Summer 8-8-2017

Clinical Considerations for Preterm Infant Growth Curves Regarding Distributions and Race

Samantha Wotiz

ABSTRACT

CLINICAL CONSIDERATIONS FOR PRETERM INFANT GROWTH CURVES
REGARDING DISTRIBUTIONS AND RACE

By

SAMANTHA LINDSEY WOTIZ

JULY 12, 2017

Clinicians use growth curves to assess infant health. Most children are measured on growth curves that contain percentiles for height, weight, and head circumference by sex. Preterm infants have their own growth curves. Infants who present with measurements below the 10th percentile are considered small-for-gestational age (SGA), and infants who present with measurements above the 90th percentile are considered large-for-gestational age (LGA). Growth curves and centiles can be generated using 3 and 4 parameter distribution models. To date, no studies have been published to investigate whether growth curves generated using a 3- or 4-parameter model differ significantly. Additionally, researchers have found mixed results when exploring the association between race and pregnancy/delivery. Black mothers may have greater risks and babies with lower weights than babies born to White mothers (Borrell, Rodriguez-Alvarez, Savitz, & Baquero, 2016), and growth curves that do not consider race may misclassify non-White babies (Buck-Louis et al., 2015). In this study, I had two specific aims: (1) to compare the preterm infant growth curves and centiles generated using 3 and 4 parameter methods (Lamba Mu Sigma [LMS] and Box-Cox Power Exponential [BCPE], respectively) and assess each model for adequate fit, and (2) to use percentile cut points from race-specific and non-race-specific LMS curves to classify babies in a validation dataset as SGA or LGA. Regarding the differences in curves generated from the LMS and BCPE distributions, the curves produced using the BCPE distribution had a lower GAIC in some cases but model fit criteria for the LMS curves were adequate. The simpler models generated by the LMS method were retained for birth length, head circumference, and weight by sex with an explanatory variable of gestational age. For aim 2, results indicated that race-specific curves classified babies within expected ranges. Non-race-specific curves overidentified Black babies as SGA and underidentified them as LGA. More research is required to test if this relationship persists for babies delivered at full term.

CLINICAL CONSIDERATIONS FOR PRETERM INFANT GROWTH CURVES
REGARDING DISTRIBUTIONS AND RACE

By

SAMANTHA LINDSEY WOTIZ

B.S., KENNESAW STATE UNIVERSITY

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA
30303

APPROVAL PAGE

CLINICAL CONSIDERATIONS FOR PRETERM INFANT GROWTH CURVES
REGARDING DISTRIBUTIONS AND RACE

By

SAMANTHA LINDSEY WOTIZ

Approved:

Katherine Masyn
Committee Chair

Louise Lawson
Committee Member

July 19, 2017
Date

Dedication

I would like to dedicate this Master's Thesis to my father, who passed away before I finished my coursework. It is because of his belief in the merits of higher education that I continued mine.

Thanks, Dad

Acknowledgements

I would like to acknowledge the following mentors and friends who helped make this project

happen.

- Dr. Lawson: for taking time to help a former student and displaying unending patience with my barrage of questions every time we spoke (especially about model selection criteria). For providing project guidance and direction.

- Dr. Masyn: for supporting me at Georgia State and providing encouragement/critique.

- The Grow, Baby, Grow project, specifically Reese H. Clark, MD, Irene E. Olsen, PhD, RD, Nicole Ferguson, PhD, Adrian Williamson, PhD, and Josip Derado, PhD : for allowing me to use their datasets and expertise.

- The "Fluff Senders": for their relentless encouragement and loving support as I sought to balance motherhood and student duties.

- To my husband, Todd: for his commitment to providing for our family as I finished my studies.

Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Samantha Lindsey Wotiz
Signature of Author

TABLE OF CONTENTS

List of Tables

List of Figures

Figure 12: Growth Curves from BCPE distribution for Preterm Infant Females—Birth Length by

Gestational Age

Introduction

Clinicians use growth curves to assess infant health. Most children are measured on growth curves that contain percentiles for height-, weight-, and head circumference-for-age with separate curves for males and females. Preterm infants have their own growth curves. In the US, clinicians use the preterm infant growth curves created by Fenton and Kim (2013) or Olsen and colleagues (2010) for tracking growth of infants in a neonatal intensive care unit (NICU). Infants who present with measurements below the 10th percentile are considered small-for-gestational age (SGA), and infants who present with measurements above the 90th percentile are considered large-for-gestational age (LGA).

Growth curves and centiles can be generated using 3- and 4-parameter distribution models. The LMS method (Cole & Green, 1992) uses 3 parameters *lambda*, *mu*, and *sigma* to account for skewness, the median, and the coefficient of variation, respectively. A 4-parameter distribution that has been used in the creation of growth curves is the Box-Cox Power Exponential (BCPE) distribution (Rigby & Stasinopoulos, 2004). The researchers found that growth curves generated with the addition of a parameter to account for kurtosis in data had a lower GAIC than curves generated using a 3-parameter distribution. To date, there are no studies that have been published to investigate how growth curves generated using a 3- or 4-parameter model differ for the clinician who will use them.

Additionally, researchers have found mixed results when exploring the association between race and infant/child growth. Although the World Health Organization (WHO) created growth curves with a diverse sample from many countries which was meant to represent the ideal growth for all children (WHO, 2006), studies indicated a lack of fit for Asian infants (Tanaka et al., 2013; Huang et al., 2016; Dwipoerwantoro et al., 2015) and European

infants/children (Natala & Rajaopalan, 2015). When researchers studied the association of race

and pregnancy/delivery, they found that Black mothers may have greater risks and babies with

lower weights than babies born to White mothers (Borrell, Rodriguez-Alvarez, Savitz, &

Baquero, 2016). A rigorous study about race and estimated fetal weight (obtained through

ultrasonography) showed that intrauterine growth curves that do not consider race may

misclassify non-White babies (Buck Louis et al., 2015).

In this study, I had two specific aims: (1) to compare the preterm infant growth curves

and centiles generated using 3- and 4-parameter methods (LMS and BCPE, respectively) and

assess each model for adequate fit, and (2) to use percentile cut points from race-specific and

non-race-specific LMS curves to classify babies in a validation dataset as SGA or LGA. I chose

to evaluate the difference in the curves created with a 3- and 4-parameter distribution by

determining whether the expected versus observed percentiles that were generated from the

curves were acceptable. The acceptable deviation in observed and expected percentiles from the

WHO Child Growth Reference were 2 percentiles points (WHO, 2006). Because the WHO

curves are considered an authority I followed this method. I used the acceptable curves from aim

1 to produce cut points for aim 2. The cut points were generated from a "general" dataset (e.g. all

males) and from a race-specific subset of the dataset (Black males, White males, and all other

males). This process was repeated for female infants. I then considered whether the race-specific

cut points classified babies as SGA and LGA any differently than cut points from the "general"

non-race-specific data.

The data were provided by Pediatrix Medical Group, Inc. The datasets were de-identified

and contained birth information about infants admitted to NICUs between 1998 and 2006.

Infants who were not singletons, presented with health conditions that would affect growth, or had implausible measurements were excluded from analysis.

PAPER I: HOW CENTILE ESTIMATIONS AND MODEL FIT CHANGE BETWEEN LMS

AND BCPE CURVES

How Curves are Used Clinically

Growth curves are a useful tool for clinicians to assess infant health. Preterm infants (those who are born before 38 weeks gestational age) require close monitoring to ensure proper growth, feeding, and health is achieved. Ideally a growth chart package will have measures for body length (height) by age, weight by age, head circumference by age, and body mass index (BMI) or some other ratio of weight and length for age. Boys and girls have separate growth charts. Children who are found to be growing too slowly or presenting with extreme measures in any category are identified by clinicians to be investigated for health abnormalities and potentially to receive targeted treatment.

Clinicians sometimes use growth curves to categorize infants into high-risk groups. Babies whose measurements fall below the 10th percentile are considered small-for-gestational age (SGA), and babies whose measurements fall above the 90th percentile are considered large-for-gestational age (LGA). These (arbitrary) cut points correlate (may indicate) with health concerns such as perinatal asphyxia, meconium aspiration, hypoglycemia, and hypothermia for SGA (Gibson & Nawab, 2015a) and respiratory distress, meconium aspiration, hypoglycemia, and polycythemia for LGA (Gibson & Nawab, 2015b).

Infants who are born before 38 weeks should be measured using intrauterine growth curves. These are curves, which usually have a starting point of 23 weeks gestational age, are generated after collecting cross-sectional birth data. It is thought that by amassing the information from infants born at 23 weeks, infants who are born 24 weeks, 25 weeks, and so on that this will represent the actual growth of infants in-utero (hence the term intrauterine growth

curves). There are other methods of reporting preterm infant growth curves (such as through ultrasound measurements) but a discussion of this is not within the scope of this paper. After 37 weeks gestational age, infants may be measured using 0-2 year age charts using the "adjusted" age of the infant.

There are several growth charts available to clinicians. Growth charts can be categorized as growth standards (how children should grow) or growth references (how children actually grow) (Bhatia, 2013). In the US, the most utilized charts are The World Health Organization (WHO) Child Growth Standards (WHO, 2006) and the 2000 Centers for Disease Control (CDC) Growth Charts. The CDC recommends the WHO Child Growth Standards be used from the age of $0 - 2$ years for children in the United States (CDC, 2000), and the CDC curves for ages 2 to 20 years. Preterm infants in the US are usually measured on curves generated by Olsen and colleagues (Olsen, Groveman, Lawson, Clark, & Zemel, 2010) or Fenton (2003; Fenton & Kim, 2013). The decision of which growth chart to use is a combination of the age of the child, the food source (primarily breastfed versus formula fed for infants), and the preference of the clinician.

Multiple growth charts exist because of differences in samples characteristics and statistical methods used to generate growth curves. I will review the differences in sample selection for the most utilized growth charts followed by an explanation of statistical methods for curve generation in these studies. Relevant to the focus of this thesis, this review will be limited to the Olsen curves and Fenton curves for preterm infants.

The Olsen curves (Olsen et al., 2010) provide weight, length, and head circumference (by gestational age) for infants who are 23 to 41 weeks gestational age. The purpose of the study was to create updated growth curves from a racially diverse set of US infants that could be compared

to older curves. The sample consisted of 391,681 infants born between 1998 and 2006 in 248 US hospitals. Data were cross-sectional, collected at the time of the infant's birth, separated by sex, and divided into subsamples—a curve (creation) sample and a validation sample. The researchers used LMSChartMakerPro2.3 to create smoothed curves. The curves were then validated and compared to older Lubchenco curves by SGA and LGA classifications (Olsen et al., 2010).

The results of the study showed that gender-specific curves classified babies differently (and better—more babies were classified correctly as SGA and LGA with the Olsen curves) than the previously used unisex Luchenco curves. In comparison with the Lubchenco curves, Olsen and colleagues produced curves that generally had lower average weights, lengths, and head circumferences at younger gestational ages until 30 and 36 weeks (Olsen et al., 2010). Growth measurements were higher at older gestational ages. These results supported the introduction of gender-specific curves and curves created from US infants from a large, racially diverse sample.

Another group of commonly used NICU growth curves in the US were created by Fenton and colleagues (Fenton, 2003; Fenton & Kim, 2013). In the original research (2003), Fenton performed a meta-analysis by combining data from three published studies to create an updated growth-chart similar in style to Babson and Benda (1976). The researcher did not create gender-specific curves. She used manual methods to create a smooth boundary (with no deceleration from pre-term curves to 2 months post 40 weeks) with CDC Growth Charts (CDC, 2000). The resulting charts for length, head circumference, and weight were described as having "better confidence in extreme intervals" (Fenton, 2003, p.e9*)* than the Babson and Benda (1976) curves.

Fenton and Kim's (2013) work revised the 2003 Fenton Preterm Growth Chart described previously. Based on [whose past work?] past work, these? researchers believed that evidence

supported a smooth transition from growth charts of pre-term infants onto full-term infant growth charts. They sought to create updated curves that would utilize existing pre-term infant growth data that could be seamlessly mapped onto the World Health Organization (WHO) Growth Standard (WHO, 2006). Six studies of pre-term infants from Germany, the United States, Canada, Australia, Scotland, and Italy were included in the meta-analysis. Researchers used the LMS method to generate curves for boys and girls separately. The resulting charts created have an age range from 22 to 50 weeks gestational age. They allow for a seamless transition from pre-term growth to the WHO Growth Standards (Fenton & Kim, 2013).

Growth Curve Creation with LMS distribution

Both the Olsen curves (Olsen et al., 2010) and Fenton curves (Fenton & Kim, 2013) used the LMS method for growth curve creation. The LMS method with a maximum penalized likelihood was created by Cole and Green (1992). LMS stands for *Lambda, Mu,* and *Sigma* which are the names of the three parameters used to create the growth curves. It is a distribution that handles data with an underlying skew normal shape.

In general, for growth curve creation methods and centile estimation, each variable is considered either a *response variable* or an *explanatory variable*. The most common explanatory variable is age; for pre-term infants, gestational age is used. The typical response variables are birth weight, length, and head circumference. Some studies have examined body mass index (BMI) in relation to age (Olsen et al., 2014; WHO, 2006). Because growth curves are an extension of a regression analysis, a typical statistical assumption for growth curve creation is that the underlying conditional distribution of the response variables are normally distributed. If data are not normally distributed then this can sometimes be "fixed" by applying a power transformation to "normalize" the response variable. Researchers may choose to apply an

additional power transformation to the explanatory variable (usually age) because children and babies may experience periods of rapid growth (growth may appear to be exponential instead of linear). The goal is to produce distribution z scores that have a mean of 0 and a standard deviation of 1.

The LMS method for centile estimation applies a re-parameterization to correct the data distribution so that z scores can be calculated and valid (Idrayan, 2014). The *lambda* parameter is a Box-Cox power transformation (defines the skewness of the distribution), the *mu* parameter accounts for the median of the distribution, and *sigma* is the coefficient of variation (Cole and Green, 1992). Using a maximum penalized likelihood, computer software can generate estimates for lambda, sigma, and mu. A mathematical representation of the LMS model is

$$z = \frac{[y/M(t)]^{L(t)} - 1}{L(t)S(t)}, \ L(t) \neq 0, \tag{1}$$

where *t* = covariate of *y*, (i.e., usually age); where λ, μ, and σ are read off the smooth curves, *L(t)*, *M(t)*, and *S(t)*, respectively.

Smoothing describes how to represent the relationship between distribution parameters and the explanatory variable (age) (Ospina, 2012). The values that are generated for *L*, *M*, and *S* are referred to as effective degrees of freedom (EDF); they are the shape of the curves. As EDFs rise, the shape of the curve takes on more "bumps". Oversmoothing will not capture the changes of a growth curve (it is over simplified) and undersmoothing will show too many variations in growth (representing the fluctuations of the sample which are not true of the population, a.k.a. "noise"). The challenge for statisticians is finding the balance between modeling the noise of the data and the actual shape of the parameter (Cole & Green, 1992). Automated methods exist for managing smoothing. The process that is relevant for this paper is the maximum likelihood method.

LMS ChartMakerPro (Cole & Green, 1992) is a user-friendly program to automatically generate growth curves with the LMS method. A user imports his/her data into the program, answers several questions about the data, and the program produces LMS EDFs and measures of model f*it.* The program features an easy to use graphical interface without any need to manually input lines of code. Users can adjust EDFs by hand or use an auto-selection method. The point-and-click program is easy to learn, fast to use, and simple to read the output. Users can export the model EDFs, fit statistics, centiles, and z scores.

Once a researcher has the EDFs for the *L*, *M*, and *S* curves, then Z scores and percentiles are easy to calculate. The mathematical relationship between z scores and percentiles is:

$$C_{100\propto}(t) = M(t)(1 + L(t)S(t)Z_\propto)^{1/L(t)}, L(t) \neq 0 \tag{2}$$

where $C_{100\propto}(t)$ is the 100th percentile of y at time *t* and Z_α is the normal equivalent deviate of size α

Z scores and percentiles are two screening tools used by clinicians that help indicate growth/nourishment problems in pre-term infants. Ease of conversion between percentiles and z scores from growth curves is important for the clinician working time sensitive and time-limited environments and for researchers who often use z scores and percentiles as quantitative variables.

Cole and Green (1992) applied the LMS method to represent the skinfolds of Gambien women from birth to 50 years and for the body weight of US women age 1 through 21 years. They found that the z scores of the modeled samples fell within the appropriate bounds with a mean close to 0 and standard deviation close to 1. The penalized likelihood function provided a non-arbitrary method for smoothing parameters. Shortcomings of the method included edge-

effects and smoothing that is not uniform by age. The authors describe these limitations as common for other smoothing techniques and related to sampling decisions.

More recently researchers have investigated how to handle data that present with underlying skewness and kurtosis. While the LMS method provides for skewness, model fit diagnoses using worm plots indicated that kurtosis can distort fit (van Buuren & Fredriks, 2001). Skewness describes how far to the right or left data are distorted. Kurtosis indicates how peaked or flat data are compared to a normal distribution. Rigby and Stasinopoulos (2004) expanded the 3 parameter LMS distribution and presented a 4 parameter distribution they called the Box-Cox Power Exponential Distribution (BCPE).

Growth Curve Creation with BCPE distribution

The BCPE distribution includes the *lambda*, *mu*, and *sigma* parameters from the LMS model. It incorporates an additional parameter, *tau*, to specifically model kurtosis. The Box-Cox *t* (BCT) distribution is a similar 4-parameter model (also by Rigby and Stasinopolis, 2006) that can be described by the parameters *nu*, *mu*, *sigma*, and *tau* where *tau* is the *t* distribution of the random variable Z. The authors used *nu* to represent skewness in the BCPE distribution parameters (instead of *lambda*). The BCT distribution is best suited at modeling leptokurtosis, and the BCPE distribution is flexible enough to model skewness and leptokurtosis/platykurtosis (Rigby & Stasinopoulis, 2004). The LMS distribution can be considered a special case of the BCPE distribution because when *tau* is equal to 2 then the equation reduces to the LMS model. The probability density function of a standard power exponential variable, Z is

$$f_Z(Z) = \frac{\tau}{c2^{\left(1+\frac{1}{\tau}\right)}\Gamma\left(\frac{1}{\tau}\right)} \exp\left\{-0.5\left|\frac{z}{c}\right|^{\tau}\right\} \tag{3}$$

for $-\infty<z<\infty$ and $\tau>0$, where $c^2 = 2^{-2/t}\Gamma(1/\tau)[\Gamma(3/\tau)]^{-1}$

To illustrate the ability of the BCPE to provide a better fitting curve than the LMS curves when data exhibit kurtosis, Rigby and Stasinopolis (2004) applied their 4-parameter distribution to model the growth of Dutch boys age 0-21 years. Worm plots indicated evidence of kurtosis in the fitted 3-parameter (LMS) model of by Cole and Green (1992) and researchers applied a BCPE distribution to the BMI of boys. After generating a suitable power transformation for the explanatory variable of age and using the lowest model generalized Akaike Information Criterion (GAIC) to determine the 4 parameter values, Rigby and Stasinopoulus assessed model fit through worm plots, QQ plots, and Q statistics, which are subjective measures. Not surprisingly, results indicated that the BCPE distribution provided the best fit for the data in comparison to the Box-Cox normal distribution and the power exponential distribution.

Researchers can create growth curves with a BCPE distribution using the GAMLSS library contained within computer program R (R Core Team, 2017). GAMLSS is an abbreviation for the Generalized Additive Model of Location, Scale, and Shape. Rigby and Stasinopoulos (2005) introduced the GAMLSS model by relaxing the exponential family distribution for the response variable and replacing it with a general distribution family (Ospina, 2012). The BCPE distribution is a special case of GAMLSS. R is a powerful computer program that responds to user-inputted lines of code. A researcher who wants to use GAMLSS will need a basic understanding of R programming and the GAMLSS handbook (Stasinopoulos et al., 2016).

The calculation of z score requires the formula

$$z = \ \Phi^{-1}[\hat{F}_Y(Y)] \tag{4}$$

where Φ^{-1} is the inverse cumulative distribution function of a standard normal variate and $\hat{F}_Y(Y)$ is the fitted model cumulative distribution function of Y (Rigby & Stasinopoulos, 2004). If a

researcher is planning to obtain z scores outside of the GAMLSS program then a working knowledge of calculus is required.

Overall, obtaining growth curves using the BCPE method are not as simple as obtaining LMS curves. The most significant hurdle for researchers and clinicians is the computer programing competency necessary to use R. Compared to LMS ChartmakerPro, which has an easy to use graphical interface, R is a blank textbox that requires a basic knowledge of R software coding. The second struggle for clinicians adopting the BCPE method is the lack of a straightforward formula to convert between percentiles and z scores.

The WHO sought to produce an international set of growth reference charts to model how healthy children grow (WHO, 2006). After considering all methods of growth curve generation, the committee decided to use the BCPE method. However, once data analysis began, researchers found that models with a *tau* value of 2 modeled the data adequately which simplified the curves to LMS curves.

The 2006 Who Child Growth Standards are considered a benchmark for rigorous methodology and study design. The published growth charts are a result of the Multicentre Growth Reference Study (MGRS) that took place between 1997 and 2003. The MGRS includes longitudinal and cross sectional data from 6 countries. Only healthy mothers and children were considered for enrollment. Inclusion criteria for the mothers/families were: no economic, health, or environmental constraints on growth, mother willing to follow feeding guidelines, term birth, single birth, absence of significant health morbidity, and nonsmoking. The anthropomorphic measures collected by trained staff were weight, length, height, head and arm circumferences, triceps and subscapular skinfold thickness, and body mass index (BMI). A novel component of

the study was the emphasis on inclusion of mothers who were primarily breastfeeding their infants during the first 4 months of life (de Onis, 2004).

Beyond Rigby and Stasinopolous's (2004) demonstration of the BCPE method compared to LMS curves, there are few publications where authors have tested both methods with data and decided on a "better" model. WHO's growth references were intended to be modeled with BCPE curves, but decided LMS curves were appropriate without providing details of this decision. When researchers compare models of varying complexity they use the principle of parsimony ("simpler is better") to aid in their model selection process. To date, there are no publications that explicitly compare model fit between BCPE curves and LMS curves for the same sample. I aim to address the question of whether the 4 parameter distributions of BCPE/BCT are significantly better at modeling intrauterine growth curves of preterm infants compared to curves generated using a 3-parameter (LMS) distribution. I used a combination of statistical measures and clinical considerations to determine the better model.

