathan and I. Bondarenko, "Diagnostics for Multiple Imputations." Rochester, NY, Nov. 21, 2007. doi: 10.2139/ssrn.1031750.
[40] "Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents: Summary Report," *Pediatrics*, vol. 128, no. Suppl 5, pp. S213–S256, Dec. 2011, doi: 10.1542/peds.2009-2107C.
[41] K. Agarwal, V. Chorsiya, D. Kaushik, and A. Yadav, "Impact of Sarcopenic obesity on body composition, physical performance and fall risk in community dwelling older adults," *Sci. Talks*, vol. 4, p. 100074, Dec. 2022, doi: 10.1016/j.sctalk.2022.100074.
[42] E. S. Lau *et al.*, "Cardiovascular Biomarkers of Obesity and Overlap With Cardiometabolic Dysfunction," *J. Am. Heart Assoc.*, vol. 10, no. 14, p. e020215, Jul. 2021, doi: 10.1161/JAHA.120.020215.
[43] R. Pate, M. Oria, L. Pillsbury, C. on F. M. and H. O. in Youth, F. and N. Board, and I. of Medicine, "Measuring Fitness in Youth," in *Fitness Measures and Health Outcomes in Youth*, National Academies Press (US), 2012. Accessed: May 06, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK241311/
[44] V. Carayanni *et al.*, "Predicting VO2max in Children and Adolescents Aged between 6 and 17 Using Physiological Characteristics and Participation in Sport Activities: A Cross-Sectional Study Comparing Different Regression Models Stratified by Gender," *Children*, vol. 9, no. 12, p. 1935, Dec. 2022, doi: 10.3390/children9121935.
[45] R. A. Rigby and D. M. Stasinopoulos, "Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution," *Stat. Med.*, vol. 23, no. 19, pp. 3053–3076, 2004, doi: 10.1002/sim.1861.
[46] F. Sohil, M. U. Sohali, and J. Shabbir, "An introduction to statistical learning with applications in R: by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, New York, Springer Science and Business Media, 2013, $41.98, eISBN: 978-1-4614-7137-7," *Stat. Theory Relat. Fields*, vol. 6, no. 1, pp. 87–87, Jan. 2022, doi: 10.1080/24754269.2021.1980261.
[47] M. Cho, "Imputation of Missing Values by Low Rank Matrix Approximation." U.S. Bureau of Labor Statistics, 2021. Accessed: May 05, 2023. [Online]. Available: https://www.bls.gov/osmr/research-papers/2021/st210010.htm
[48] S. C. Government of Canada, "The Daily — The Canadian census: A rich portrait of the country's religious and ethnocultural diversity," Oct. 26, 2022. https://www150.statcan.gc.ca/n1/daily-quotidien/221026/dq221026b-eng.htm (accessed May 06, 2023).
[49] "Canada vs. United States - demographics comparison." https://www.indexmundi.com/factbook/compare/canada.united-states/demographics (accessed May 06, 2023).
[50] G. Ridic, S. Gleason, and O. Ridic, "Comparisons of Health Care Systems in the United States, Germany and Canada," *Mater. Socio-Medica*, vol. 24, no. 2, pp. 112–120, 2012, doi: 10.5455/msm.2012.24.112-120.

Appendices

Appendix A: L, M, S, and decile values

Table A1: Appendicular skeletal mass index (kg/m^2) decile values for females

				Appendicular skeletal mass index (kg/m^2) deciles								
Age (years)	**L**	**M**	**S**	**10^{th}**	**20^{th}**	**30^{th}**	**40^{th}**	**50^{th}**	**60^{th}**	**70^{th}**	**80^{th}**	**90^{th}**
8	-1.460	4.793	0.155	4.026	4.254	4.439	4.614	4.793	4.991	5.226	5.540	6.060
9	-0.694	5.159	0.154	4.287	4.556	4.769	4.964	5.159	5.368	5.607	5.911	6.382
10	-0.511	5.489	0.154	4.548	4.841	5.071	5.281	5.489	5.709	5.961	6.276	6.758
11	-0.880	5.751	0.154	4.794	5.087	5.320	5.535	5.751	5.984	6.254	6.599	7.143
12	-1.082	5.962	0.154	4.985	5.281	5.518	5.739	5.962	6.205	6.488	6.856	7.445
13	-1.066	6.109	0.154	5.106	5.410	5.653	5.880	6.109	6.358	6.648	7.026	7.628
14	-0.900	6.197	0.155	5.163	5.479	5.730	5.963	6.197	6.450	6.742	7.118	7.711
15	-0.842	6.236	0.156	5.184	5.506	5.762	5.998	6.236	6.492	6.788	7.167	7.762
16	-1.160	6.228	0.158	5.193	5.504	5.755	5.989	6.228	6.488	6.794	7.195	7.843
17	-1.737	6.194	0.160	5.199	5.490	5.728	5.957	6.194	6.460	6.781	7.219	7.977
18	-0.733	6.487	0.172	5.302	5.664	5.952	6.219	6.487	6.776	7.111	7.540	8.214
19	-0.728	6.508	0.173	5.316	5.680	5.970	6.238	6.508	6.799	7.136	7.568	8.247
Note: Values for age 18 and 19 are obtained from Prado et al.'s (2014) results.												