Method

Subjects

The data were provided by Pediatrix Medical Group, Inc. The de-identified data were collected from medical records from 33 NICU's in the US. The sample included 391,681 babies who were admitted to a NICU (in the Pediatrix network) between the years 1998-2006. The infant measurements were conducted by hospital nurses. These data were cross-sectional and all measurements were collected at the time of birth.

As these datasets have been previously used in publication (Olsen et al., 2010), they were already "cleaned" and ready for use. Briefly, implausible outliers were removed and babies with characteristics known to impact fetal growth (like congenital birth abnormalities) were removed

(see Groveman, 2008, for full details). Table 1 lists the variables contained in the datasets. The data was randomly split into training and validation datasets. For this analysis I used the training dataset.

Growth curve generation

First, I imported the data for males into R version 3.3.3. using the `read.table` command. I loaded the GAMLSS library and used `na.omit` to delete missing observations because GAMLSS will not run with missing data. For each anthropomorphic variable of interest (birth weight, birth head circumference, and birth length) I asked GAMLSS to generate growth curves with gestational age as the explanatory variable. GAMLSS includes a function to automatically select the best distribution to model the relationship between an explanatory and response variable based on the lowest AIC between models. The function prompt name, LMS, is somewhat misleading/confusing because the function tests the fit of the LMS, BCPE, and BCT distributions.

I used the `LMS` call to generate a power transformation for age and generate an auto-selected best distribution (which was usually a 4 parameter distribution). The age transformation was applied to a GAMLSS call with a specified 3-parameter distribution in order to force GAMLSS to use the LMS distribution (called BCCGo in GAMLSS). I then ran a 3- and 4-parameter model without a power transformation for age. I selected the P-splines smoother by using the call `pb()`. See below for a portion of the R code for curve generation for boys and birth weight. The complete R code can be found in Appendix 1.

```
*GAMLSS fitted object, no age transformation, 4 parameter distribution;
```

```
mBCPE_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),

nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCPEo,

data=boys2);

*autoselect best distribution with a power transformation for age;

mBCPE_T<-lms(BirthWeight, GestAge, data=boys2, trans.x=T);

Tage=(boys2$GestAge)^(mBCPE_T$power);

*GAMLSS fitted object, age transformation, 3-parameter dist;

mBCCG_T<- gamlss(BirthWeight~pb(Tage), sigma.formula=~pb(Tage),

nu.formula=~pb(Tage), tau.fomula=~pb(Tage), family=BCCGo, data=boys2);

*GAMLSS fitted object, NO age transformation, 3-parameter dist;

mBCCG_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),

nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,

data=boys2);

*print results of autoselected distribution model to screen;
```

summary(mBCPE_T);

Model Comparison

To facilitate model fit comparison, I called the degrees of freedom for each parameter (mu, sigma, nu, and tau when applicable) by using the model name followed by the prompts `$mu.df`, `$sigma.df`, `$tau.df`, or `$nu.df`. I collected model GAIC scores by using the `GAIC` command. To produce the differences in expected versus observed centiles, I used the `Centiles` call plus a specification for the 3rd, 10th, 25th, 50th, 75th, 90th, and 97[h] percentiles. I plotted the model GAIC scores and differences, degrees of freedom for the parameters, and the differences in expected versus observed centiles in a series of tables in Excel. I have included an example of my R code for model fit comparison criteria below.

```
GAIC(mBCPE_noT, mBCPE_T, mBCCG_noT, mBCCG_T);
centiles(mBCPE_noT, xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
```

```
centiles(mBCPE_T, xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90, 97));
centiles(mBCCG_noT,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90, 97));
centiles(mBCCG_T,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90, 97));
mBCPE_T$mu.df;
mBCPE_T$sigma.df;
mBCPE_T$nu.df;
mBCPE_T$tau.df;
mBCPE_noT$mu.df;
mBCPE_noT$sigma.df;
mBCPE_noT$nu.df;
mBCPE_noT$tau.df;
mBCCG_noT$mu.df;
mBCCG_noT$sigma.df;
mBCCG_noT$nu.df;
mBCCG_noT$tau.df;
mBCCG_T$mu.df;
mBCCG_T$sigma.df;
mBCCG_T$nu.df;
mBCCG_T$tau.df;
```

Decision Criteria

The traditional method for model comparison (and the one used by Rigby &

Stasinopoulos, 2004) is to select the model with the lowest GAIC. GAIC is calculated by

$$GAIC(\#) = \hat{D} + \#.df \qquad (5)$$

where *df* is the total EDFs of the model and

$$\hat{D} = -2(\hat{l}) \qquad (6)$$

in which $\hat{D}$ is the fitted deviance and $\hat{l}$ is the fitted log likelihood. By substituting in different

"penalties" for # in Equation 4, the GAIC (when # = 3) becomes a special case of the AIC (# =

2) and SBC (# = log[*n*]). Caution should be used when choosing between models based on

lowest GAIC as a sole indicator. It is possible to include more parameters than are *realistically*

beneficial to the overall fit of the model given the principle of parsimony.

Because clinicians are interested in percentiles for measuring growth, I explored how

percentiles changed between curves generated with a 3-parameter and 4-parameter distribution.

I considered the acceptability of the differences between observed and expected centiles in an aggregate table across gestational ages. In the WHO Technical Report, I noticed that the acceptable difference between observed and expected percentile was 2 percentile points (WHO, 2006). Therefore, I utilized this same criteria to judge whether the differences in the observed and expected centiles generated by the 3- and 4-parameter distributions were considered acceptable. Following the principle of parsimony, if the centiles generated from the LMS curves and the centiles generated from the BCPE curves both met the criteria for acceptability then I retained the simpler (LMS) model's curves. I chose this method for model selection (over the traditional GAIC model selection method) because I was interested in the clinical/practical significance of the difference in the models.

This method was repeated for birth length, head circumference, and birth weight with an explanatory variable of gestational age for the male and female datasets.

Results

Descriptive statistics for males and females can be found in Table 2.

Addressing age transformation

Tables 3, 6, and 9 contain the GAIC by model for males for length, head circumference, and weight, respectively. Tables 12, 15, and 18 contain the GAIC by model for females for length, head circumference, and weight, respectively.

For (all) boys, the difference in model GAIC between same distributions with and without a power transformation for age had a range of 0.66—25.5. For (all) girls, the difference in model GAIC between same distributions with and without a power transformation for age had a range of 1.22—25.5. Changes in GAIC were small, but ultimately there was no difference in

the observed centiles (versus the expected). I decided that the power transformation for gestational age did not contribute clinically significant changes.

Males

Tables 4, 5, 7, 8, 10, and 11 indicate the EDFs and the expected versus observed centiles. There were no differences in centiles by power transformation or model distribution.

For length, the LMS method provided acceptable fit. There were no differences in expected and observed centiles greater than 1.36 percentile points for the LMS model. The BCPE model had a GAIC that was 43 lower than the GAIC of the LMS model.

For head circumference, the LMS method provided acceptable fit. I included the BCT distribution (a 4-parameter distribution) because GAMLSS automatically suggested it as the best distribution. There were no differences in expected and observed centiles greater than 0.72 percentile points for the LMS model. The model with the lowest GAIC was the BCT model and it was lower than the LMS model by only 7.2.

For weight, the LMS method provided acceptable fit. There was no difference in the expected and observed centiles greater than 1.28 percentile points for the LMS model. The BCPE model had a GAIC that was 278 lower than the GAIC of the LMS model.

Females

Tables 13, 14, 16, 17, 19, and 20 indicate the EDFs and the expected versus observed centiles. There were no differences in centiles by power transformation or model distribution.

For length, the LMS method provided acceptable fit. There was no difference in the expected and observed centiles greater than 0.91percentile points for the LMS model. The BCPE model had a GAIC that was 28 lower than the GAIC of the LMS model.

For head circumference, the LMS method provided acceptable fit. There was no difference in the expected and observed centiles greater than 1.16 percentile points for the LMS model. The BCPE model had a GAIC that was 7.4 lower than the GAIC of the LMS model.

For weight, the LMS method provided acceptable fit. There was no difference in the expected and observed centiles greater than 0.91 percentile points for the LMS model. The BCPE model has a GAIC that was 131.85 lower than the GAIC of the LMS model.

Discussion

The results of this analysis support that for clinical considerations the (simpler) 3-parameter model LMS distribution (called BCCGo in R) is adequate for growth curve modeling with birth length, head circumference, and weight by gestational age in preterm infants. From a strictly statistical model comparison perspective, GAIC was lower for the models with the added parameter to model kurtosis. However, the defining criteria of centile difference (no greater than 2 percentile points) between observed and expected indicated that the LMS curves provided adequate fit in all instances. The principle of parsimony indicates that the simpler solution is the better solution. In this case because the 3-parameter LMS method models provided adequate fit I determined that the slight difference in model GAIC was not enough to justify the addition of the 4th parameter.

The benefits of using the LMS method are 1) the LMS ChartmakerPro computer software is user-friendly. It has a simple graphical interface where a researcher or clinician can open the program, upload data, click buttons, and produce LMS growth curves. The time it takes to learn the program is brief. 2) The LMS method provides a simple formula to calculate z scores and percentiles once the EDFs are chosen. There is no calculus involved, unlike the BCPE method.

3) Clinicians are familiar with the LMS method. 4) I have shown that LMS curves model this data accurately and fulfil the principle of parsimony.

This analysis is the first to explicitly demonstrate the differences between a 4-parameter model and a 3-parameter model with model fit statistics and a clinical application perspective. The WHO Child Growth Standards Technical Methods (WHO, 2006) indicated that only small differences in global deviance their decision to use the LMS curves over the BCPE curves. I have provided the individual model fit statistics for all models that were considered in this study. I followed the WHO acceptability rule of 2 percentile points deviation between observed and expected percentiles to judge whether a model had adequate fit.

The racially diverse sample composition and sample size is a strength of this study. A simultaneous strength and weakness for this study include the fact that these data are from health records and therefore not "research quality" data. Using "real world" data are messy but allow for high external validity. Health record data may contain errors/variability by site (hospital, in this case), training methods, and manual transcription.

Using 2 percentile points for the definition of acceptable deviation for observed versus expected percentiles is a limitation of the study because a deviation of 2 percentile points at the lowest percentiles (3rd and 5th) would have a greater weight than deviations in the middle centiles (50th). This can be addressed by using z scores and empirical centiles. In future studies I will examine the differences in expected and observed centiles by model in full instead of through an aggregate table, and I will use z scores and empirical centiles to assess model fit.

Clinically, LMS curves classify preterm babies into percentiles no differently than the BCPE curves. The statistical measure of lowest GAIC is not enough to justify using a 4-parameter distribution for growth curve generation and centile estimation. Researchers may wish

to explore whether this finding can be replicated for term-babies. For now clinicians can close

their old calculus textbooks because growth curves generated by the LMS method are sufficient.

References

Babson, S., & Benda, G. (1976). Growth graphs for the clinical assessment of infants of varying gestational age. *The Journal of Pediatrics*, *89*(5), 814—820.

Bhatia, J. (2013). Growth Curves: How to Best Measure Growth of the Preterm Infant. *The Journal of Pediatrics*, *162*(3), S2—S6. doi: 10.1016/j.jpeds.2012.11.047

Cole, T., & Green, P. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, *11*, 1305—1319. doi: 10.1002/sim.4780111005

CDC. (2002). 2000 CDC Growth Charts for the United States: Methods and Development. *Vital and Health Statistics*, *11*(246).

de Onis, M., Garza, C., Victora, C., Onyango, A., Frongillo, E., & Martines, J. (2004). The WHO Multicentre Growth Reference Study: Planning, study design, and methodology. *Food and Nutrition Bulletin*, *25*(1), S15—S26. doi: 10.1177/15648265040251S103

Fenton, T. (2003). A new growth chart for preterm babies: Babson and Benda's chart updated with recent data and a new format. *BMC Pediatrics*, *3*(13), e1—e10. doi: 10.1186/1471-2431-3-13

Fenton, T., & Kim, J. (2013). A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatrics*, *13*(59), 1—13. doi: 10.1186/1471-2431-13-59

Gibson, E., & Nawab, U. (2015a). *Small-for-Gestational Age (SGA) Infant*. Retrieved from https://www.merckmanuals.com/professional/pediatrics/perinatal-problems/small-for-gestational-age-sga-infant

Gibson, E., & Nawab, U. (2015b). *Large-for-Gestational Age (LGA) Infant*. Retrieved from

      http://www.merckmanuals.com/professional/pediatrics/perinatal-problems/large-for-

      gestational-age-lga-infant

Groveman, S. (2008). *New Preterm Infant Growth Curves Influence of Gender and Race on

      Birth Size*. Retrieved from http://hdl.handle.net/1860/2877

Indrayan, A. (2014). Demystifying LMS and BCPE Methods of Centile Estimation for Growth

      and Other Health Parameters. *Indian Pediatrics*, *51*(1), 37—43. Doi: 10.1007/s13312-

      014-0310-6

Olsen, I., Groveman, S., Lawson, M., Clark, R., & Zemel, B. (2010). New Intrauterine Growth

      Curves Based on United States Data. *Pediatrics, 125*, e214—e224. doi:

      10.1542/peds.2009-0913

Olsen, I., Lawson, M., Ferguson, A., Cantrell, R., Grabich, S., Zemel, B, & Clark, R. (2014).

      BMI curves for preterm infants. *Pediatrics*, *135*(3), e1—e10. doi: 10.1542/peds.2014-

      2777.

Ospina, R. (2012) *A brief introduction to GAMLSS modeling* [PowerPoint slides]. Retrieved

      from http://www.de.ufpe.br/~raydonal/Talks/Gamlss-Theory.pdf

Pan, H., & Cole, T. (2001). *User's Guide to LMS Chartmaker*. Medical Research Council, UK:

      Author.

R Core Team (2017). *R: A language and environment for statistical computing.* R Foundation for

      Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rigby, R., & Stasinopoulos, D. (2004). Smooth centile curves for skew and kurtotic data

      modelled using the Box–Cox power exponential distribution. *Statistics in Medicine*, *23*,

      3053-3076. doi: 10.1002/sim.1861

Rigby, R., & Stasinopoulos, D. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, *54*(3), 507—554. doi: 10.1111/j.1467-9876.2005.00510.x

Rigby, R., & Stasinopoulos, D. (2006). Using the Box–Cox t distribution in GAMLSS to model skewness and kurtosis. *Statistical Modelling*, *6*, 209-229. Doi: 10.1191/1471082X06st122oa

van Burren, S., & Fredriks, M. (2001). Worm plot: A simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, *20*, 1259-1277. doi: 10.1002/sim.746

WHO. (2006). *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, Switzerland: Author.

Tables for Paper I

Table 1: Variables Contained in the Pediatrix Dataset				
Variable	Label	Unit	Range	Categories
Gestational Age	GestAge	Completed weeks	22-42	
Race/Ethnicity	Race			Black, White, Hispanic, Other
Facility State	Facility_State	US state name		
Facility ID	Facility_ID		1-3797	
How many days old was the infant upon admission	Admit_DSB	Days	0-366	
Unknown and unused	GestAgeDays		1-6	
Transferred in to NICU from adjoining hospital or separate	AdmitGroup			Inborn, Outborn, Missing
Birth Weight	BirthWeight	kg	0.42-5.3	
Birth Length	BirthLength	cm	25-59.5	
Head circumference at birth	BirthHC	cm	18-40.5	
Age of mother during pregnancy	MatAgePreg	years	11-54	
Birth year of the infant	BirthYear	years	1998-2006	
Delivery method	Delivery			Cesarean section, forcepts extraction, missing data, vacuum extraction, vaginal delivery
Antenatal steriods	AntenatalSteriods	0 = no, 1 = yes	0-1	
Did the mother report smoking	RptofSmoking			yes, no
Was there a diagnosis of Intrauterine growth restriction	RPT_IUGR			yes, no
Was there a diagnosis of diabetes	DiabetesRPT			yes, no
Was the mother taking insulin	InsulinRPT			yes, no
Was there a diagnosis of Pre-Eclampsia or Eclampsia	PreOREclapsia			yes, no
Number of children delivered	BirthNumber		1	
APGAR score at 1 minute after delivery	APGAR1	0 points to 10 points	0-10	
APGAR score at 5 minutes after delivery	APGAR5	0 points to 10 points	0-10	

note: highlighted variables are the variables used for this analysis

Table 2: Descriptive Statistics							
Males, *N* = 62, 822							
Variable	**Unit**	**Min**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Gestational Age	Weeks	22	33	36	35.4	38	42
Birth Weight	kg	0.42	2.01	2.71	2.66	3.35	5.3
Birth Length	cm	25	44	48	46.84	51	59.5
Head Circumference	cm	18	30.5	33	32.23	34.5	40.5
Females, *N* = 47, 019							
Variable	**Unit**	**Min**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Gestational Age	Weeks	22	33	35	35.12	38	42
Birth Weight	kg	0.38	1.83	2.51	2.49	3.19	5.03
Birth Length	cm	25.5	43	47	45.75	49.5	59.5
Head Circumference	cm	17	29.5	32	31.45	34	39.5

Table 3: GAIC Comparisons by Model for Males Birth Length				
	BCPE PTrans	BCPE No PTrans	LMS No PTrans	LMS PTrans
BCPE PTrans	**296698.4**	-13.4	-56.5	-57.7
BCPE No PTrans	13.4	296711.8	-43.1	-44.3
LMS No PTrans	56.5	43.1	296754.9	-1.2
LMS PTrans	57.7	44.3	1.2	296756.1

Table 4: Parameter EDF by Model for Length for Males Birth Length				
Model	Mu	Sigma	Nu	Tau
BCPE with PTrans	12	11.05	6.42	8.58
BCPE no Ptrans	12.1	11.55	6.36	1
LMS with no Ptrans	12.08	11.68	6.2	
LMS with Ptrans	11.96	11.26	6.25	

Table 5: Expected and Observed Centiles by Model for Males Birth Length								
	BCPE with PTrans		BCPE no PTrans		LMS no PTrans		LMS with PTrans	
Expected	Observed	Difference	Observed	Difference	Observed	Difference	Observed	Difference
Below 3	3.32	-0.32	3.32	-0.32	3.32	-0.32	3.32	-0.32
Below 10	9.86	0.14	9.85	**0.15**	9.77	0.23	9.77	0.23
Below 25	25.33	-0.33	25.17	**-0.17**	24.71	0.29	24.71	0.29
Below 50	50.47	-0.47	50.43	-0.43	50.43	-0.43	50.43	-0.43
Below 75	75.34	-0.34	75.34	**-0.34**	76.36	-1.36	76.36	-1.36
Below 90	90.68	-0.68	90.84	-0.84	90.83	**-0.83**	90.84	-0.84
Below 97	97.15	-0.15	97.24	-0.24	97.14	**-0.14**	97.14	-0.14

Table 6: GAIC Comparisons by Model for Males Head Circumference						
	BCT with PTrans	BCT no PTrans	BCPE with PTrans	BCPE no PTrans	LMS with PTrans	LMS no PTrans
BCT with PTrans	**235525.1**	-5.6	-8.5	-9.8	-12.8	-14.1
BCT no PTrans	5.6	235530.7	-2.9	-4.2	-7.2	-8.5
BCPE with PTrans	8.5	2.9	235533.6	-1.3	-4.3	-5.6
BCPE no PTrans	9.8	4.2	1.3	235534.9	-3	-4.3
LMS with PTrans	12.8	7.2	4.3	3	235537.9	-1.3
LMS no PTrans	14.1	8.5	5.6	4.3	1.3	235539.2

Table 7: Parameter EDF by Model for Head Circumference for Males Head Circumference				
Model	Mu	Sigma	Nu	Tau
BCT with PTrans	13.04	7.83	2.39	1
BCT no PTrans	13.24	7.99	2.66	1
BCPE with PTrans	13.01	7.78	2.38	1
BCPE no PTrans	13.22	7.99	2.61	1
LMS with PTrans	13.02	7.77	2.36	
LMS no PTrans	13.21	8	2.59	

Table 8: Expected and Observed Centiles by Model for Males Head Circumference												
	BCPE with PTrans		BCPE no PTrans		BCT no PTrans		LMS no Ptrans		BCT with PTrans		LMS with PTrans	
Expected	Observed	Difference	Observed	Difference	Observed	Difference	Observed	Difference	Observed	Difference	Observed	Difference
Below 3	3.30	-0.30	3.30	-0.30	3.30	-0.30	3.30	-0.30	3.30	-0.30	3.30	-0.30
Below 10	10.27	-0.27	10.27	-0.27	10.27	-0.27	10.07	**-0.07**	10.27	-0.27	10.09	-0.09
Below 25	25.64	-0.64	25.72	-0.72	25.72	-0.72	25.64	**-0.64**	25.72	-0.72	25.64	-0.64
Below 50	49.38	0.62	49.38	0.62	49.38	0.62	49.38	0.62	49.38	0.62	49.38	0.62
Below 75	74.82	0.18	74.81	0.19	74.81	0.19	74.82	**0.18**	74.64	0.36	74.82	0.18
Below 90	89.70	0.30	89.70	0.30	89.47	0.53	90.16	**-0.16**	89.70	0.30	90.16	-0.16
Below 97	96.98	0.02	96.99	0.01	96.98	0.02	96.98	0.02	96.98	0.02	96.98	0.02

Table 9: GAIC Differences by Model for Birth Length in Males Birth Weight				
	BCPE with PTrans	BCPE no PTrans	LMS no PTrans	LMS with PTrans
BCPE with PTrans	**62938.48**	-25.58	-303.65	-304.31
BCPE no PTrans	25.58	62964.06	-278.07	-278.73
LMS no PTrans	303.65	278.07	63242.13	-0.66
LMS with PTrans	304.31	278.73	0.66	63242.79

Table 10: Parameter EDF by Model for Weight for Males Birth Weight				
Model	Mu	Sigma	Nu	Tau
BCPE with PTrans	13.9	12.15	4.91	5.36
BCPE no Ptrans	14.43	12.69	4.65	1
LMS with no Ptrans	14.5	12.88	4.4	
LMS with Ptrans	13.91	12.26	4.39	

Table 11: Expected and Observed Centiles by Model for Males Length								
	BCPE with PTrans		BCPE no PTrans		LMS no PTrans		LMS with PTrans	
Expected	Observed	Difference	Observed	Difference	Observed	Difference	Observed	Difference
Below 3	3.32	-0.32	3.31	**-0.31**	3.45	-0.45	3.45	-0.45
Below 10	9.97	0.03	9.99	**0.01**	9.69	0.31	9.69	0.31
Below 25	24.74	0.26	24.78	**0.22**	23.72	1.28	23.71	1.29
Below 50	49.92	0.08	49.97	**0.03**	50.07	-0.07	50.07	-0.07
Below 75	75.19	-0.19	75.23	**-0.23**	76.24	-1.24	76.26	-1.26
Below 90	90.1	-0.1	90.08	**-0.08**	90.37	-0.37	90.36	-0.36
Below 97	96.88	0.12	96.87	**0.13**	96.64	0.36	96.64	0.36

Table 12: GAIC Difference by Model for Females Birth Length				
	BCPE with PTrans	BCPE no PTrans	LMS with PTrans	LMS no PTrans
BCPE with PTrans	221695.6	-2	-28	-29.8
BCPE no PTrans	2	221697.6	-26	-27.8
LMS with PTrans	28	26	221723.6	-1.8
LMS no PTrans	29.8	27.8	1.8	221725.4

Table 13: Parameter EDF by Model for Females Birth Length				
Model	Mu	Sigma	Nu	Tau
BCPE with PTrans	13.77	10.27	2	3.18
BCPE no Ptrans	14.45	10.59	2	1
LMS no Ptrans	14.51	10.63	2	
LMS with Ptrans	13.83	10.33	2	

Table 14: Expected and Observed Centiles by Model for Females Birth Length								
	BCPE with PTrans		BCPE no PTrans		LMS no PTrans		LMS with PTrans	
Expected	Observed	Difference	Observed	Difference	Observed	Difference	Observed	Difference
Below 3	3.09	-0.09	3.09	**-0.09**	3.1	-0.1	3.09	-0.09
Below 10	9.96	0.04	9.66	**0.34**	9.65	0.35	9.66	0.34
Below 25	25.32	-0.32	25.32	**-0.32**	24.58	0.42	24.58	0.42
Below 50	50.91	-0.91	50.91	-0.91	50.81	**-0.81**	50.91	-0.91
Below 75	74.47	0.53	75.47	**-0.47**	74.47	0.53	74.47	0.53
Below 90	89.48	0.52	89.48	0.52	89.48	0.52	89.48	0.52
Below 97	96.92	0.08	97.13	**-0.13**	97.91	-0.91	96.92	0.08

Table 15: GAIC Difference by Model for Females Head Circumference				
	BCPE with PTrans	BCPE no PTrans	LMS with PTrans	LMS no PTrans
BCPE with PTrans	**173403.3**	-11.5	-17.5	-18.9
BCPE no PTrans	11.5	173414.8	-6	-7.4
LMS with PTrans	17.5	6	173420.8	-1.4
LMS no PTrans	18.9	7.4	1.4	173422.2

Table 16: Parameter EDF by Model for Females Head Circumference				
Model	Mu	Sigma	Nu	Tau
BCPE with PTrans	12.31	9.81	2	6.06
BCPE no Ptrans	12.37	10.11	2	1
LMS with Ptrans	12.27	9.8	2	
LMS no Ptrans	12.38	10.14	2	

Table 17: Expected and Observed Centiles by Model for Females Head Circumference								
	BCPE with PTrans		BCPE no PTrans		LMS no PTrans		LMS with PTrans	
Expected	Observed	Difference	Observed	Difference	Observed	Difference	Observed	Difference
Below 3	3.22	-0.22	3.22	**-0.22**	3.32	-0.32	3.32	-0.32
Below 10	10.2	-0.2	10.19	-0.19	9.94	**0.06**	9.95	0.05
Below 25	24.95	0.05	24.97	**0.03**	24.15	0.85	24.13	0.87
Below 50	49.56	0.44	49.63	0.37	49.64	**0.36**	49.62	0.38
Below 75	75.26	-0.26	75.32	**-0.32**	76.16	-1.16	76.16	-1.16
Below 90	90.1	-0.1	90.07	**-0.07**	90.33	-0.33	90.33	-0.33
Below 97	96.85	0.15	96.86	**0.14**	96.7	0.3	96.7	0.3

Table 18: GAIC Difference by Model for Females Birth Weight				
	BCPE with PTrans	BCPE no PTrans	LMS with PTrans	LMS no PTrans
BCPE with PTrans	**43265.05**	-25.54	-156.17	-157.39
BCPE no PTrans	25.54	43290.59	-130.63	-131.85
LMS with PTrans	156.17	130.63	43421.22	-1.22
LMS no PTrans	157.39	131.85	1.22	43422.44

Table 19: Parameter EDF by Model for Females Birth Weight				
Model	Mu	Sigma	Nu	Tau
BCPE with PTrans	12.8	12.9	5.98	7
BCPE no Ptrans	13.36	13.59	6.05	1
LMS no Ptrans	13.39	13.66	5.9	
LMS with Ptrans	12.8	12.9	5.7	

Table 20: Expected and Observed Centiles by Model for Females Weight								
	BCPE with PTrans		BCPE no PTrans		LMS no PTrans		LMS with PTrans	
Expected	Observed	Difference	Observed	Difference	Observed	Difference	Observed	Difference
Below 3	3.09	-0.09	3.09	**-0.09**	3.1	-0.1	3.09	-0.09
Below 10	9.66	0.34	9.66	**0.34**	9.65	0.35	9.66	0.34
Below 25	25.32	-0.32	25.32	**-0.32**	24.58	0.42	24.58	0.42
Below 50	50.91	-0.91	50.91	-0.91	50.91	-0.91	50.91	-0.91
Below 75	74.47	0.53	74.47	0.53	74.47	0.53	75.47	-0.47
Below 90	89.48	0.52	89.48	-0.52	89.48	0.52	89.48	0.52
Below 97	96.92	0.08	97.13	-0.13	96.92	**0.08**	96.92	0.08

Figures for Paper I

Figure 1: Growth Curves from LMS distribution for Preterm Infant Males—Birth Weight by
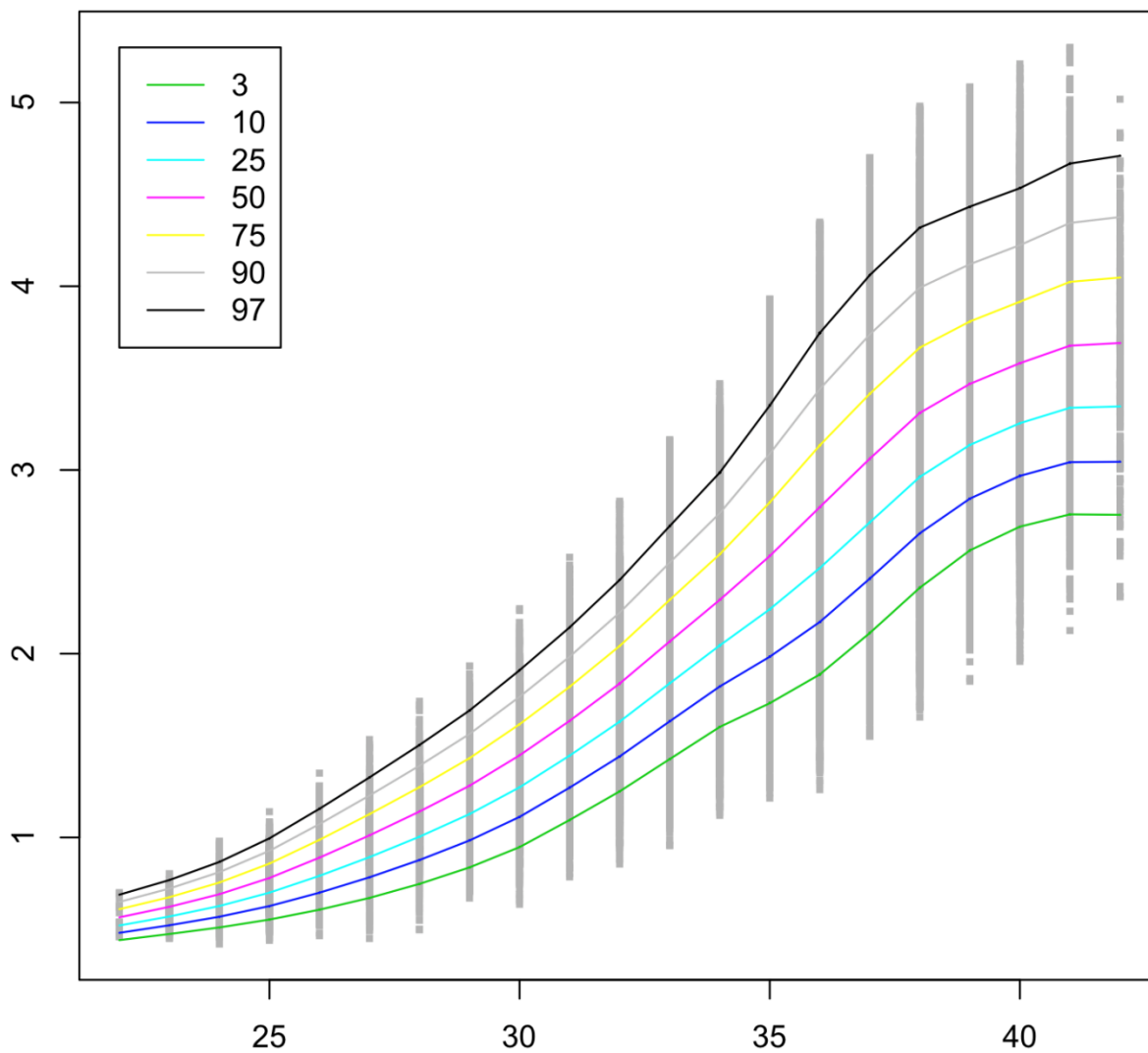
Gestational Age

Note: The x axis represents gestational age measured in weeks and the y axis represents weight

measured in kg

Figure 2: Growth Curves from LMS distribution for Preterm Infant Males—Birth Head

Circumference by Gestational Age

Note: The x axis represents gestational age measured in weeks and the y axis represents head

circumference measured in cm

Figure 3: Growth Curves from LMS distribution for Preterm Infant Males—Birth Length by Gestational Age

Note: The x axis represents gestational age measured in weeks and the y axis represents length measured in cm

Figure 4: Growth Curves from LMS distribution for Preterm Infant Females—Birth Weight by

Gestational Age

Note: The x axis represents gestational age measured in weeks and the y axis represents length

measured in kg

Figure 5: Growth Curves from LMS distribution for Preterm Infant Females—Birth Head

Circumference by Gestational Age

Note: The x axis represents gestational age measured in weeks and the y axis represents head

circumference measured in cm

Figure 6: Growth Curves from LMS distribution for Preterm Infant Females—Birth Length by

Gestational Age

Note: The x axis represents gestational age measured in weeks and the y axis represents length

measured in cm

Figure 7: Growth Curves from BCPE distribution for Preterm Infant Males—Birth Weight by Gestational Age

Note: The x axis represents gestational age measured in weeks and the y axis represents weight measured in kg

Figure 8: Growth Curves from BCPE distribution for Preterm Infant Males—Birth Head

Circumference by Gestational Age

Note: The x axis represents gestational age measured in weeks and the y axis represents head
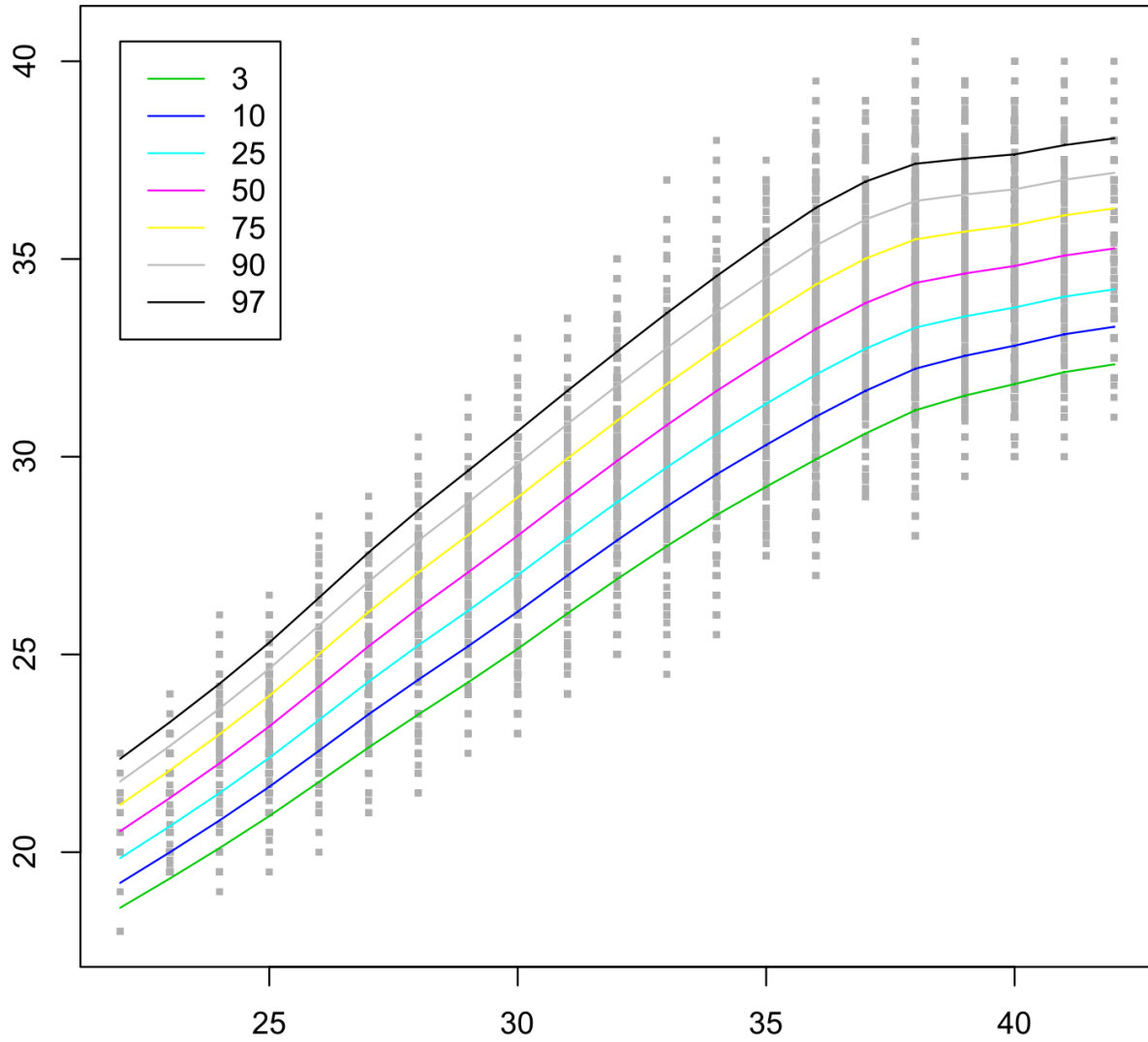
circumference measured in cm

Figure 9: Growth Curves from BCPE distribution for Preterm Infant Males—Birth Length by

Gestational Age

Note: The x axis represents gestational age measured in weeks and the y axis represents length

measured in cm

Figure 10: Growth Curves from BCPE distribution for Preterm Infant Females—Birth Weight by Gestational Age
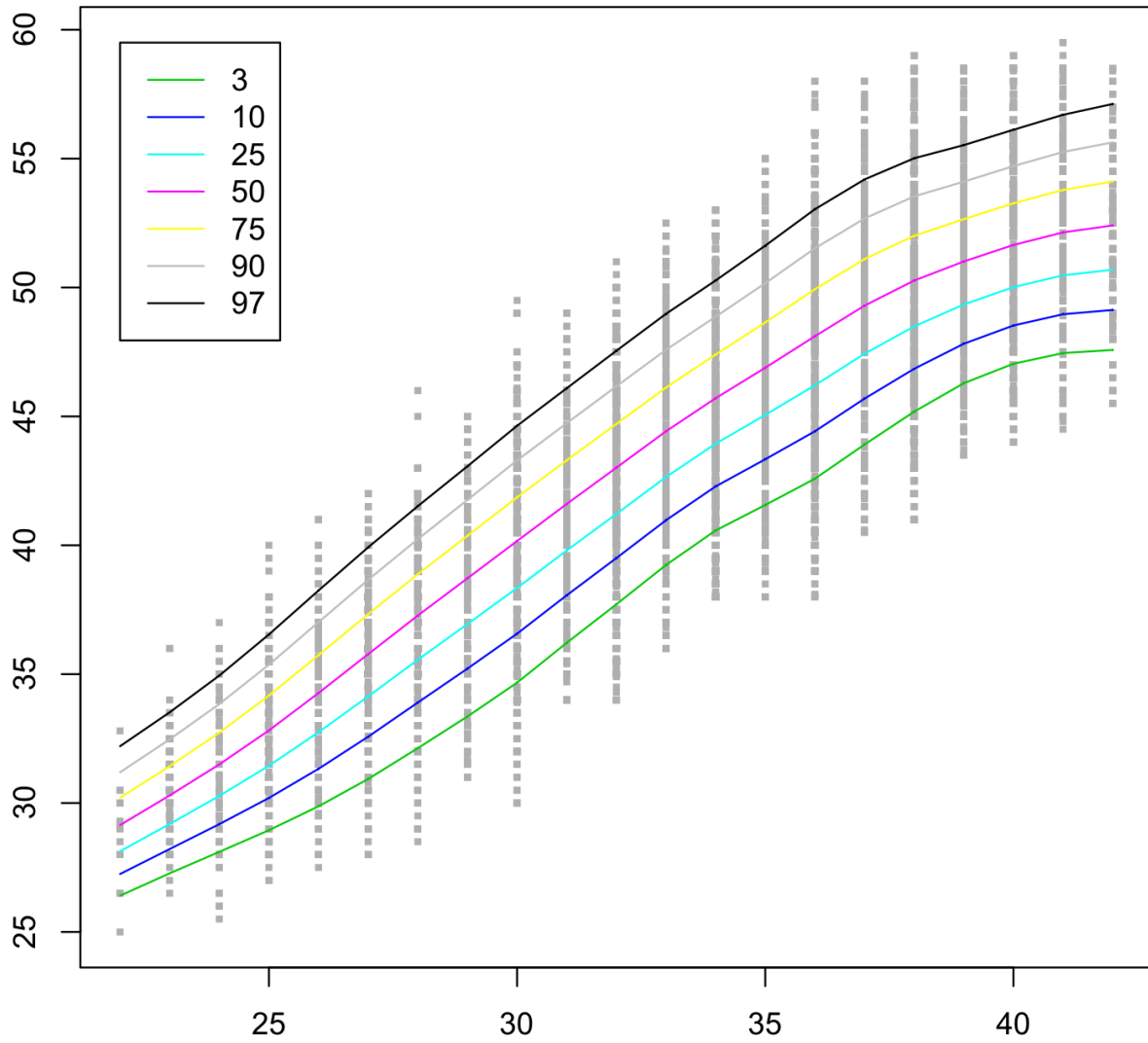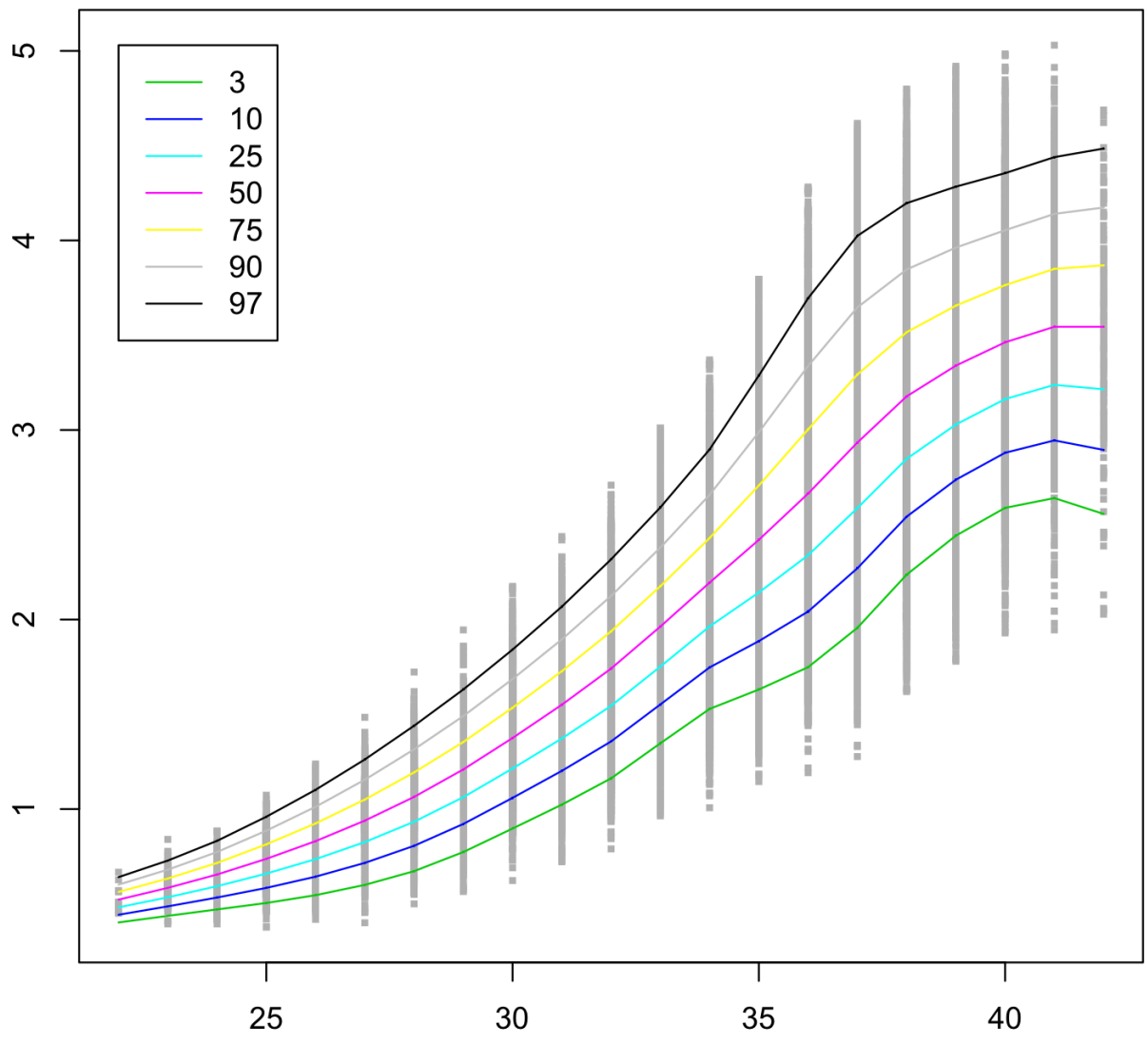