Table A2: Fat mass index (kg/m^2) decile values for females

				Fat mass index (kg/m^2) deciles								
Age (years)	**L**	**M**	**S**	**10th**	**20th**	**30th**	**40th**	**50th**	**60th**	**70th**	**80th**	**90th**
8	-0.529	5.322	0.385	3.432	3.945	4.392	4.838	5.322	5.883	6.589	7.595	9.432
9	-0.444	5.759	0.391	3.663	4.236	4.733	5.227	5.759	6.373	7.141	8.220	10.152
10	-0.379	6.092	0.396	3.829	4.450	4.988	5.520	6.092	6.748	7.563	8.700	10.704
11	-0.344	6.289	0.398	3.929	4.579	5.140	5.695	6.289	6.969	7.811	8.979	11.024
12	-0.337	6.495	0.397	4.059	4.731	5.310	5.883	6.495	7.195	8.061	9.261	11.355
13	-0.357	6.760	0.393	4.254	4.945	5.541	6.130	6.760	7.481	8.373	9.611	11.777
14	-0.401	6.966	0.386	4.438	5.134	5.735	6.329	6.966	7.696	8.601	9.861	12.081
15	-0.464	7.233	0.376	4.684	5.385	5.990	6.590	7.233	7.973	8.894	10.183	12.474
16	-0.542	7.482	0.365	4.932	5.632	6.237	6.837	7.482	8.226	9.156	10.467	12.825
17	-0.625	7.618	0.354	5.111	5.798	6.392	6.982	7.618	8.354	9.278	10.590	12.981
18	-0.196	8.658	0.419	5.195	6.155	6.981	7.794	8.658	9.639	10.840	12.480	15.275
19	-0.185	8.760	0.419	5.253	6.227	7.064	7.887	8.760	9.751	10.961	12.611	15.414
Note: Values for age 18 and 19 are obtained from Prado et al.'s (2014) results.												

Table A3: Appendicular skeletal mass index (kg/m^2) decile values for males

				Appendicular skeletal mass index (kg/m^2) deciles								
Age (years)	**L**	**M**	**S**	**10th**	**20th**	**30th**	**40th**	**50th**	**60th**	**70th**	**80th**	**90th**
8	0.297	5.337	0.136	4.462	4.750	4.966	5.156	5.337	5.524	5.728	5.975	6.328
9	-0.748	5.458	0.133	4.652	4.904	5.101	5.280	5.458	5.647	5.862	6.133	6.546
10	-1.109	5.678	0.130	4.873	5.121	5.317	5.497	5.678	5.872	6.095	6.380	6.827
11	-0.902	5.989	0.136	5.093	5.370	5.589	5.789	5.989	6.203	6.448	6.759	7.241
12	-0.603	6.361	0.150	5.304	5.633	5.891	6.127	6.361	6.611	6.895	7.254	7.804
13	-0.396	6.847	0.173	5.538	5.944	6.264	6.556	6.847	7.156	7.509	7.953	8.633
14	-0.558	7.340	0.158	6.056	6.455	6.768	7.055	7.340	7.644	7.991	8.430	9.104
15	-0.891	7.729	0.144	6.514	6.888	7.184	7.456	7.729	8.021	8.357	8.786	9.456
16	-1.028	7.979	0.148	6.711	7.098	7.406	7.691	7.979	8.290	8.650	9.115	9.850
17	-1.093	8.130	0.146	6.860	7.246	7.555	7.841	8.130	8.442	8.804	9.273	10.016
18	0.288	8.329	0.146	6.888	7.361	7.716	8.029	8.329	8.637	8.976	9.384	9.971
19	0.314	8.360	0.146	6.918	7.391	7.746	8.059	8.360	8.668	9.007	9.416	10.003
Note: Values for age 18 and 19 are obtained from Prado et al.'s (2014) results.												