Note: The x axis represents gestational age measured in weeks and the y axis represents weight measured in kg

Figure 11: Growth Curves from BCPE distribution for Preterm Infant Females—Birth Head
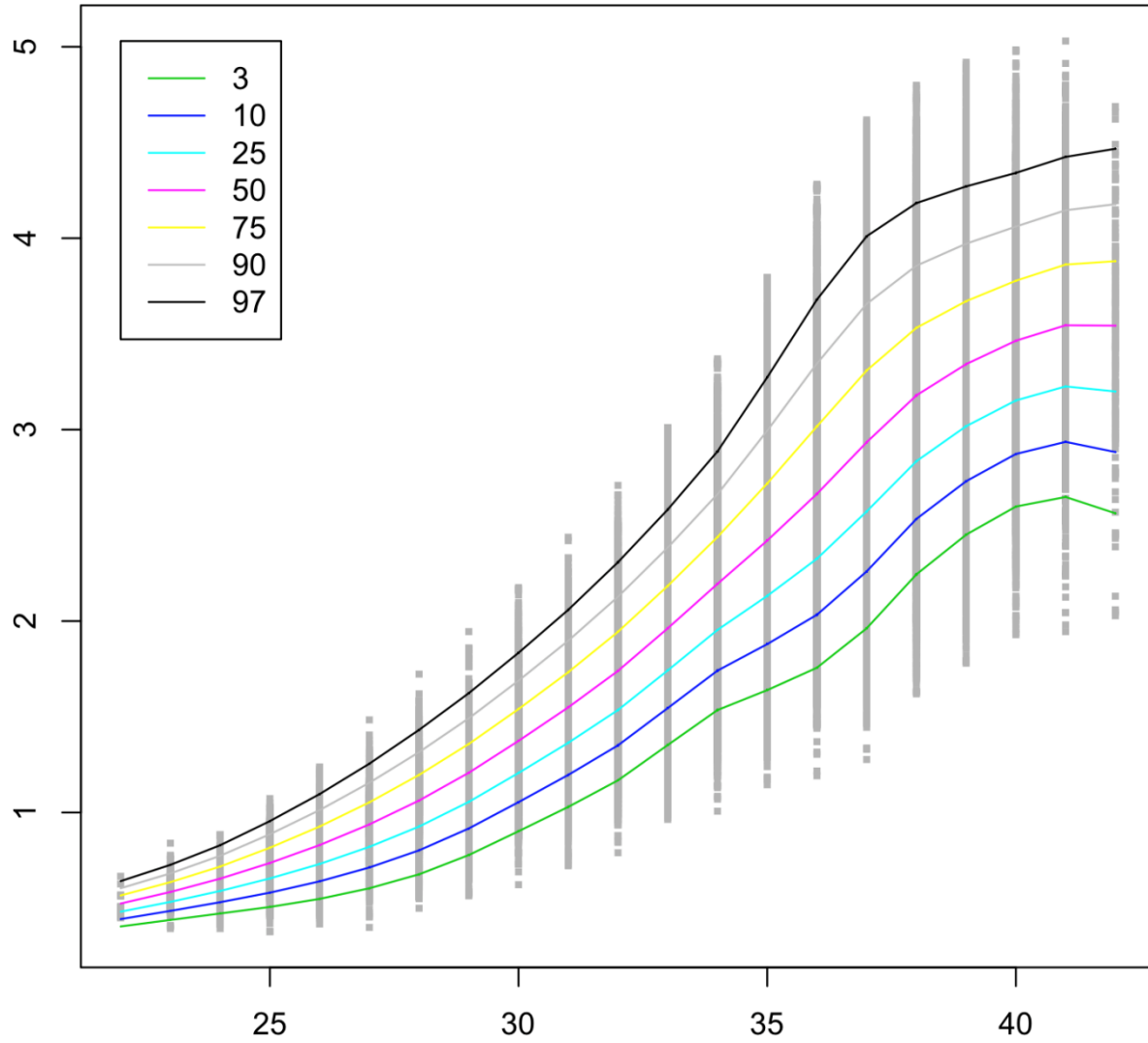
Circumference by Gestational Age
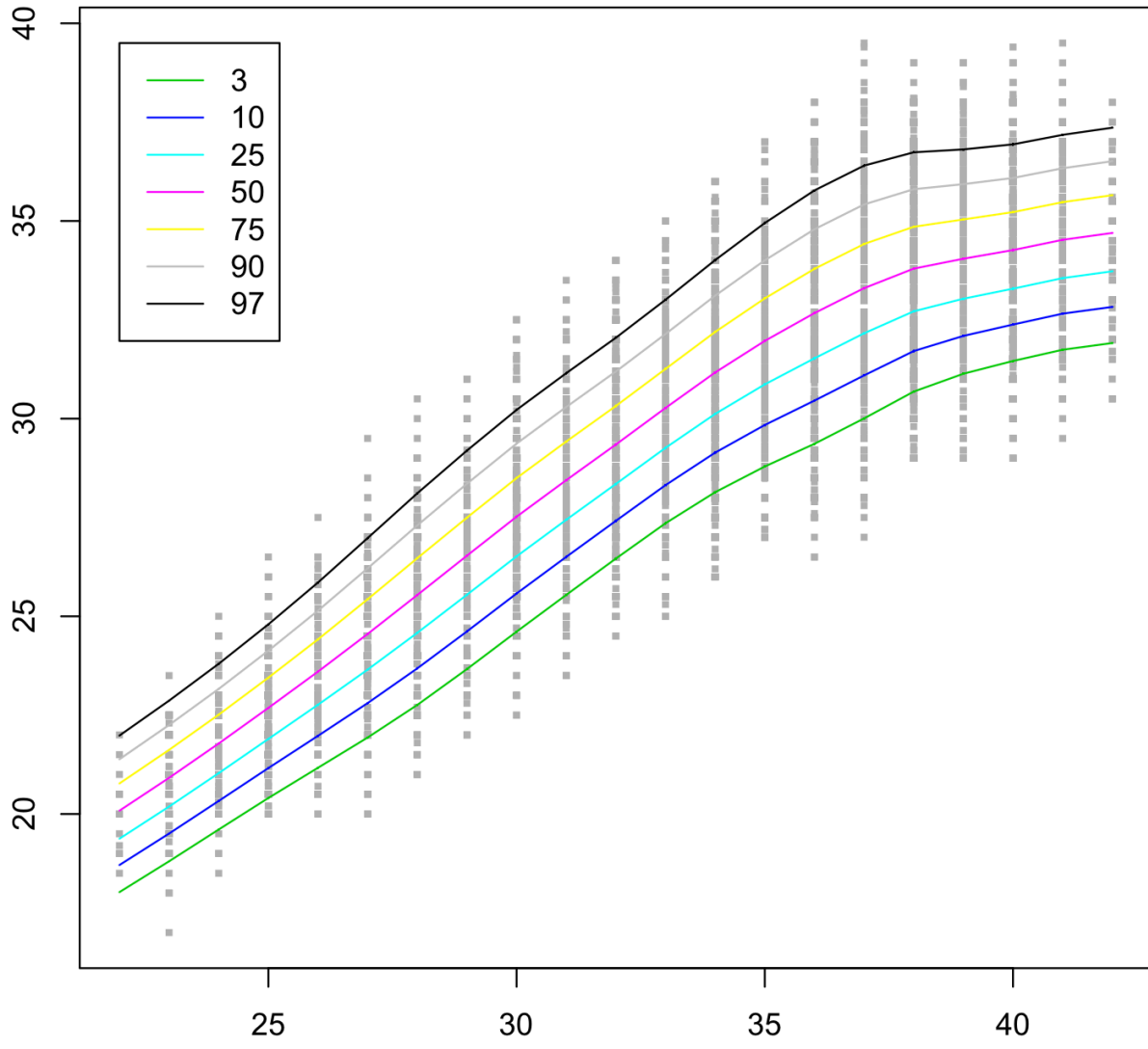
Note: The x axis represents gestational age measured in weeks and the y axis represents head

circumference measured in cm

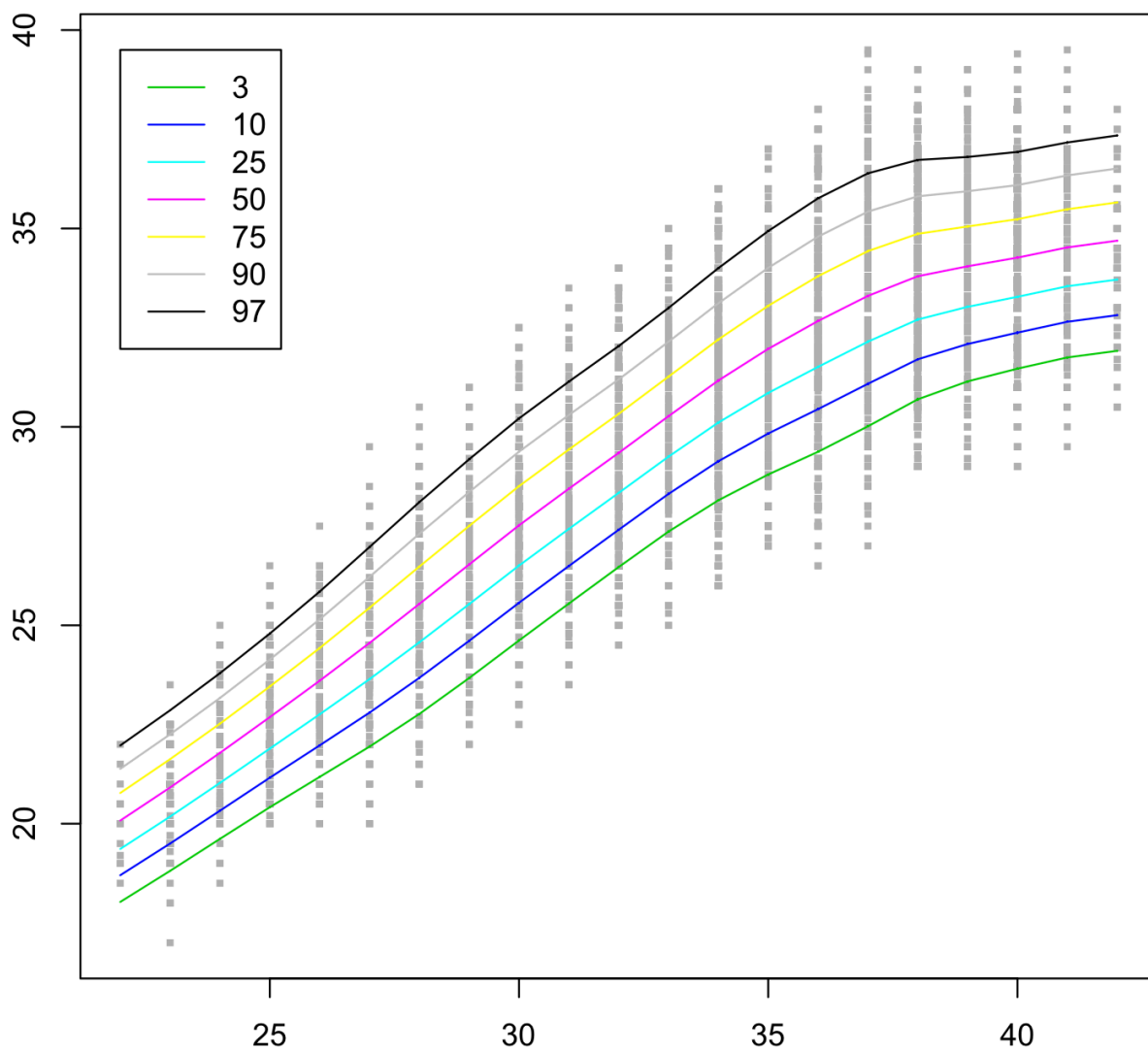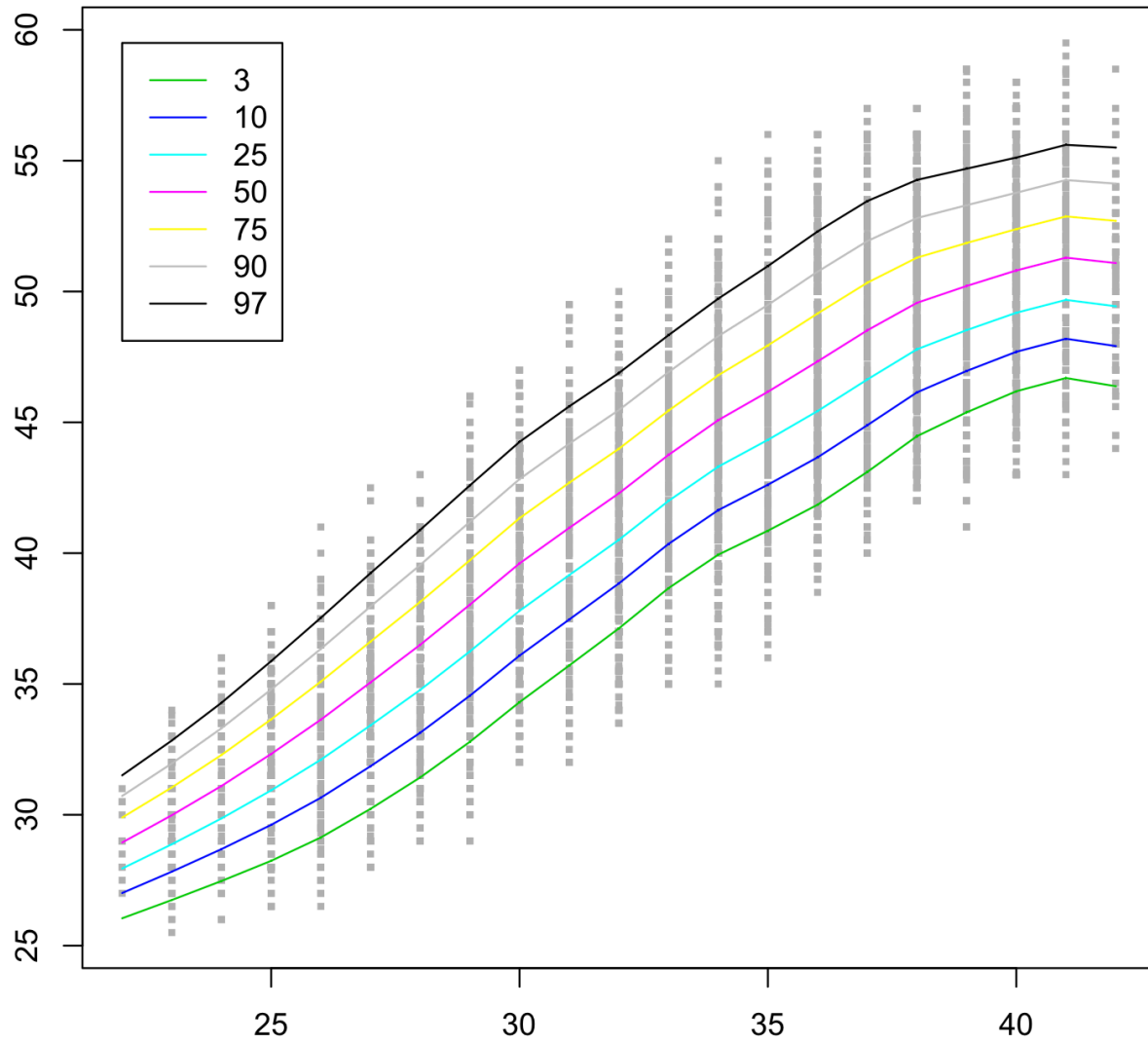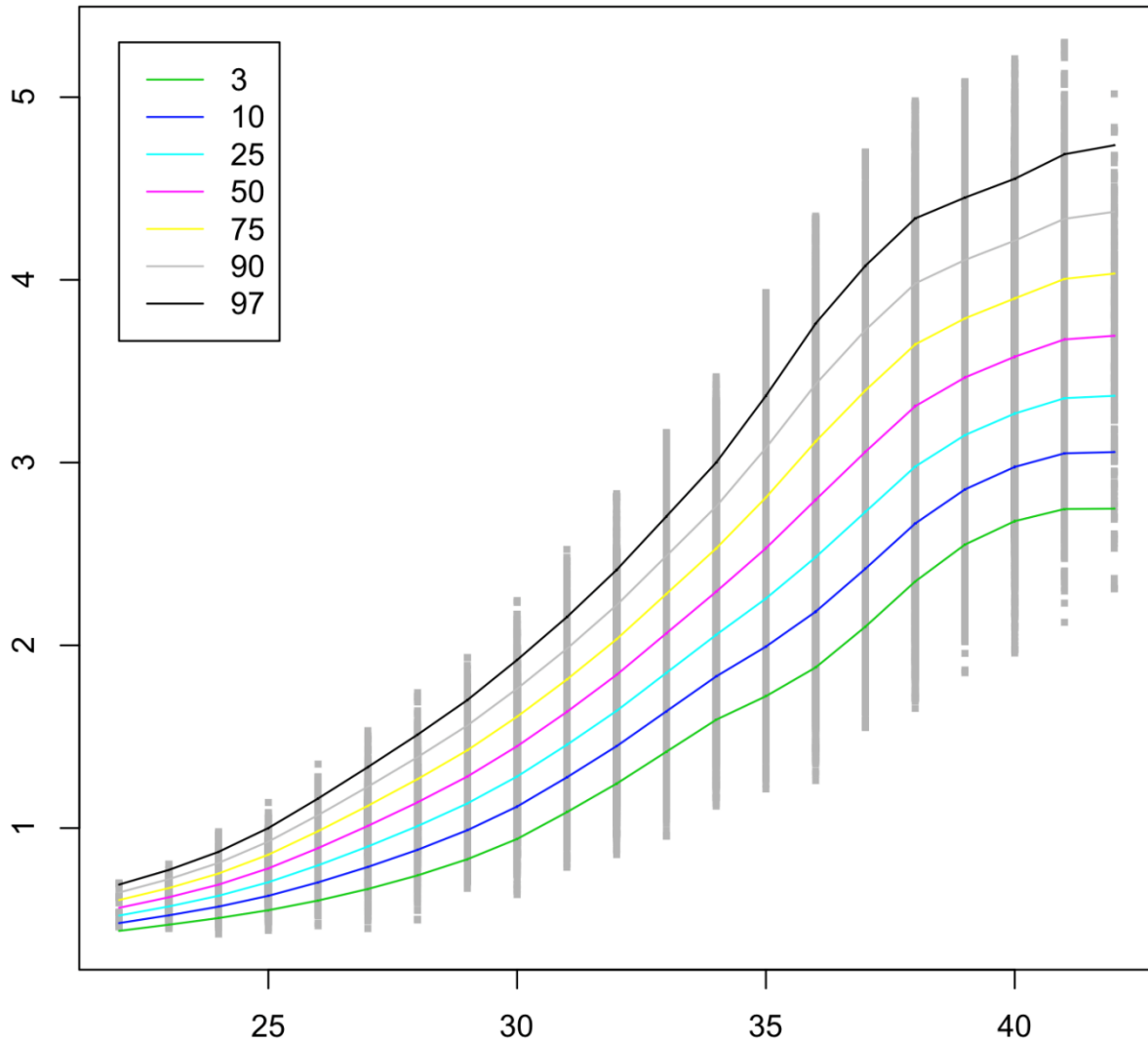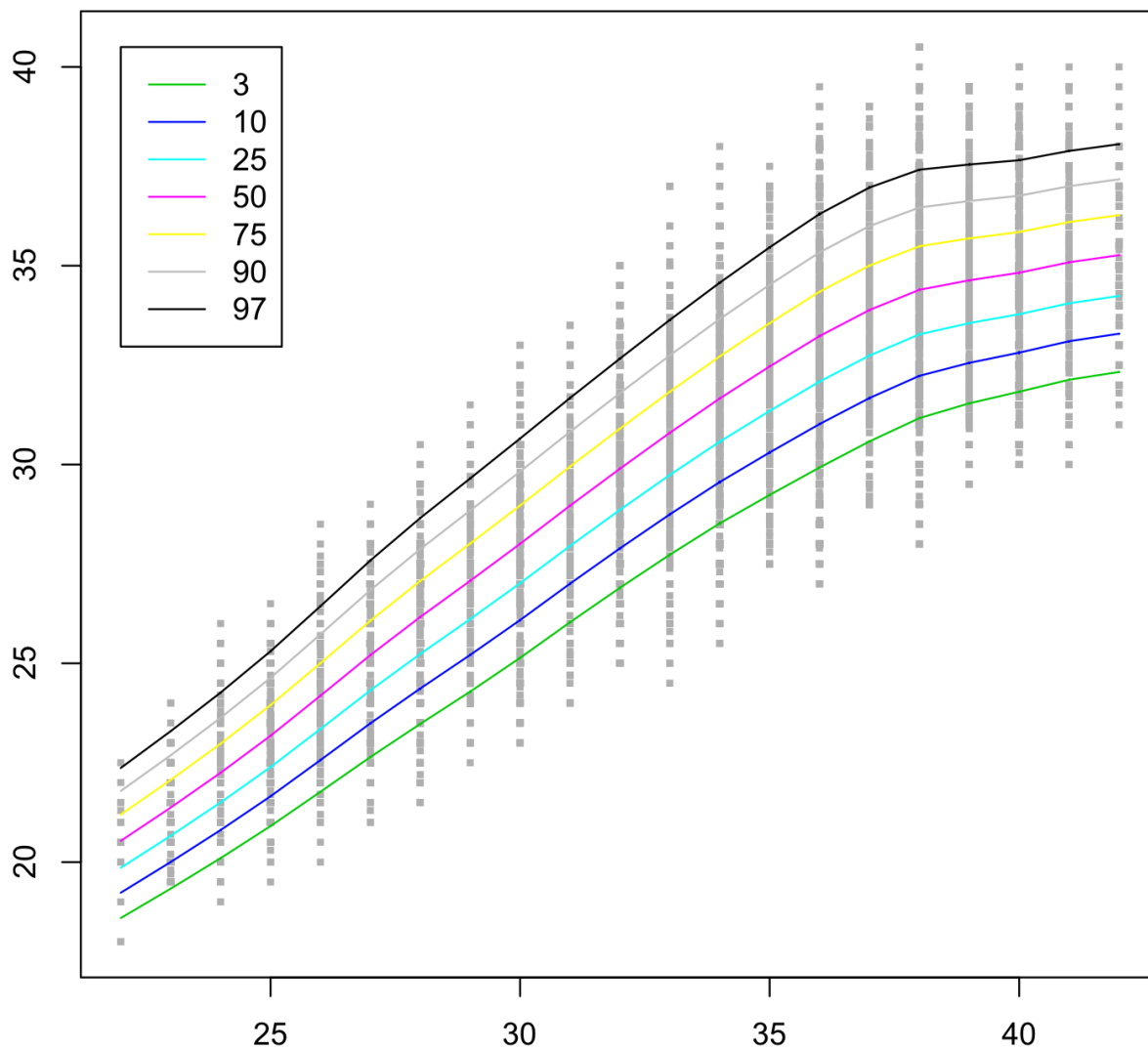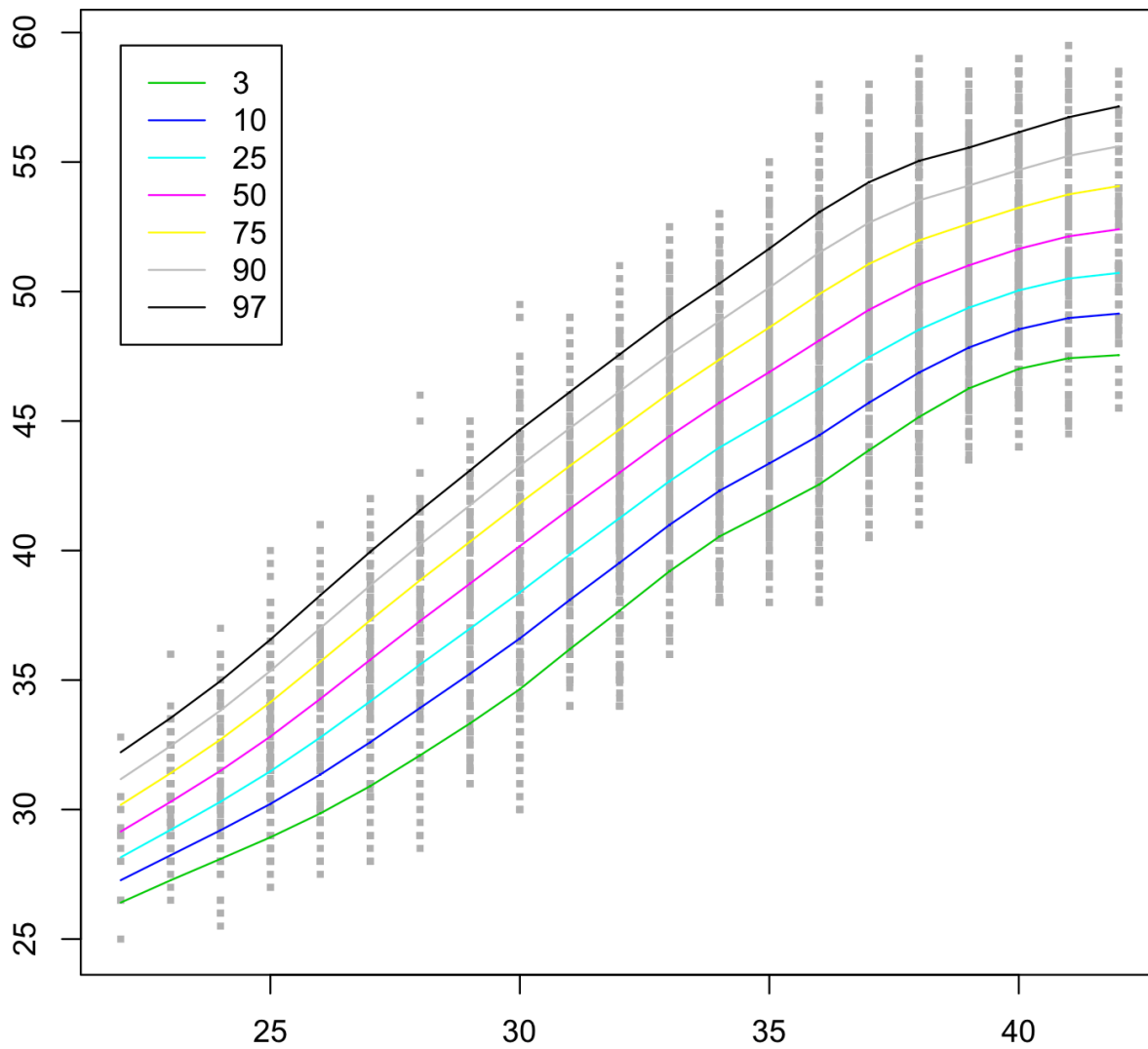Figure 12: Growth Curves from BCPE distribution for Preterm Infant Females—Birth Length by Gestational Age
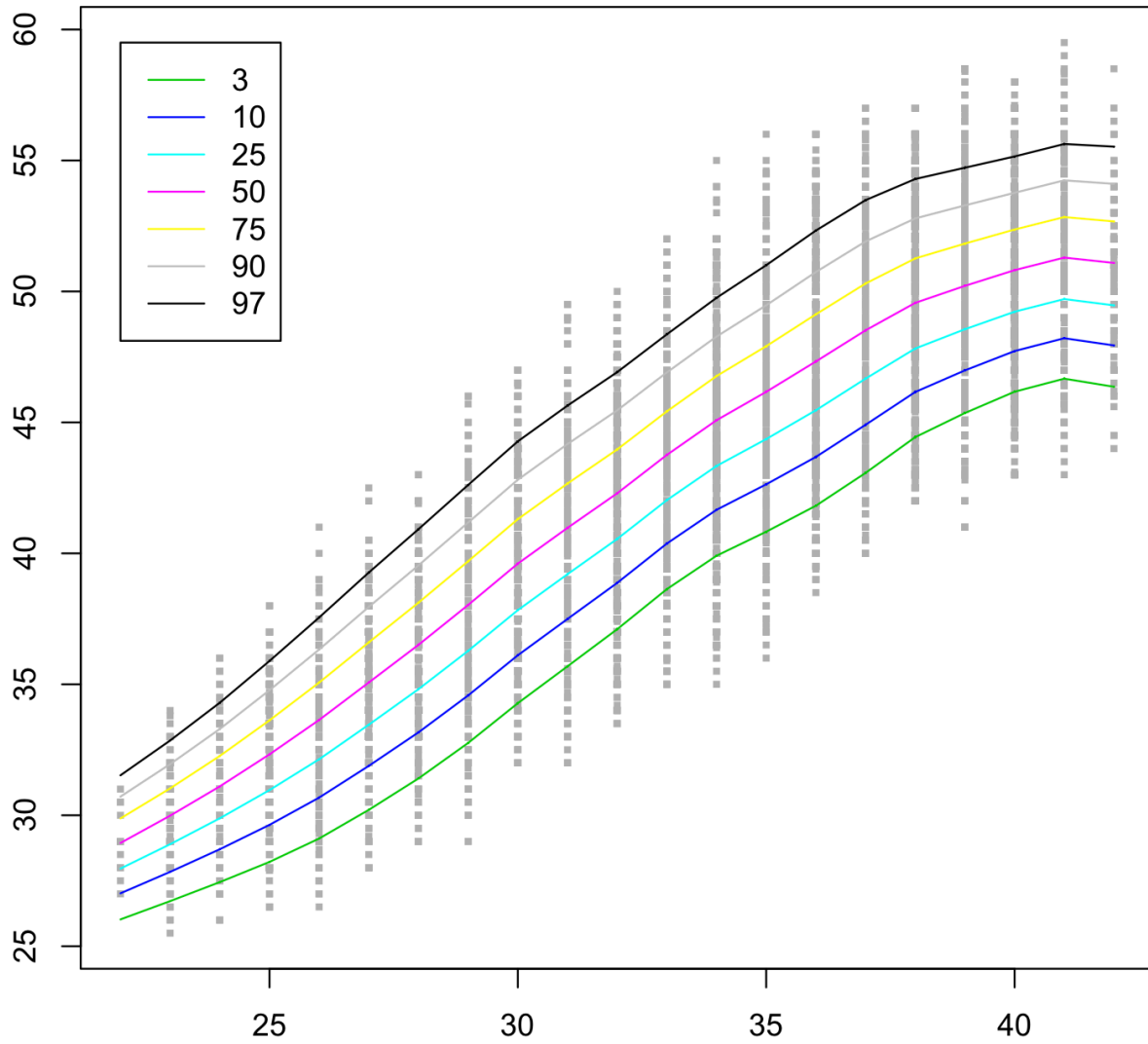
Note: The x axis represents gestational age measured in weeks and the y axis represents length circumference measured in cm

PAPER II: SGA AND LGA CLASSIFICATION FROM RACE-SPECIFIC AND NON-RACE-SPECIFIC GROWTH CURVES

The topic of the usefulness of race/ethnicity in the clinical setting is a continuing dialogue. In the Executive Summary of the World Health Organization's (WHO) Child Growth Standards Methods and Development, WHO asserted that the growth charts they produced were a tool that "can be used to assess children everywhere, regardless of ethnicity, socioeconomic status and type of feeding" (WHO, 2006, p. *xx*). WHO provided growth charts for males and females from birth to five years for a variety of anthropomorphic measurements that are frequently used by clinicians. Some countries adopted the WHO Child Growth Standards quickly and some were hesitant to use the new references.

The details of the WHO study are as follows: as a part of the Multicentre Growth Reference Study (MGRS), the WHO collected longitudinal and cross-sectional data from 6 developed countries for children ages 0 to 71 months between the years of 1997 to 2003 (deOnis, 2004). The aim of their study was to develop growth standards that represented the actual growth of healthy children who were exposed to optimal conditions. Participants were selected from Brazil, Ghana, India, Norway, Oman, and the United States. Only healthy mothers and children were considered for enrollment. Inclusion criteria for the mothers/families were: no economic, health, or environmental constraints on growth, mother willing to follow feeding guidelines, term birth, single birth, absence of significant health morbidity, and nonsmoking. The anthropomorphic measures collected by trained staff were weight, length, height, head and arm circumferences, triceps and subscapular skinfold thickness, and body mass index (BMI). A novel component of the study was the emphasis on inclusion of mothers who were primarily breastfeeding their infants.

The WHO put together a committee of experts to decide how the data should be analyzed. The committee members considered all available statistical methods for growth curve creation (the details of which can be found in Bhorgi et al., 2006). Briefly, after developing primary and secondary model/method selection criteria, the advisory group selected three models for testing from 30 methods. The final selected model was the Box-Cox Power Exponential distribution ([BCPE] Rigby & Stasinopouls, 2004) because of its flexibility to model skewness and kurtosis and produce centiles and z scores. Once data analysis began, the researchers used a reduced BCPE model (the LMS model [Cole & Green, 1992]) to create their final curves.

The efforts of WHO to develop growth references that could be used globally may have been methodologically sound but unrealistic. A notable criticism of the WHO Child Growth Standards is their lack of inclusion of children from Asian countries. A study of healthy 647 full-term Japanese babies was conducted to assess for potential differences between national-specific percentiles and the WHO growth standards (Tanaka et al., 2013). Researchers used second-order polynomial regression models to create sex-specific growth curves for length, weight, and head circumference of exclusively breastfed infants. Japanese infants were found to have lower mean percentiles for body weight and length. The authors state that they do not believe that the WHO growth standards are valid for Japanese children (Tanaka et al., 2013).

Similarly, a multilevel growth curve modeling study of breastfed Chinese infants indicated that Chinese infants were heavier and longer on average than the WHO growth standards (Huang et al., 2016). Using the WHO standards on Chinese infants would lead to an increase in the classification of overweight infants and a decreased classification of underweight. A sample of 160 healthy full-term Indonesian infants found were found to be lighter and smaller than the WHO growth standards (Dwipoerwantoro et al., 2015). The authors indicated that using

the WHO growth standards would lead to an increase in diagnosed underweight or stunted infants who are healthy but become assigned supplemental nutrition resources when it isn't necessary which could put a strain on government allocated resources for intervention with underweight children.

Natale and Rajagopalan (2015) remarked that WHO never released site-specific data for weight and head circumference from the MGRS and questioned whether the assumption that all children who experience unconstrained growth can be measured accurately on the same centiles. The authors conducted a systematic review of available data for countries and ethnicities (n = 78) to investigate whether significant differences existed between the mean weight, height, and head circumference from the WHO Child Growth References in samples of children from other countries/ethnicities. Researchers used the criteria of a sample mean being .5 SD different from the WHO reference mean to determine whether differences were significant.

Head circumferences varied the most, but researchers found significant differences in all measurements by race/ethnicity (Natale & Rajaopalan, 2015). For height 20% of the total means were considered significantly different. For weight, 30% of the total means were considered significantly different. Fifty-six percent of male means and 51% of females means were significantly different for head circumference.

The authors noted that Europeans represented those most often further than .5 SD in either direction for the anthropomorphic variables. The researchers conclude that WHO may have intended to measure optimal growth but for various complex reasons (including whether a population as a whole has reached the maximum height for that group) may have missed the mark (Natale & Rajaopalan, 2015).

Research involving race/ethnicity and infant growth is varied in both study focus and conclusions. Researchers have investigated the role of race primarily with a focus on predicting pregnancy outcomes or fetal outcomes. Usually a study will consider race when investigating if SGA/LGA categorization can predict health outcome. The results of these studies are mixed. Sometimes researchers find that only one or two ethnicities out of several may benefit from a race-specific perinatal mortality curves (Kierans et al., 2008), or that ethnic-specific curves seem to classify SGA/LGA babies better but may fail a test of clinical significance/usefulness (Norris et al., 2015).

In a recent study of all infants born alive in New York City from 2000-2010 (n = 984, 807), researchers investigated the association between race and infant outcome using log-binomial regression. Results indicated that infants of Black women were at a higher risk of having low birth weight, being delivered before term, and had an increased risk of infant mortality than infants of White women (Borrell, Rodriguez-Alvarez, Savitz, & Baquero, 2016). The risk of a Black couple having a preterm delivery was 2.1 times the risk of White couples. These results highlight the importance of considering parental race during pregnancy.

A compelling and thorough analysis of the association between race and infant growth was conducted by Buck Louis and colleagues (2015). The authors chose to investigate racial/ethnic patterns for estimated fetal weight (EFW) and measured birthweight. The inclusion criteria was strict in order to permit only healthy women with low-risk pregnancies to enroll in the study. Women self-identified their race (n = 481 non-Hispanic White, n = 428 non-Hispanic Black, n = 488 Hispanic, n = 342 Asian/Pacific Islander) and were randomized to 1 of 4 ultrasonography schedules There was one clinic that administered all sonographic exams. The statistical analysis included linear mixed models with cubic splines, likelihood ratio tests,

multiple imputations for missing data, and a comparison of different linear mixed models and smoothing methods (Buck Louis et al., 2015).

For EFW, curves differed significantly by race/ethnicity starting at 16 weeks gestational age. For head circumference, significant differences were detected by 21 weeks. When authors investigated misclassification rates of nonwhite fetuses for EFW they found significant misclassifications for Black, Hispanic, and Asian fetuses for the $<5^{th}$ percentile. These findings were the result of adequate sample sizes and a rigorous statistical analysis. They indicate that babies of different races may grow differently when in the womb (Buck Louis, 2015).

Clinical classifications of SGA and LGA are important to the care of newborns admitted to NICUs. Clinicians use growth charts to categorize infants into small-for-gestation age (SGA) and large-for-gestational age (LGA). Infants who are considered SGA are at risk of developing perinatal asphyxia, meconium aspiration, hypoglycemia, and hypothermia (Gibson & Nawab, 2015a); Infants who are LGA are at risk of developing respiratory distress, meconium aspiration, hypoglycemia, and polycythemia (Gibson & Nawab, 2015b). Infants risk being exposed to unnecessary testing or missed intervention opportunities when they are not able to be properly screened. Preterm infants are a vulnerable subgroup with few studies existing to examine how race-specific growth curves may change the validity of SGA and LGA classifications. I used the LMS method (Cole & Green, 1992) to generate race and non-race-specific growth curves and centile cut points from the training sub-sample of a large, racially diverse dataset of US infants.

My aim was to compare SGA and LGA age classification for growth curve centile cut points generated from these non-race specific curves and race specific curves on a validation dataset.

Method

Subjects

The data were provided by Pediatrix Medical Group, Inc. The de-identified birth data were collected from medical records from 33 NICU's in the US. The sample included 391,681 babies who were admitted to a NICU (in the Pediatrix network) between the years 1998-2006. The infant measurements were conducted by hospital staff at the time of the infant's birth.

As these datasets have been previously used in publication (Olsen, et al., 2010), they were already "cleaned" and ready for use. Briefly, implausible outliers were removed and babies with characteristics known to impact fetal growth (like congenital birth abnormalities) were removed (see Groveman, 2008 for full details). Table 1 lists the variables contained in the datasets. The original dataset was cleaned, separated by gender (male or female), and then separated further into random "training" and "validation" datasets.

For the purpose of my analysis, I further separated the training and validation datasets by race. I had six training files and six validation files: White males, Black males, and all other males, White females, Black females, and all other females. The "all other" males/females race category contained the information for infants with a labeled race of "Hispanic" or "other".

Non-race-specific curve and SGA/LGA classification generation

In R (R Core Team, 2017), I loaded the GAMLSS library. I used the `na.omit` function to delete missing observations because GAMLSS will not run with missing data. I used the GAMLSS call for birth length, head circumference, and weight (separately) with gestational age as the explanatory variable to generate LMS (3-parameter) distribution growth curves. I requested the estimated centiles for the 3rd, 10th, 25th, 50th, 75th, 90th, and 97th percentiles with the `centiles` function. To identify the **non-race-specific cut points** for the centiles I created a new variable in the form of a matrix with values 20-42 representing gestational ages weeks 20-42. I

then asked for the predicted values for the new variable numbers 20-42 using `centiles.pred`. I

transferred the predicted percentile cut points into a csv file.

For each race separately (Black, White, other) in SAS (SAS Institute Inc, 2013) I

imported the boys validation data and the **non-race-specific boys cut point data** for length. I

used proc sort to organize each file by gestational age. I merged the files and used an "if…do"

loop to categorize the validation data by centile using the centile cut points. I then used the `proc`

`freq` function to generate a frequency table of infants in each of the centiles. I imported the

frequency table into excel and calculated the percentage of observations below percentile 10 as

SGA and the percentage of observations greater than centile 90 as LGA. The SAS process was

then repeated for weight and head circumference.

I followed these same steps in R and SAS for girls.

Here is an example of my R code for the training dataset for boys length. A full collection

of all code can be found the Appendix 3.

```
boys<-read.table("desktop/BoysData_full.csv", header=TRUE, sep=",");
boys2<-na.omit(boys);
L_mBCCG_noT<- gamlss(BirthLength~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boys2);
centiles(L_mBCCG_noT,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
Mat5 <- centiles.pred(L_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat5;
```

Here is an example of the SAS code I used for the dataset merging and centile

categorization with an exported frequency table.

```
PROC IMPORT OUT= WORK.BoysB_V
          DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\Samantha va
lidboys.csv"
          DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.BoyslengthBCCG
          DATAFILE=
```

```
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysLengthBCCG.xlsx"
           DBMS=EXCEL REPLACE;
     RANGE="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=validboys; by gestage;
proc sort data=BoysLengthBCCG; by gestage;

data sam.boyslengthbccg(keep=gestage birthlength lengthile); merge
validboys boyslengthbccg; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then lengthile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then
lengthile=i;
end;
run;

PROC EXPORT DATA= SAM.Boyslengthbccg
           OUTFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_L
engthiles_BCCG.xlsx"
           DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Lengthile BCCG";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_L
engthiles_BCCG Freqs.csv";
proc freq data=SAM.Boyslengthbccg; tables lengthile; where gestage ne 22
and gestage ne 42;
run;
ods tagsets.csv close;
```

Race x Sex curve generation and SGA/LGA classification

I followed the same process as described above for the non-race-specific curve and

SGA/LGA classification generation apart from the first step. I used GAMLSS to generate LMS

curves for each race category (Black, White, and other) for birth length, head circumference, and

weight (separately) with gestational age as the explanatory variable. From that point I requested

estimated centiles, predicted values for weeks 20-42, and continued my analysis in SAS.

I repeated all of the above for females.

Results

The descriptive statistics for the training and validation datasets can be found in Tables 21-24.

Tables 25 and 27 show the results for the race-specific curve classifications. The ideal (or expected) percentage for SGA and LGA is 10%. The race-specific curves provided adequate fit for SGA and LGA classification for males and females. All classifications were within the acceptable range of 8—12%.

Table 26 shows the results for the non-race-specific curve classification for males. These curves did not provide adequate fit for Black babies. For example, SGA and LGA classification ranged between 5 and 18% for Black males. Table 28 indicates similar results for Black females, with SGA and LGA categorization ranging between 6 and 16%. Likewise, for length, and head circumference, non-race-specific curves did not provide adequate fit for any SGA or LGA categorization for Black males or females. Babies who were White or other race were classified as SGA or LGA within the acceptable range for all measures with the non-race specific curves, although the fit was not as good as with the race-specific curves

Discussion

The misclassification of preterm infants as appropriate-for-gestational age (AGA) when an infant is actually SGA or LGA may lead to a lack of intervention or health screenings; a misclassification of an infant of SGA or LGA when the infant is in fact AGA may trigger a cascade of unnecessary testing. Parents and clinicians want to spend time and healthcare efficiently on newborns. This analysis suggests that Black US preterm infants may be misclassified by using a non-race-specific growth curve.