Table A4: Fat mass index (kg/m^2) decile values for males

				Fat mass index (kg/m^2) deciles								
Age (years)	**L**	**M**	**S**	**10th**	**20th**	**30th**	**40th**	**50th**	**60th**	**70th**	**80th**	**90th**
8	-0.789	4.539	0.382	3.001	3.408	3.769	4.135	4.539	5.020	5.645	6.576	8.426
9	-0.682	4.748	0.400	3.058	3.506	3.902	4.305	4.748	5.275	5.956	6.964	8.930
10	-0.598	4.978	0.419	3.126	3.616	4.050	4.492	4.978	5.555	6.299	7.394	9.506
11	-0.556	5.097	0.434	3.139	3.654	4.113	4.581	5.097	5.709	6.501	7.668	9.917
12	-0.558	5.066	0.444	3.090	3.608	4.070	4.543	5.066	5.691	6.502	7.705	10.049
13	-0.597	4.961	0.447	3.032	3.534	3.984	4.446	4.961	5.578	6.385	7.594	9.996
14	-0.643	4.786	0.445	2.943	3.421	3.850	4.292	4.786	5.381	6.164	7.349	9.747
15	-0.667	4.679	0.444	2.889	3.352	3.769	4.198	4.679	5.259	6.026	7.191	9.571
16	-0.669	4.760	0.442	2.944	3.414	3.837	4.273	4.760	5.348	6.125	7.303	9.707
17	-0.657	4.815	0.439	2.984	3.459	3.886	4.325	4.815	5.404	6.179	7.349	9.713
18	-0.0003	5.531	0.464	3.053	3.744	4.337	4.918	5.531	6.221	7.054	8.172	10.022
19	0.008	5.646	0.457	3.137	3.839	4.441	5.028	5.646	6.339	7.174	8.292	10.133
Note: Values for age 18 and 19 are obtained from Prado et al.'s (2014) results.												

Appendix B

Table B1: MLR results based on complete dataset without imputation

	Phenotype Classification	**Total cholesterol**[1,2]	**HDL**[1,2]	**LDL**[1,2]	**Triglycerides**[1,2]	**Glucose**[1,2]	**HOMA-IR**[1,2]
V1 ($n = 611$)	HA-HM	0.15 (0.09)	**-0.17 (0.03)**	**0.17 (0.07)**	**0.31 (0.05)**	0.03 (0.04)	**0.76 (0.08)**
	HA-LM	0.11 (0.11)	**-0.07 (0.04)**	0.13 (0.09)	**0.13 (0.06)**	0.03 (0.05)	**0.32 (0.09)**
	LA-LM	-0.06 (0.08)	0.01 (0.03)	-0.05 (0.07)	-0.05 (0.05)	**-0.10 (0.04)**	-0.08 (0.07)
	LA-HM	Reference	Reference	Reference	Reference	Reference	Reference
V2 ($n = 334$)	HA-HM	0.18 (0.11)	**-0.20 (0.05)**	**0.21 (0.10)**	**0.38 (0.07)**	0.02 (0.06)	**0.87 (0.14)**
	HA-LM	0.12 (0.12)	-0.08 (0.05)	0.12 (0.10)	**0.17 (0.08)**	-0.02 (0.06)	**0.32 (0.14)**
	LA-LM	-0.02 (0.11)	0.02 (0.04)	-0.05 (0.09)	0.02 (0.07)	-0.09 (0.06)	-0.20 (0.13)
	LA-HM	Reference	Reference	Reference	Reference	Reference	Reference
V3 ($n = 278$)	HA-HM	0.19 (0.12)	**-0.12 (0.04)**	0.18 (0.10)	**0.29 (0.07)**	0.12 (0.07)	**1.08 (0.16)**
	HA-LM	0.22 (0.14)	0.01 (0.05)	0.13 (0.11)	0.16 (0.09)	0.09 (0.08)	**0.66 (0.18)**
	LA-LM	0.12 (0.12)	0.05 (0.04)	0.06 (0.09)	0.02 (0.07)	-0.07 (0.06)	-0.05 (0.15)
	LA-HM	Reference	Reference	Reference	Reference	Reference	Reference

[1]Boldface indicates the phenotype had a p-value below 0.05.
[2]Models adjusted for age, sex, and Tanner's stage.