The race-specific cut points generated in this study classified SGA and LGA babies in the validation dataset babies remarkably close to the expected percentage of 10% for each. The

differences in expected and observed classification rates is further from 10% but still within the acceptable range for White and other infants when measured against a non-race-specific growth curve. However, Black babies of both sexes would be overidentified as SGA and underidentified as LGA when using cut points from a non-race-specific curve. Because clinicians make decisions to provide interventions and/or conduct further testing based on SGA/LGA classifications, it is important to accurately identify these infants.

These results are in agreement with Buck Louis and colleagues (2015) who noted significant differences in head circumference in utero by race at 21 weeks' gestation. They found that Black and Asian fetuses/neonates had lowered weights than White and Hispanic fetuses/neonates. Both studies indicate that Black infants are at risk of being categorized as underweight using non-race-specific centiles.

The strengths of this study are the large sample size and the use of a validation dataset to test differences in observed and expected centile categorization of preterm infants. A limitation to this study is that the labels for race were included in medical records without a definition of how race was assessed. It could be that a nurse or doctor used his/her own judgement or race could have been recorded after asking the mother of the baby. There is no way to know if the definition of a "White" baby was consistent between NICUs.

Researchers may wish to investigate if the relationship between race and differences in SGA/LGA classification remains for healthy infants delivered at full term (>38 weeks gestational age).

References for Paper II

Bhorgi, E., de Onis, M., Garza, C., Van den Broeck, J., Frongillo, A., … & Martines, J. (2006). Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Statistics in Medicine*, *25*, 247—265. doi: 10.1002/sim.2227

Borrell, L., Rodriguez-Alvarez, E., Savitz, D., & Baquero, M. (2016). Parental Race/Ethnicity and Adverse Birth Outcomes in New York City: 2000–2010. *American Journal of Public Health*, *106*(8), 1491—1497. doi: 10.2105/ AJPH.2016.303242

Buck Louis, G., Grewal, J., Albert, P., Sciscione A., Wing, D., … & Grantz, K. (Racial/ethnic standards for fetal growth: the NICHD Fetal Growth Studies. *American Journal of Obstetrics and Gynecology, 213*(4), 1491—1497. Doi: 10.1016/j.ajog.2015.08.032.

Cole, T., & Green, P. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, *11*, 1305—1319. doi: 10.1002/sim.4780111005

Dwipoerwantoro, P., Mansyur, M., Oswari, H., Makrides, M., Cleghorn, G., & Firmansyah, A. (2015). Growth of Indonesian Infants Compared With World Health Organization Growth Standards. *Journal of Pediatric Gastroenterology*, *61*(2), 248—252. doi: 10.1097/MPG.0000000000000770.

Gibson, E., & Nawab, U. (2015a). *Small-for-Gestational Age (SGA) Infant*. Retrieved from https://www.merckmanuals.com/professional/pediatrics/perinatal-problems/small-for-gestational-age-sga-infant

Gibson, E., & Nawab, U. (2015b). *Large-for-Gestational Age (LGA) Infant*. Retrieved from

http://www.merckmanuals.com/professional/pediatrics/perinatal-problems/large-for-

gestational-age-lga-infant

Groveman, S. (2008). *New Preterm Infant Growth Curves Influence of Gender and Race on

Birth Size*. Retrieved from http://hdl.handle.net/1860/2877

Huang, X., Chang, J., Feng, W., Xul, Y., Xu, T., …, & Pan, X. (2016). Development of a New

Growth Standard for Breastfed Chinese Infants: What Is the Difference from the WHO

Growth Standards? *PLOS One, 11*(12), 1—18. doi: 10.1371/journal.pone.0167816

Kierans, W., Joseph, K., Luo, Z., Platt, R., Wilkins, R., & Kramer, M. (2008). Does one size fit

all? The case for ethnic-specific standards of fetal growth. *MBC Pregnancy and

Childbirth*, *8*(1), e1—e9. doi: 10.1186/1471-2393-8-1

de Onis, M., Garza, C., Victora, C., Onyango, A., Frongillo, E., & Martines, J. (2004). The WHO

Multicentre Growth Reference Study: Planning, study design, and methodology. *Food

and Nutrition Bulletin*, *25*(1), S15—S26. doi: 10.1177/15648265040251S103

Natale, V., & Rajagopalan, A. (2014). Worldwide variation in human growth and the World

Health Organization growth standards: a systematic review. *BMJ Open*, *4*, 1—11. doi:

10.1136/bmjopen-2013-003735

SAS Institute Inc (2013). SAS/ACCESS® 9.4 Interface to ADABAS: Reference. Cary, NC: SAS

Institute Inc.

Tanaka, H., Ishnii, H., Yamada, T., Akazawa, K., Nagata, S., & Yamashiro, Y. (2013). Growth

of Japanese breastfed infants compared to national references and World Health

Organization growth standards. *Acta Pædiatrica, 102,* 739—743. doi: 10.1111/apa.12262

WHO. (2006). *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, Switzerland: Author.

Tables for Paper II

Table 21: Descriptive Statistics for Males, Training Sample							
White, _N_ = 32,485							
Variable	**Unit**	**Min**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Gestational Age	Weeks	22	33	36	35.45	38	42
Birth Weight	kg	0.46	2.09	2.78	2.72	3.39	5.26
Birth Length	cm	25.5	44.5	48	47.16	51	59.5
Head Circumference	cm	19.5	31	33	32.53	35	40.5
Black, _N_ = 9,282							
Variable	**Unit**	**Min**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Gestational Age	Weeks	22	32	35	34.63	38	42
Birth Weight	kg	0.42	1.67	2.36	2.4	3.15	5.3
Birth Length	cm	25	42	46	45.28	49.5	58.5
Head Circumference	cm	19.5	29	31.8	31.05	33.7	40
Other, _N_ = 21,055							
Variable	**Unit**	**Min**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Gestational Age	Weeks	22	34	36	35.66	39	42
Birth Weight	kg	0.45	2.06	2.74	2.68	3.34	5.3
Birth Length	cm	26	44	48	47.02	51	58.5
Head Circumference	cm	18	30.5	33	32.27	34.5	40.5

Table 22: Descriptive Statistics for Females, Training Sample							
White, *N* = 23,488							
Variable	Unit	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Gestational Age	Weeks	22	33	35	35.17	38	42
Birth Weight	kg	0.38	1.89	1.57	2.54	3.23	4.98
Birth Length	cm	25.5	43	47	46.03	50	59.5
Head Circumference	cm	18	30	32.5	31.72	34	39.5
Black, *N* = 7,991							
Variable	Unit	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Gestational Age	Weeks	22	31	34	34.14	38	42
Birth Weight	kg	0.41	1.49	2.13	2.21	2.96	4.92
Birth Length	cm	26.5	40.5	45	44.05	48.5	58.4
Head Circumference	cm	18	28	31	30.24	33	39.5
Other, *N* = 15,540							
Variable	Unit	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Gestational Age	Weeks	22	33	36	35.55	39	42
Birth Weight	kg	0.4	1.94	2.61	2.56	3.23	5.03
Birth Length	cm	26	43	47	46.22	50	59
Head Circumference	cm	17	30	32.4	31.67	34	39.4

Table 23: Descriptive Statistics for Males, Validation Sample							
White, *N* = 32,197							
Variable	Unit	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Gestational Age	Weeks	22	34	36	35.49	38	42
Birth Weight	kg	0.42	2.11	2.8	2.74	3.4	5.22
Birth Length	cm	25	44.5	48	47.21	51	59.5
Head Circumference	cm	19.5	31	33	32.6	35	40
Black, *N* = 8,914							
Variable	Unit	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Gestational Age	Weeks	22	32	35	34.68	38	42
Birth Weight	kg	0.43	1.67	2.36	2.41	3.16	5.11
Birth Length	cm	26	42	46	45.35	49.5	58.5
Head Circumference	cm	20	29	31.5	31.09	34	40
Other, *N* = 20,425							
Variable	Unit	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Gestational Age	Weeks	22	34	36	35.72	39	42
Birth Weight	kg	0.4	2.07	2.75	2.67	3.33	5.25
Birth Length	cm	25.5	44.5	48	47.08	51	59.5
Head Circumference	cm	19.5	30.5	33	32.31	34.5	40

Table 24: Descriptive Statistics for Females, Validation Sample							
White, *N* = 23,325							
Variable	**Unit**	**Min**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Gestational Age	Weeks	23	33	35	35.22	38	42
Birth Weight	kg	0.38	1.9	2.57	2.55	2.24	5.14
Birth Length	cm	26	43	47	46.09	50	59
Head Circumference	cm	19	30	32.5	31.78	34	39
Black, *N* = 7,777							
Variable	**Unit**	**Min**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Gestational Age	Weeks	22	32	34	34.21	38	42
Birth Weight	kg	0.39	1.52	2.13	2.22	2.93	5.15
Birth Length	cm	26	40.5	45	44.18	48.5	57
Head Circumference	cm	18	28	31	30.28	33	39
Other, *N* = 14,976							
Variable	**Unit**	**Min**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Gestational Age	Weeks	22	33	36	35.65	39	42
Birth Weight	kg	0.4	1.94	2.62	2.57	3.23	5.01
Birth Length	cm	26.7	43.2	47	46.31	50	59
Head Circumference	cm	19	30	32.5	31.72	34	39.5

Table 25: **Race-specific** Curves Classification of Size on Validation Sample for Males					
Black		White		Other	
Head Circumference					
Category	Percentage	Category	Percentage	Category	Percentage
SGA	11	SGA	10	SGA	10
LGA	10	LGA	11	LGA	9
Birth Weight					
SGA	11	SGA	10	SGA	10
LGA	9	LGA	10	LGA	9
Length					
SGA	10	SGA	11	SGA	11
LGA	10	LGA	9	LGA	9
N = 10,610		*N* = 37,717		*N* = 24,518	

Table 26: **Non-Race-specific** Curves Classification of Size on Validation Sample for Males					
Black		White		Other	
Head Circumference					
Category	Percentage	Category	Percentage	Category	Percentage
SGA	18	SGA	8	SGA	11
LGA	5	LGA	12	LGA	8
Birth Weight					
SGA	15	SGA	8	SGA	11
LGA	6	LGA	12	LGA	8
Length					
SGA	14	SGA	8	SGA	11
LGA	6	LGA	12	LGA	8
N = 10,610		*N* = 37,717		*N* = 24,518	

Table 27: **Race-specific** Curves Classification of Size on Validation Sample for Females					
Black		White		Other	
Head Circumference					
Category	Percentage	Category	Percentage	Category	Percentage
SGA	10	SGA	10	SGA	10
LGA	9	LGA	9	LGA	8
Birth Weight					
SGA	10	SGA	10	SGA	10
LGA	9	LGA	9	LGA	9
Length					
SGA	10	SGA	10	SGA	10
LGA	10	LGA	10	LGA	9
N = 9,151		*N* = 27,518		*N* = 18,043	

Table 28: **Non-Race-specific** Curves Classification of Size on Validation Sample for Females					
Black		White		Other	
Head Circumference					
Category	Percentage	Category	Percentage	Category	Percentage
SGA	16	SGA	8	SGA	11
LGA	6	LGA	11	LGA	8
Birth Weight					
SGA	14	SGA	9	SGA	10
LGA	7	LGA	11	LGA	8
Length					
SGA	13	SGA	9	SGA	10
LGA	7	LGA	12	LGA	9
N = 9,151		*N* = 27,518		*N* = 18,043	

Conclusions

Traditional methods of model comparison indicate selecting the model with the lowest GAIC. The BCPE method curves always had a lower GAIC which can be explained due to the inclusion of the parameter to account for kurtosis. For clinical significance and practicality, the LMS method curves provided adequate fit (as defined by differences in the observed and expected percentiles from the curves to be no greater than 2 percentage points). Further, there was no difference in the centiles that were generated using both methods. The simpler models generated by the LMS method were retained for birth length, head circumference, and weight by sex with an explanatory variable of gestational age.

I used the LMS curves to produce cut points for males and females. Then I stratified the data by race and created race-specific cut points. When the cut points were imposed on a validation dataset and I checked the SGA and LGA classification, I found that race-specific curves classified babies within expected ranges. Once again I used the WHO acceptability of ± 2 percentage points to identify acceptable SGA and LGA classification rates. Non-race-specific curves overidentified Black babies as SGA and underidentified them as LGA.

The results of this analysis support that LMS curves and BCPE curves do not create different percentile estimates for preterm infants. While kurtosis may be present in the data used to create growth curves, modeling for it provides no advantage for the clinician. Clinicians may want to consider race when tracking the growth of preterm infants. This study supports that Black babies may present with different growth patterns that may cause clinicians to misclassify them as SGA too often and LGA not often enough. More research is required to test if this relationship persists for babies delivered at full term.

Strengths of this study include the large sample size and use of medical record data. Some limitations to this study include that it is not recorded how the race of the baby was determined in the medical record. Identifying race presents a challenge in clinical settings. However we do have these labels (however they were ascertained) and the sample size should prevent any single clinician or center misclassification from having undue influence. Further, any classification errors that may be present would be expected to mirror the classification error made in the population.

APPENDICES

Appendix 1: R Code for Paper I

```
boys<-read.table("desktop/BoysData_full.csv", header=TRUE, sep=",");
boys2<-na.omit(boys);

mBCPE_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCPEo,
data=boys2);
mBCPE_T<-lms(BirthWeight, GestAge, data=boys2, trans.x=T);
Tage=(boys2$GestAge)^(mBCPE_T$power);
mBCCG_T<- gamlss(BirthWeight~pb(Tage), sigma.formula=~pb(Tage),
nu.formula=~pb(Tage), tau.formula=~pb(Tage), family=BCCGo, data=boys2);
mBCCG_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boys2);
summary(mBCPE_T);
GAIC(mBCPE_noT, mBCPE_T, mBCCG_noT, mBCCG_T);
centiles(mBCPE_noT, xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
centiles(mBCPE_T, xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90, 97));
centiles(mBCCG_noT,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90, 97));
centiles(mBCCG_T,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90, 97));
mBCPE_T$mu.df;
mBCPE_T$sigma.df;
mBCPE_T$nu.df;
mBCPE_T$tau.df;
mBCPE_noT$mu.df;
mBCPE_noT$sigma.df;
mBCPE_noT$nu.df;
mBCPE_noT$tau.df;
mBCCG_noT$mu.df;
mBCCG_noT$sigma.df;
mBCCG_noT$nu.df;
mBCCG_noT$tau.df;
mBCCG_T$mu.df;
mBCCG_T$sigma.df;
mBCCG_T$nu.df;
mBCCG_T$tau.df;

 L_mBCPE_noT<- gamlss(BirthLength~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCPEo,
data=boys2);
L_mBCPE_T<-lms(BirthLength, GestAge, data=boys2, trans.x=T);
Tage=(boys2$GestAge)^(L_mBCPE_T$power);
L_mBCCG_T<- gamlss(BirthLength~pb(Tage), sigma.formula=~pb(Tage),
nu.formula=~pb(Tage), tau.formula=~pb(Tage), family=BCCGo, data=boys2);
L_mBCCG_noT<- gamlss(BirthLength~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boys2);
 summary(L_mBCPE_T);
centiles(L_mBCPE_noT,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
centiles(L_mBCPE_T,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90, 97));
```

```
centiles(L_mBCCG_noT,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
centiles(L_mBCCG_T,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90, 97));
GAIC(L_mBCPE_noT, L_mBCPE_T, L_mBCCG_noT, L_mBCCG_T);
L_mBCPE_T$mu.df;
L_mBCPE_T$sigma.df;
L_mBCPE_T$nu.df;
L_mBCPE_T$tau.df;
L_mBCPE_noT$mu.df;
L_mBCPE_noT$sigma.df;
L_mBCPE_noT$nu.df;
L_mBCPE_noT$tau.df;
L_mBCCG_noT$mu.df;
L_mBCCG_noT$sigma.df;
L_mBCCG_noT$nu.df;
L_mBCCG_noT$tau.df;
L_mBCCG_T$mu.df;
L_mBCCG_T$sigma.df;
L_mBCCG_T$nu.df;
L_mBCCG_T$tau.df;

HC_mBCT_noT<- gamlss(BirthHC~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCTo,
data=boys2);
HC_mBCT_T<-lms(BirthHC, GestAge, data=boys2, trans.x=T);
Tage=(boys2$GestAge)^(HC_mBCT_T$power);
HC_mBCCG_T<- gamlss(BirthHC~pb(Tage), sigma.formula=~pb(Tage),
nu.formula=~pb(Tage), tau.fomula=~pb(Tage), family=BCCGo, data=boys2);
HC_mBCCG_noT<- gamlss(BirthHC~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boys2);
summary(HC_mBCT_T);
HC_mBCPE_T<- gamlss(BirthHC~pb(Tage), sigma.formula=~pb(Tage),
nu.formula=~pb(Tage), tau.fomula=~pb(Tage), family=BCPEo, data=boys2);
HC_mBCPE_noT<- gamlss(BirthHC~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCPEo,
data=boys2);

GAIC(HC_mBCPE_noT, HC_mBCPE_T, HC_mBCCG_noT, HC_mBCCG_T, HC_mBCT_T,
HC_mBCT_noT);
centiles(HC_mBCPE_noT,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
centiles(HC_mBCPE_T,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
centiles(HC_mBCCG_noT,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
centiles(HC_mBCCG_T,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97)); centiles(HC_mBCT_noT,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75,
90, 97));
centiles(HC_mBCT_T,xvar=boys2$GestAge, cent=c(3, 10, 25, 50, 75, 90, 97));

HC_mBCPE_T$mu.df;
HC_mBCPE_T$sigma.df;
HC_mBCPE_T$nu.df;
HC_mBCPE_T$tau.df;
```

```
HC_mBCPE_noT$mu.df;
HC_mBCPE_noT$sigma.df;
HC_mBCPE_noT$nu.df;
HC_mBCPE_noT$tau.df;
HC_mBCCG_noT$mu.df;
HC_mBCCG_noT$sigma.df;
HC_mBCCG_noT$nu.df;
HC_mBCCG_noT$tau.df;
HC_mBCCG_T$mu.df;
HC_mBCCG_T$sigma.df;
HC_mBCCG_T$nu.df;
HC_mBCCG_T$tau.df;
HC_mBCT_noT$mu.df;
HC_mBCT_noT$sigma.df;
HC_mBCT_noT$nu.df;
HC_mBCT_noT$tau.df;
HC_mBCT_T$mu.df;
HC_mBCT_T$sigma.df;
HC_mBCT_T$nu.df;
HC_mBCT_T$tau.df;


girls<-read.table("desktop/GirlsData_full.csv", header=TRUE, sep=",");
girls2<-na.omit(girls);

  L_mBCPE_noT<- gamlss(BirthLength~pb(GestAge),
sigma.formula=~pb(GestAge), nu.formula=~pb(GestAge),
tau.fomula=~pb(GestAge), family=BCPEo, data=girls2);
  L_mBCPE_T<-lms(BirthLength, GestAge, data=girls2, trans.x=T);
  Tage=(girls2$GestAge)^(L_mBCPE_T$power);
  L_mBCCG_T<- gamlss(BirthLength~pb(Tage), sigma.formula=~pb(Tage),
nu.formula=~pb(Tage), tau.formula=~pb(Tage), family=BCCGo, data=girls2);
  L_mBCCG_noT<- gamlss(BirthLength~pb(GestAge),
sigma.formula=~pb(GestAge), nu.formula=~pb(GestAge),
tau.fomula=~pb(GestAge), family=BCCGo, data=girls2);
  summary(L_mBCPE_T)
centiles(L_mBCPE_noT,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
  centiles(L_mBCPE_T,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
  centiles(L_mBCCG_noT,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
  centiles(L_mBCCG_T,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
  GAIC(L_mBCPE_noT, L_mBCPE_T, L_mBCCG_noT, L_mBCCG_T);
L_mBCPE_T$mu.df;
L_mBCPE_T$sigma.df;
L_mBCPE_T$nu.df;
L_mBCPE_T$tau.df;
L_mBCPE_noT$mu.df;
L_mBCPE_noT$sigma.df;
L_mBCPE_noT$nu.df;
L_mBCPE_noT$tau.df;
L_mBCCG_noT$mu.df;
L_mBCCG_noT$sigma.df;
```

```
L_mBCCG_noT$nu.df;
L_mBCCG_noT$tau.df;
L_mBCCG_T$mu.df;
L_mBCCG_T$sigma.df;
L_mBCCG_T$nu.df;
L_mBCCG_T$tau.df;


  HC_mBCPE_noT<- gamlss(BirthHC~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCPEo,
data=girls2);
  HC_mBCPE_T<-lms(BirthHC, GestAge, data=girls2, trans.x=T);
  Tage=(girls2$GestAge)^(HC_mBCPE_T$power);
  HC_mBCCG_T<- gamlss(BirthHC ~pb(Tage), sigma.formula=~pb(Tage),
nu.formula=~pb(Tage), tau.fomula=~pb(Tage), family=BCCGo, data=girls2);
  HC_mBCCG_noT<- gamlss(BirthHC ~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=girls2);
  summary(HC_mBCPE_T)
  centiles(HC_mBCPE_noT,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
  centiles(HC_mBCPE_T,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
  centiles(HC_mBCCG_noT,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
  centiles(HC_mBCCG_T,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
GAIC(HC_mBCPE_noT, HC_mBCPE_T, HC_mBCCG_noT, HC_mBCCG_T);
HC_mBCPE_T$mu.df;
HC_mBCPE_T$sigma.df;
HC_mBCPE_T$nu.df;
HC_mBCPE_T$tau.df;
HC_mBCPE_noT$mu.df;
HC_mBCPE_noT$sigma.df;
HC_mBCPE_noT$nu.df;
HC_mBCPE_noT$tau.df;
HC_mBCCG_noT$mu.df;
HC_mBCCG_noT$sigma.df;
HC_mBCCG_noT$nu.df;
HC_mBCCG_noT$tau.df;
HC_mBCCG_T$mu.df;
HC_mBCCG_T$sigma.df;
HC_mBCCG_T$nu.df;
HC_mBCCG_T$tau.df;



mBCPE_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCPEo,
data=girls2);
  mBCPE_T<-lms(BirthWeight, GestAge, data=girls2, trans.x=T);
  Tage=(girls2$GestAge)^(mBCPE_T$power);
  mBCCG_T<- gamlss(BirthWeight~pb(Tage), sigma.formula=~pb(Tage),
nu.formula=~pb(Tage), tau.fomula=~pb(Tage), family=BCCGo, data=girls2);
```

```
   mBCCG_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=girls2);
   GAIC(mBCPE_noT, mBCPE_T, mBCCG_noT, mBCCG_T);
summary(mBCPE_T);
   centiles(mBCPE_noT, xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
   centiles(mBCPE_T, xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
   centiles(mBCCG_noT,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
   centiles(mBCCG_T,xvar=girls2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
mBCPE_T$mu.df;
mBCPE_T$sigma.df;
mBCPE_T$nu.df;
mBCPE_T$tau.df;
mBCPE_noT$mu.df;
mBCPE_noT$sigma.df;
mBCPE_noT$nu.df;
mBCPE_noT$tau.df;
mBCCG_noT$mu.df;
mBCCG_noT$sigma.df;
mBCCG_noT$nu.df;
mBCCG_noT$tau.df;
mBCCG_T$mu.df;
mBCCG_T$sigma.df;
mBCCG_T$nu.df;
mBCCG_T$tau.df;
```

Appendix 2: R Code for Paper II

```
*I segmented my dataset to make "GirlsDataRB" = Girls, reduced, black;
GirlsDataRW = Girls, reduced, white, and GirlsRO= Girls, not black and not
white;

                                White curves
girlsRW<-read.table("desktop/GirlsData_RW.csv", header=TRUE, sep=",");
girlsRW2<-na.omit(girlsRW);
newx<-seq(22,42,1);

  L_mBCCG_noT<- gamlss(BirthLength~pb(GestAge),
sigma.formula=~pb(GestAge), nu.formula=~pb(GestAge),
tau.fomula=~pb(GestAge), family=BCCGo, data=girlsRW2);
  centiles(L_mBCCG_noT,xvar=girlsRW2$GestAge, cent=c(3, 10, 25, 50, 75,
90, 97));
L_mBCCG_noT$mu.df;
L_mBCCG_noT$sigma.df;
L_mBCCG_noT$nu.df;
  centiles(L_mBCCG_noT,xvar=girlsRW2$GestAge, cent=c(3, 10, 25, 50, 75,
90, 97));
Mat7 <- centiles.pred(L_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat7;

HC_mBCCG_noT<- gamlss(BirthHC ~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=girlsRW2);
  centiles(HC_mBCCG_noT,xvar=girlsRW2$GestAge, cent=c(3, 10, 25, 50, 75,
90, 97));
HC_mBCCG_noT$mu.df;
HC_mBCCG_noT$sigma.df;
HC_mBCCG_noT$nu.df;
 Mat8 <- centiles.pred(HC_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat8;

  mBCCG_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=girlsRW2);
  centiles(mBCCG_noT,xvar=girlsRW2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
mBCCG_noT$mu.df;
mBCCG_noT$sigma.df;
mBCCG_noT$nu.df;
mat9 <- centiles.pred(mBCCG_noT, xname="GestAge", xvalues=newx, cent=c(3,
10, 25, 50, 75, 90, 97) ); mat9;

                                Black curves

girlsRB<-read.table("desktop/GirlsData_RB.csv", header=TRUE, sep=",");
girlsRB2<-na.omit(girlsRB);

  L_mBCCG_noT<- gamlss(BirthLength~pb(GestAge),
sigma.formula=~pb(GestAge), nu.formula=~pb(GestAge),
tau.fomula=~pb(GestAge), family=BCCGo, data=girlsRB2);
```

```
  centiles(L_mBCCG_noT,xvar=girlsRB2$GestAge, cent=c(3, 10, 25, 50, 75,
90, 97));
L_mBCCG_noT$mu.df;
L_mBCCG_noT$sigma.df;
L_mBCCG_noT$nu.df;
Mat10 <- centiles.pred(L_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat10;

  HC_mBCCG_noT<- gamlss(BirthHC ~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=girlsRB2);
  centiles(HC_mBCPE_noT,xvar=girlsRB2$GestAge, cent=c(3, 10, 25, 50, 75,
90, 97));
HC_mBCCG_noT$mu.df;
HC_mBCCG_noT$sigma.df;
HC_mBCCG_noT$nu.df;
Mat11 <- centiles.pred(HC_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat11;

  mBCCG_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=girlsRB2);
  centiles(mBCCG_noT,xvar=girlsRB2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
mBCCG_noT$mu.df;
mBCCG_noT$sigma.df;
mBCCG_noT$nu.df;
mat12 <- centiles.pred(mBCCG_noT, xname="GestAge", xvalues=newx, cent=c(3,
10, 25, 50, 75, 90, 97) ); mat12;
                                Other Curves

girlsRO<-read.table("desktop/GirlsNBNW.csv", header=TRUE, sep=",");
girlsRO2<-na.omit(girlsRO);

  L_mBCCG_noT<- gamlss(BirthLength~pb(GestAge),
sigma.formula=~pb(GestAge), nu.formula=~pb(GestAge),
tau.fomula=~pb(GestAge), family=BCCGo, data=girlsRO2);
  centiles(L_mBCCG_noT,xvar=girlsRO2$GestAge, cent=c(3, 10, 25, 50, 75,
90, 97));
L_mBCCG_noT$mu.df;
L_mBCCG_noT$sigma.df;
L_mBCCG_noT$nu.df;
Mat13 <- centiles.pred(L_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) );
Mat13;

  HC_mBCCG_noT<- gamlss(BirthHC ~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=girlsRO2);
HC_mBCCG_noT$mu.df;
HC_mBCCG_noT$sigma.df;
HC_mBCCG_noT$nu.df;
Mat14 <- centiles.pred(HC_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat14;
```

```
   mBCCG_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=girlsRO2);
   centiles(mBCCG_noT,xvar=girlsRO2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
mBCCG_noT$mu.df;
mBCCG_noT$sigma.df;
mBCCG_noT$nu.df;
mat15 <- centiles.pred(mBCCG_noT, xname="GestAge", xvalues=newx, cent=c(3,
10, 25, 50, 75, 90, 97) ); mat15;


*I segmented my dataset to make "BoysDataRB" = Boys, reduced, black;
BoysDataRW = Boys, reduced, white, and BoysRO= Boys, not black and not
white;


                              White curves
boysRW<-read.table("desktop/BoysData_RW.csv", header=TRUE, sep=",");
boysRW2<-na.omit(boysRW);
newx<-seq(22,42,1);


   L_mBCCG_noT<- gamlss(BirthLength~pb(GestAge),
sigma.formula=~pb(GestAge), nu.formula=~pb(GestAge),
tau.fomula=~pb(GestAge), family=BCCGo, data=boysRW2);
   centiles(L_mBCCG_noT,xvar=boysRW2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
L_mBCCG_noT$mu.df;
L_mBCCG_noT$sigma.df;
L_mBCCG_noT$nu.df;
   centiles(L_mBCCG_noT,xvar=boysRW2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
Mat16 <- centiles.pred(L_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat16;

HC_mBCCG_noT<- gamlss(BirthHC ~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boysRW2);
   centiles(HC_mBCCG_noT,xvar=boysRW2$GestAge, cent=c(3, 10, 25, 50, 75,
90, 97));
HC_mBCCG_noT$mu.df;
HC_mBCCG_noT$sigma.df;
HC_mBCCG_noT$nu.df;
 Mat17 <- centiles.pred(HC_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat17;

   mBCCG_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boysRW2);
   centiles(mBCCG_noT,xvar=boysRW2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
mBCCG_noT$mu.df;
mBCCG_noT$sigma.df;
mBCCG_noT$nu.df;
mat18 <- centiles.pred(mBCCG_noT, xname="GestAge", xvalues=newx, cent=c(3,
10, 25, 50, 75, 90, 97) ); mat18;
```

Black curves

```
boysRB<-read.table("desktop/BoysData_RB.csv", header=TRUE, sep=",");
boysRB2<-na.omit(boysRB);

  L_mBCCG_noT<- gamlss(BirthLength~pb(GestAge),
sigma.formula=~pb(GestAge), nu.formula=~pb(GestAge),
tau.fomula=~pb(GestAge), family=BCCGo, data=boysRB2);
  centiles(L_mBCCG_noT,xvar=boysRB2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
L_mBCCG_noT$mu.df;
L_mBCCG_noT$sigma.df;
L_mBCCG_noT$nu.df;
Mat19 <- centiles.pred(L_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat19;

  HC_mBCCG_noT<- gamlss(BirthHC ~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boysRB2);
  centiles(HC_mBCPE_noT,xvar=boysRB2$GestAge, cent=c(3, 10, 25, 50, 75,
90, 97));
HC_mBCCG_noT$mu.df;
HC_mBCCG_noT$sigma.df;
HC_mBCCG_noT$nu.df;
Mat20 <- centiles.pred(HC_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat20;

  mBCCG_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boysRB2);
  centiles(mBCCG_noT,xvar=boysRB2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
mBCCG_noT$mu.df;
mBCCG_noT$sigma.df;
mBCCG_noT$nu.df;
mat21<- centiles.pred(mBCCG_noT, xname="GestAge", xvalues=newx, cent=c(3,
10, 25, 50, 75, 90, 97) ); mat21;
```

Other Curves

```
boysRO<-read.table("desktop/BoysNBNW.csv", header=TRUE, sep=",");
boysRO2<-na.omit(boysRO);

  L_mBCCG_noT<- gamlss(BirthLength~pb(GestAge),
sigma.formula=~pb(GestAge), nu.formula=~pb(GestAge),
tau.fomula=~pb(GestAge), family=BCCGo, data=boysRO2);
  centiles(L_mBCCG_noT,xvar=boysRO2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
L_mBCCG_noT$mu.df;
L_mBCCG_noT$sigma.df;
L_mBCCG_noT$nu.df;
Mat22 <- centiles.pred(L_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) );
Mat22;
```

```
  HC_mBCCG_noT<- gamlss(BirthHC ~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boysRO2);
HC_mBCCG_noT$mu.df;
HC_mBCCG_noT$sigma.df;
HC_mBCCG_noT$nu.df;
Mat23 <- centiles.pred(HC_mBCCG_noT, xname="GestAge", xvalues=newx,
cent=c(3, 10, 25, 50, 75, 90, 97) ); Mat23;

  mBCCG_noT<- gamlss(BirthWeight~pb(GestAge), sigma.formula=~pb(GestAge),
nu.formula=~pb(GestAge), tau.fomula=~pb(GestAge), family=BCCGo,
data=boysRO2);
  centiles(mBCCG_noT,xvar=boysRO2$GestAge, cent=c(3, 10, 25, 50, 75, 90,
97));
mBCCG_noT$mu.df;
mBCCG_noT$sigma.df;
mBCCG_noT$nu.df;
mat24 <- centiles.pred(mBCCG_noT, xname="GestAge", xvalues=newx, cent=c(3,
10, 25, 50, 75, 90, 97) ); mat24;
```

Appendix 3: SAS Code for Paper II

```
***********************************************GIRLS********************************
***********************;
*******************Length*************************************************;
*****BCCG;
PROC IMPORT OUT= WORK.GirlslengthBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_
Length_BCCG.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=validgirls; by gestage;
proc sort data=GirlsLengthBCCG; by gestage;

data sam.girlslengthbccg(keep=gestage birthlength lengthile); merge validgirls
girlslengthbccg; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then lengthile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then lengthile=i;
end;
run;

PROC EXPORT DATA= SAM.Girlslengthbccg
            OUTFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_L
engthiles_BCCG.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Lengthile BCCG";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_L
engthiles_BCCG Freqs.csv";
proc freq data=SAM.Girlslengthbccg; tables lengthile; where gestage ne 22 and gestage
ne 42;
run;
ods tagsets.csv close;

************BCPE***************************;
PROC IMPORT OUT= WORK.GirlsLengthBCPE
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_
Length_BCPEe.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Girls_Length_BCPE$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=validgirls; by gestage;
proc sort data=GirlsLengthBCPE; by gestage;

data sam.girlslengthbcpe(keep=gestage birthlength lengthile); merge validgirls
girlslengthbcpe; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
```

```
if birthlength ge c97 then lengthile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then lengthile=i;
end;
run;

PROC EXPORT DATA= SAM.Girlslengthbcpe
            OUTFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_L
engthiles_BCPE.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Lengthile BCPE";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_L
engthiles_BCPE Freqs.csv";
proc freq data=SAM.Girlslengthbcpe; tables lengthile; where gestage ne 22 and gestage
ne 42;
run;
ods tagsets.csv close;

***************************Head circumference***************************;
***BCCG;
PROC IMPORT OUT= WORK.GirlsHCBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_
HC_BCCGe.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Girls_HC_BCCG$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=validgirls; by gestage;
proc sort data=GirlsHCBCCG; by gestage;

data sam.girlsHCBCCG(keep=gestage birthHC centile); merge validgirls GirlsHCBCCG; by
gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.GirlsHCBCCG
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_BCCG.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Centile BCCG";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_BCCG
Freqs.csv";
proc freq data=SAM.GirlsHCBCCG; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;
****BCPE;
PROC IMPORT OUT= WORK.GirlsHCBCPE
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_
HC_BCPEe.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Girls_HC_BCPE$";
     GETNAMES=YES;
```

```
        MIXED=NO;
        SCANTEXT=YES;
        USEDATE=YES;
        SCANTIME=YES;
RUN;

proc sort data=validgirls; by gestage;
proc sort data=GirlsHCBCPE; by gestage;

data sam.girlsHCBCPE(keep=gestage birthHC centile); merge validgirls GirlsHCBCPE; by
gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.GirlsHCBCPE
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_BCPE.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="Girls Centile BCPE";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_BCPE
Freqs.csv";
proc freq data=SAM.GirlsHCBCPE; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;

***********************Weight*******************************************;
****BCCG;
PROC IMPORT OUT= WORK.GirlsWeightBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_
Weight_BCCGe.xlsx"
            DBMS=EXCEL REPLACE;
      RANGE="Girls_Weight_BCCG$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
RUN;

proc sort data=validgirls; by gestage;
proc sort data=GirlsWeightBCCG; by gestage;

data sam.girlsWeightBCCG(keep=gestage birthweight centile); merge validgirls
GirlsWeightBCCG; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.GirlsweightBCCG
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weightcentiles_BCCG.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="Girls Weight Centile BCCG";
RUN;
```

```
ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weightcentiles_BCCG
Freqs.csv";
proc freq data=SAM.GirlsWeightBCCG; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;


****BCPE;
PROC IMPORT OUT= WORK.GirlsWeightBCPE
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_
Weight_BCPEe.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Girls_Weight_BCPE$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;


proc sort data=validgirls; by gestage;
proc sort data=GirlsWeightBCPE; by gestage;

data sam.girlsWeightBCPE(keep=gestage birthweight centile); merge validgirls
GirlsWeightBCPE; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;


PROC EXPORT DATA= SAM.GirlsweightBCPE
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weightcentiles_BCPE.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Weight Centile BCPE";
RUN;


ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weightcentiles_BCPE
Freqs.csv";
proc freq data=SAM.GirlsWeightBCPE; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;

****************************************************************************************
***************
****************************************************************************************
*****************
**********************************************************************BOYS******************
***************;
PROC IMPORT OUT= WORK.BoysB_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\Samantha va
lidboys.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.BoyslengthBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysLengthBCCG.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
```

```
        SCANTEXT=YES;
        USEDATE=YES;
        SCANTIME=YES;
RUN;


proc sort data=validboys; by gestage;
proc sort data=BoysLengthBCCG; by gestage;

data sam.boyslengthbccg(keep=gestage birthlength lengthile); merge validboys
boyslengthbccg; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then lengthile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then lengthile=i;
end;
run;


PROC EXPORT DATA= SAM.Boyslengthbccg
            OUTFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_L
engthiles_BCCG.xlsx"
            DBMS=EXCEL LABEL REPLACE;
    SHEET="Boys Lengthile BCCG";
RUN;


ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_L
engthiles_BCCG Freqs.csv";
proc freq data=SAM.Boyslengthbccg; tables lengthile; where gestage ne 22 and gestage
ne 42;
run;
ods tagsets.csv close;


***********BCPE*************************;
PROC IMPORT OUT= WORK.BoysLengthBCPE
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysLengthBCPE.xlsx"
            DBMS=EXCEL REPLACE;
    RANGE="sheet1$";
    GETNAMES=YES;
    MIXED=NO;
    SCANTEXT=YES;
    USEDATE=YES;
    SCANTIME=YES;
RUN;


proc sort data=validboys; by gestage;
proc sort data=BoysLengthBCPE; by gestage;

data sam.boyslengthbcpe(keep=gestage birthlength lengthile); merge validboys
boyslengthbcpe; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then lengthile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then lengthile=i;
end;
run;


PROC EXPORT DATA= SAM.Boyslengthbcpe
            OUTFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_L
engthiles_BCPE.xlsx"
            DBMS=EXCEL LABEL REPLACE;
    SHEET="Boys Lengthile BCPE";
RUN;


ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_L
engthiles_BCPE Freqs.csv";
proc freq data=SAM.Boyslengthbcpe; tables lengthile; where gestage ne 22 and gestage
```

```
ne 42;
run;
ods tagsets.csv close;

***************************Head circumference***************************;
***BCCG;
PROC IMPORT OUT= WORK.BoysHCBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysHCBCCG.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=validboys; by gestage;
proc sort data=BoysHCBCCG; by gestage;

data sam.boysHCBCCG(keep=gestage birthHC centile); merge validboys BoysHCBCCG; by
gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.BoysHCBCCG
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_BCCG.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Centile BCCG";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_BCCG
Freqs.csv";
proc freq data=SAM.BoysHCBCCG; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
****BCT;
PROC IMPORT OUT= WORK.BoysHCBCT
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysHCBCT.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=validboys; by gestage;
proc sort data=BoysHCBCT; by gestage;

data sam.boysHCBCT(keep=gestage birthHC centile); merge validboys BoysHCBCT; by
gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;
```

```
PROC EXPORT DATA= SAM.BoysHCBCT
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_BCT.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Centile BCT";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_BCT
Freqs.csv";
proc freq data=SAM.BoysHCBCT; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
***BCPE;
PROC IMPORT OUT= WORK.BoysHCBCPE
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysHCBCPE.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=validboys; by gestage;
proc sort data=BoysHCBCPE; by gestage;

data sam.boysHCBCPE(keep=gestage birthHC centile); merge validboys BoysHCBCPE; by
gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.BoysHCBCPE
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_BCPE.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Centile BCPE";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_BCPE
Freqs.csv";
proc freq data=SAM.BoysHCBCPE; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
************************Weight*********************************************;
****BCCG;
PROC IMPORT OUT= WORK.BoysWeightBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysWeightBCCG.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=validboys; by gestage;
proc sort data=BoysWeightBCCG; by gestage;

data sam.boysWeightBCCG(keep=gestage birthweight centile); merge validboys
```

```
BoysWeightBCCG; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.BoysweightBCCG
             OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Weightcentiles_BCCG.xlsx"
             DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Weight Centile BCCG";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Weightcentiles_BCCG
Freqs.csv";
proc freq data=SAM.BoysWeightBCCG; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;


****BCPE;
PROC IMPORT OUT= WORK.BoysWeightBCPE
             DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysWeightBCPE.xlsx"
             DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=validboys; by gestage;
proc sort data=BoysWeightBCPE; by gestage;

data sam.boysWeightBCPE(keep=gestage birthweight centile); merge validboys
BoysWeightBCPE; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.BoysweightBCPE
             OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Weightcentiles_BCPE.xlsx"
             DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Weight Centile BCPE";
RUN;

ods tagsets.csv file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Weightcentiles_BCPE
Freqs.csv";
proc freq data=SAM.BoysWeightBCPE; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;



***************************************************Boys
Black*********************************************;
```

```
**********Length;
PROC IMPORT OUT= WORK.BoysB_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamBoysV_B.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.BoysBlength
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boysblength.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=BoysB_V; by gestage;
proc sort data=Boysblength; by gestage;

data sam.BoysB_V_L(keep=gestage birthlength centile); merge BoysB_V boysblength; by
gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.BoysB_V_L
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Lengthtiles_Black.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Black Centiles BCCG";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Lengthtiles_Black_Freqs.csv";
proc freq data=SAM.BoysB_V_L; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.BoysBhc
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysBhc.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=boysb_v; by gestage;
proc sort data=boysbhc; by gestage;

data sam.boys_V_hc(keep=gestage birthHC centile); merge boysb_V boysbhc; by gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;
```

```
PROC EXPORT DATA= SAM.Boys_V_hc
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_Black.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Black Centiles HC HCCG";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_Black_Freqs.csv";
proc freq data=SAM.boys_v_hc; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.BoysBweight
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysBweight.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=boysb_v; by gestage;
proc sort data=boysbweight; by gestage;

data sam.boysb_v_W(keep=gestage birthweight centile); merge boysb_v boysbweight; by
gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.boysb_v_w
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Weightcentiles_black.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Black Weight Centiles";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_weightcentiles_Black_Freqs.csv";
proc freq data=SAM.boysb_v_w; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;



*********************************************Boys
White***********************************;
**********Length;
PROC IMPORT OUT= WORK.BoysW_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamBoysV_W.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.BoysWlength
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysWlength.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
```

```
        SCANTEXT=YES;
        USEDATE=YES;
        SCANTIME=YES;
RUN;


proc sort data=BoysW_V; by gestage;
proc sort data=BoysWlength; by gestage;

data sam.BoysW_V_L(keep=gestage birthlength centile); merge BoysW_V boysWlength; by
gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;


PROC EXPORT DATA= SAM.BoysW_V_L
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Lengthtiles_White.xlsx"
            DBMS=EXCEL LABEL REPLACE;
        SHEET="Boys White Centiles BCCG";
RUN;


ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Lengthtiles_White_Freqs.csv";
proc freq data=SAM.BoysW_V_L; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.BoysWhc
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysWhc.xlsx"
            DBMS=EXCEL REPLACE;
        RANGE="sheet1$";
        GETNAMES=YES;
        MIXED=NO;
        SCANTEXT=YES;
        USEDATE=YES;
        SCANTIME=YES;
RUN;


proc sort data=boysW_v; by gestage;
proc sort data=boysWhc; by gestage;

data sam.boysW_V_hc(keep=gestage birthHC centile); merge boysW_V boysWhc; by gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;


PROC EXPORT DATA= SAM.BoysW_V_hc
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_White.xlsx"
            DBMS=EXCEL LABEL REPLACE;
        SHEET="Boys White Centiles HC HCCG";
RUN;


ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_White_Freqs.csv";
proc freq data=SAM.boysW_v_hc; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
**************Weight;
```

```
PROC IMPORT OUT= WORK.BoysWweight
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysWweight.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=boysW_v; by gestage;
proc sort data=boysWweight; by gestage;

data sam.boysW_v_W(keep=gestage birthweight centile); merge boysW_v boysWweight; by
gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.boysW_v_w
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Weightcentiles_white.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys White Weight Centiles";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_weightcentiles_White_Freqs.csv";
proc freq data=SAM.boysW_v_w; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;


*******************************Boys Other (not black, not
white)********************************;
**********Length;
PROC IMPORT OUT= WORK.BoysNBNW_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamBoysV_NBNW.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.BoysNBNWlength
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysNBNWlength.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=BoysNBNW_V; by gestage;
proc sort data=BoysNBNWlength; by gestage;

data sam.BoysNBNW_V_L(keep=gestage birthlength centile); merge BoysNBNW_V
boysNBNWlength; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
```

```
end;
run;

PROC EXPORT DATA= SAM.BoysNBNW_V_L
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Lengthtiles_Other.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="Boys Other Centiles BCCG";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Lengthtiles_Other_Freqs.csv";
proc freq data=SAM.BoysNBNW_V_L; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;
*******Head Circumference;

PROC IMPORT OUT= WORK.BoysNBNWhc
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysNBNWhc.xlsx"
            DBMS=EXCEL REPLACE;
      RANGE="sheet1$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
RUN;

proc sort data=boysNBNW_v; by gestage;
proc sort data=boysNBNWhc; by gestage;

data sam.boysNBNW_V_hc(keep=gestage birthHC centile); merge boysNBNW_V boysNBNWhc; by
gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.BoysNBNW_V_hc
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_Other.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="Boys Other Centiles HC HCCG";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_HCcentiles_Other_Freqs.csv";
proc freq data=SAM.boysNBNW_v_hc; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.BoysNBNWweight
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysNBNWweight.xlsx"
            DBMS=EXCEL REPLACE;
      RANGE="sheet1$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
RUN;
```

```
proc sort data=boysNBNW_v; by gestage;
proc sort data=boysNBNWweight; by gestage;

data sam.boysNBNW_v_W(keep=gestage birthweight centile); merge boysNBNW_v
boysNBNWweight; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.boysNBNW_v_w
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_Weightcentiles_Other.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Boys Other Weight Centiles";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_weightcentiles_Other_Freqs.csv";
proc freq data=SAM.boysNBNW_v_w; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;


**************************************************Girls
Black*********************************************;

**********Length;
PROC IMPORT OUT= WORK.GirlsB_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamGirlsV_B.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.GirlsBlength
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girlsblength.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=GirlsB_V; by gestage;
proc sort data=Girlsblength; by gestage;

data sam.GirlsB_V_L(keep=gestage birthlength centile); merge GirlsB_V girlsblength; by
gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.GirlsB_V_L
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Lengthtiles_Black.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Black Centiles BCCG";
```

```
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Lengthtiles_Black_Freqs.csv";
proc freq data=SAM.GirlsB_V_L; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.GirlsBhc
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\GirlsBhc.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;


proc sort data=girlsb_v; by gestage;
proc sort data=girlsbhc; by gestage;

data sam.girls_V_hc(keep=gestage birthHC centile); merge girlsb_V girlsbhc; by
gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.Girls_V_hc
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_Black.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Black Centiles HC HCCG";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_Black_Freqs.csv";
proc freq data=SAM.girls_v_hc; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.GirlsBweight
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\GirlsBweight.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;


proc sort data=girlsb_v; by gestage;
proc sort data=girlsbweight; by gestage;

data sam.girlsb_v_W(keep=gestage birthweight centile); merge girlsb_v girlsbweight; by
gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
```

```
run;

PROC EXPORT DATA= SAM.girlsb_v_w
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weightcentiles_black.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Black Weight Centiles";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_weightcentiles_Black_Freqs.csv";
proc freq data=SAM.girlsb_v_w; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;




*********************************************Girls
White***********************************;
**********Length;
PROC IMPORT OUT= WORK.GirlsW_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamGirlsV_W.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.GirlsWlength
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\GirlsWlength.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=GirlsW_V; by gestage;
proc sort data=GirlsWlength; by gestage;

data sam.GirlsW_V_L(keep=gestage birthlength centile); merge GirlsW_V girlsWlength; by
gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.GirlsW_V_L
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Lengthtiles_White.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls White Centiles BCCG";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Lengthtiles_White_Freqs.csv";
proc freq data=SAM.GirlsW_V_L; tables centile; where gestage ne 22 and gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.GirlsWhc
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\GirlsWhc.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
```

```
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=girlsW_v; by gestage;
proc sort data=girlsWhc; by gestage;

data sam.girlsW_V_hc(keep=gestage birthHC centile); merge girlsW_V girlsWhc; by
gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.GirlsW_V_hc
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_White.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls White Centiles HC HCCG";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_White_Freqs.csv";
proc freq data=SAM.girlsW_v_hc; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.GirlsWweight
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\GirlsWweight.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=girlsW_v; by gestage;
proc sort data=girlsWweight; by gestage;

data sam.girlsW_v_W(keep=gestage birthweight centile); merge girlsW_v girlsWweight; by
gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.girlsW_v_w
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weightcentiles_white.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls White Weight Centiles";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_weightcentiles_White_Freqs.csv";
proc freq data=SAM.girlsW_v_w; tables centile; where gestage ne 22 and gestage ne 42;
```

```
run;
ods tagsets.csv close;


******************************Girls Other (not black, not
white)******************************;
**********Length;
PROC IMPORT OUT= WORK.GirlsNBNW_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamGirlsV_NBNW.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.GirlsNBNWlength
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\GirlsNBNWlength.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=GirlsNBNW_V; by gestage;
proc sort data=GirlsNBNWlength; by gestage;

data sam.GirlsNBNW_V_L(keep=gestage birthlength centile); merge GirlsNBNW_V
girlsNBNWlength; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.GirlsNBNW_V_L
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Lengthtiles_Other.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Other Centiles BCCG";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Lengthtiles_Other_Freqs.csv";
proc freq data=SAM.GirlsNBNW_V_L; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;
*******Head Circumference;

PROC IMPORT OUT= WORK.GirlsNBNWhc
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\GirlsNBNWhc.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=girlsNBNW_v; by gestage;
proc sort data=girlsNBNWhc; by gestage;
```

```
data sam.girlsNBNW_V_hc(keep=gestage birthHC centile); merge girlsNBNW_V girlsNBNWhc;
by gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.GirlsNBNW_V_hc
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_Other.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Other Centiles HC HCCG";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HCcentiles_Other_Freqs.csv";
proc freq data=SAM.girlsNBNW_v_hc; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.GirlsNBNWweight
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\GirlsNBNWweight.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=girlsNBNW_v; by gestage;
proc sort data=girlsNBNWweight; by gestage;

data sam.girlsNBNW_v_W(keep=gestage birthweight centile); merge girlsNBNW_v
girlsNBNWweight; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.girlsNBNW_v_w
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weightcentiles_Other.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="Girls Other Weight Centiles";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_weightcentiles_Other_Freqs.csv";
proc freq data=SAM.girlsNBNW_v_w; tables centile; where gestage ne 22 and gestage ne
42;
run;
ods tagsets.csv close;




********************************************************************************
***********************
********************************************************************************
```

```
************************
**************************************************GENERAL BOYS ON Boys Black
**********************************************;

**********Length;
PROC IMPORT OUT= WORK.BoysB_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamBoysV_B.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.BoysLengthBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysLengthBCCG.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=BoysB_V; by gestage;
proc sort data=BoysLengthBCCG; by gestage;

data sam.Boys_BVal_GeneralLength(keep=gestage birthlength centile); merge BoysB_V
BoysLengthBCCG; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.Boys_BVal_GeneralLength
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_BVal_GeneralLength_.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="BoysGen on BlackVal Length";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_BVal_GeneralLength_Freqs.csv";
proc freq data=SAM.Boys_BVal_GeneralLength; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.BoysHCbccg
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysHCbccg.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=boysb_v; by gestage;
proc sort data=BoysHCbccg; by gestage;

data sam.Boys_BVal_GeneralHC(keep=gestage birthHC centile); merge boysb_V BoysHCbccg;
by gestage;
c0=0;
```

```
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.Boys_BVal_GeneralHC
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_BVal_GeneralHC.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="BoysGen on BlackVal HC";
RUN;


ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_BVal_GeneralHC_Freqs.csv";
proc freq data=SAM.Boys_BVal_GeneralHC; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.BoysWeightBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysWeightBCCG.xlsx"
            DBMS=EXCEL REPLACE;
      RANGE="sheet1$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
RUN;


proc sort data=boysb_v; by gestage;
proc sort data=BoysWeightBCCG; by gestage;

data sam.Boys_BVal_GeneralWeight(keep=gestage birthweight centile); merge boysb_v
BoysWeightBCCG; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.Boys_BVal_GeneralWeight
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_BVal_GeneralWeight.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="BoysGen on BlackVal Weight";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_BVal_GeneralWeight_Freqs.csv";
proc freq data=SAM.Boys_BVal_GeneralWeight; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
*****************************GENERAL BOYS ON BOYS WHITE Length;
PROC IMPORT OUT= WORK.BoysW_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamBoysV_W.csv"
            DBMS=CSV REPLACE;
      GETNAMES=YES;
      DATAROW=2;
RUN;

PROC IMPORT OUT= WORK.BoysLengthBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysLengthBCCG.xlsx"
```

```
            DBMS=EXCEL REPLACE;
      RANGE="sheet1$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
RUN;


proc sort data=BoysW_V; by gestage;
proc sort data=BoysLengthBCCG; by gestage;


data sam.Boys_WVal_GeneralLength(keep=gestage birthlength centile); merge BoysW_V
BoysLengthBCCG; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Boys_WVal_GeneralLength
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_WVal_GeneralLength.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="BoysGen on WhiteVal Length";
RUN;


ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_WVal_GeneralLength_Freqs.csv";
proc freq data=SAM.Boys_WVal_GeneralLength; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.BoysHCbccg
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysHCbccg.xlsx"
            DBMS=EXCEL REPLACE;
      RANGE="sheet1$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
RUN;
proc sort data=boysW_v; by gestage;
proc sort data=BoysHCbccg; by gestage;
data sam.Boys_WVal_GeneralHC(keep=gestage birthHC centile); merge boysW_V BoysHCbccg;
by gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Boys_WVal_GeneralHC
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_WVal_GeneralHC.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="BoysGen on WhiteVal HC";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_WVal_GeneralHC_Freqs.csv";
proc freq data=SAM.Boys_WVal_GeneralHC; tables centile; where gestage ne 22 and
gestage ne 42;
run;
```

```
ods tagsets.csv close;
*************Weight;
PROC IMPORT OUT= WORK.BoysWeightBCCG
             DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysWeightBCCG.xlsx"
             DBMS=EXCEL REPLACE;
      RANGE="sheet1$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
RUN;

proc sort data=boysW_v; by gestage;
proc sort data=BoysWeightBCCG; by gestage;
data sam.boys_WVal_GeneralWeight(keep=gestage birthweight centile); merge boysW_v
BoysWeightBCCG; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Boys_WVal_GeneralWeight
             OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_WVal_GeneralWeight.xlsx"
             DBMS=EXCEL LABEL REPLACE;
      SHEET="BoysGen on WhiteVal Weight";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_WVal_GeneralWeight_Freqs.csv";
proc freq data=SAM.Boys_WVal_GeneralWeight; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;




**************************General Boys on NotBlackNotWhite (defined as "Other");
PROC IMPORT OUT= WORK.BoysNBNW_V
             DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamBoysV_NBNW.csv"
             DBMS=CSV REPLACE;
      GETNAMES=YES;
      DATAROW=2;
RUN;

PROC IMPORT OUT= WORK.BoysLengthBCCG
             DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysLengthBCCG.xlsx"
             DBMS=EXCEL REPLACE;
      RANGE="sheet1$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
RUN;

proc sort data=BoysNBNW_V; by gestage;
proc sort data=BoysLengthBCCG; by gestage;

data sam.Boys_NBNWVal_GeneralLength(keep=gestage birthlength centile); merge
BoysNBNW_V BoysLengthBCCG; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
```

```
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Boys_NBNWVal_GeneralLength
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_NBNWVal_GeneralLength.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="BoysGen on NBNWVal Length";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_NBNWVal_GeneralLength_Freqs.csv";
proc freq data=SAM.Boys_NBNWVal_GeneralLength; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.BoysHCbccg
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysHCbccg.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;
proc sort data=boysNBNW_v; by gestage;
proc sort data=BoysHCbccg; by gestage;
data sam.Boys_NBNWVal_GeneralHC(keep=gestage birthHC centile); merge boysNBNW_V
BoysHCbccg; by gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Boys_NBNWVal_GeneralHC
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_NBNWVal_GeneralHC.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="BoysGen on NBNWVal HC";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_NBNWVal_GeneralHC_Freqs.csv";
proc freq data=SAM.Boys_NBNWVal_GeneralHC; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.BoysWeightBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\BoysWeightBCCG.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=boysNBNW_v; by gestage;
proc sort data=BoysWeightBCCG; by gestage;
data sam.boys_NBNWVal_GeneralWeight(keep=gestage birthweight centile); merge
```

```
boysNBNW_v BoysWeightBCCG; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Boys_NBNWVal_GeneralWeight
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_NBNWVal_GeneralWeight.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="BoysGen on NBNWVal Weight";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Boys_NBNWVal_GeneralWeight_Freqs.csv";
proc freq data=SAM.Boys_NBNWVal_GeneralWeight; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;

********************************************************************************
********
********************************************************************************
*
*********************Girls General on Race Specific************************;
**********Length;
PROC IMPORT OUT= WORK.GirlsB_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamGirlsV_B.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.GirlsLengthBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Length_BCCG.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=GirlsB_V; by gestage;
proc sort data=GirlsLengthBCCG; by gestage;

data sam.Girls_BVal_GeneralLength(keep=gestage birthlength centile); merge GirlsB_V
GirlsLengthBCCG; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;

PROC EXPORT DATA= SAM.Girls_BVal_GeneralLength
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_BVal_GeneralLength_.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="GirlsGen on BlackVal Length";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_BVal_GeneralLength_Freqs.csv";
proc freq data=SAM.Girls_BVal_GeneralLength; tables centile; where gestage ne 22 and
```

```
gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.GirlsHCbccg
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HC_bccge.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Girls_HC_BCCG$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;


proc sort data=girlsb_v; by gestage;
proc sort data=GirlsHCbccg; by gestage;

data sam.Girls_BVal_GeneralHC(keep=gestage birthHC centile); merge girlsb_V
GirlsHCbccg; by gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;


PROC EXPORT DATA= SAM.Girls_BVal_GeneralHC
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_BVal_GeneralHC.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="GirlsGen on BlackVal HC";
RUN;


ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_BVal_GeneralHC_Freqs.csv";
proc freq data=SAM.Girls_BVal_GeneralHC; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.GirlsWeightBCCG
            DATAFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weight_BCCGe.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Girls_Weight_BCCG$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;


proc sort data=girlsb_v; by gestage;
proc sort data=GirlsWeightBCCG; by gestage;

data sam.Girls_BVal_GeneralWeight(keep=gestage birthweight centile); merge girlsb_v
GirlsWeightBCCG; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;
```

```
PROC EXPORT DATA= SAM.Girls_BVal_GeneralWeight
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_BVal_GeneralWeight.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="GirlsGen on BlackVal Weight";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_BVal_GeneralWeight_Freqs.csv";
proc freq data=SAM.Girls_BVal_GeneralWeight; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
*****************************GENERAL GIRLS ON GIRLS WHITE Length;
PROC IMPORT OUT= WORK.GirlsW_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamGirlsV_W.csv"
            DBMS=CSV REPLACE;
      GETNAMES=YES;
      DATAROW=2;
RUN;

PROC IMPORT OUT= WORK.GirlsLengthBCCG
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Length_BCCG.xlsx"
            DBMS=EXCEL REPLACE;
      RANGE="sheet1$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
RUN;

proc sort data=GirlsW_V; by gestage;
proc sort data=GirlsLengthBCCG; by gestage;

data sam.Girls_WVal_GeneralLength(keep=gestage birthlength centile); merge GirlsW_V
GirlsLengthBCCG; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Girls_WVal_GeneralLength
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_WVal_GeneralLength.xlsx"
            DBMS=EXCEL LABEL REPLACE;
      SHEET="GirlsGen on WhiteVal Length";
RUN;

ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_WVal_GeneralLength_Freqs.csv";
proc freq data=SAM.Girls_WVal_GeneralLength; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.GirlsHCbccg
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HC_bccge.xlsx"
            DBMS=EXCEL REPLACE;
      RANGE="Girls_HC_BCCG$";
      GETNAMES=YES;
      MIXED=NO;
      SCANTEXT=YES;
      USEDATE=YES;
      SCANTIME=YES;
```

```
RUN;
proc sort data=girlsW_v; by gestage;
proc sort data=GirlsHCbccg; by gestage;
data sam.Girls_WVal_GeneralHC(keep=gestage birthHC centile); merge girlsW_V
GirlsHCbccg; by gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Girls_WVal_GeneralHC
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_WVal_GeneralHC.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="GirlsGen on WhiteVal HC";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_WVal_GeneralHC_Freqs.csv";
proc freq data=SAM.Girls_WVal_GeneralHC; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.GirlsWeightBCCG
            DATAFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weight_BCCGe.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="Girls_Weight_BCCG$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc sort data=girlsW_v; by gestage;
proc sort data=GirlsWeightBCCG; by gestage;
data sam.girls_WVal_GeneralWeight(keep=gestage birthweight centile); merge girlsW_v
GirlsWeightBCCG; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Girls_WVal_GeneralWeight
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_WVal_GeneralWeight.xlsx"
            DBMS=EXCEL LABEL REPLACE;
     SHEET="GirlsGen on WhiteVal Weight";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_WVal_GeneralWeight_Freqs.csv";
proc freq data=SAM.Girls_WVal_GeneralWeight; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;




***************************General Girls on NotBlackNotWhite (defined as "Other");
PROC IMPORT OUT= WORK.GirlsNBNW_V
            DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Data\SamGirlsV_NBNW.csv"
```

```
                    DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;

PROC IMPORT OUT= WORK.GirlsLengthBCCG
          DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Length_BCCG.xlsx"
          DBMS=EXCEL REPLACE;
     RANGE="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;


proc sort data=GirlsNBNW_V; by gestage;
proc sort data=GirlsLengthBCCG; by gestage;

data sam.Girls_NBNWVal_GeneralLength(keep=gestage birthlength centile); merge
GirlsNBNW_V GirlsLengthBCCG; by gestage;
c0=0;
array length {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthlength ge c97 then centile=8;
else do i=1 to 7; if length{i} le birthlength lt length{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Girls_NBNWVal_GeneralLength
          OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_NBNWVal_GeneralLength.xlsx"
          DBMS=EXCEL LABEL REPLACE;
     SHEET="GirlsGen on NBNWVal Length";
RUN;


ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_NBNWVal_GeneralLength_Freqs.csv";
proc freq data=SAM.Girls_NBNWVal_GeneralLength; tables centile; where gestage ne 22
and gestage ne 42;
run;
ods tagsets.csv close;
*******Head Circumference;
PROC IMPORT OUT= WORK.GirlsHCbccg
          DATAFILE= "C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_HC_bccge.xlsx"
          DBMS=EXCEL REPLACE;
     RANGE="Girls_HC_BCCG$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;
proc sort data=girlsNBNW_v; by gestage;
proc sort data=GirlsHCbccg; by gestage;
data sam.Girls_NBNWVal_GeneralHC(keep=gestage birthHC centile); merge girlsNBNW_V
GirlsHCbccg; by gestage;
c0=0;
array HC {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthHC ge c97 then centile=8;
else do i=1 to 7; if HC{i} le birthHC lt HC{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Girls_NBNWVal_GeneralHC
          OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_NBNWVal_GeneralHC.xlsx"
          DBMS=EXCEL LABEL REPLACE;
```

```
        SHEET="GirlsGen on NBNWVal HC";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_NBNWVal_GeneralHC_Freqs.csv";
proc freq data=SAM.Girls_NBNWVal_GeneralHC; tables centile; where gestage ne 22 and
gestage ne 42;
run;
ods tagsets.csv close;
**************Weight;
PROC IMPORT OUT= WORK.GirlsWeightBCCG
            DATAFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_Weight_BCCGe.xlsx"
            DBMS=EXCEL REPLACE;
    RANGE="Girls_Weight_BCCG$";
    GETNAMES=YES;
    MIXED=NO;
    SCANTEXT=YES;
    USEDATE=YES;
    SCANTIME=YES;
RUN;

proc sort data=girlsNBNW_v; by gestage;
proc sort data=GirlsWeightBCCG; by gestage;
data sam.girls_NBNWVal_GeneralWeight(keep=gestage birthweight centile); merge
girlsNBNW_v GirlsWeightBCCG; by gestage;
c0=0;
array weight {8} c0 c3 c10 c25 c50 c75 c90 c97;
if birthweight ge c97 then centile=8;
else do i=1 to 7; if weight{i} le birthweight lt weight{i+1}then centile=i;
end;
run;
PROC EXPORT DATA= SAM.Girls_NBNWVal_GeneralWeight
            OUTFILE=
"C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_NBNWVal_GeneralWeight.xlsx"
            DBMS=EXCEL LABEL REPLACE;
    SHEET="GirlsGen on NBNWVal Weight";
RUN;
ods tagsets.csv
file="C:\Users\Todd\Dropbox\Thesis\Cutpoints\Girls_NBNWVal_GeneralWeight_Freqs.csv";
proc freq data=SAM.Girls_NBNWVal_GeneralWeight; tables centile; where gestage ne 22
and gestage ne 42;
run;
ods tagsets.csv close;
```