A proposed extension to the WHO weight-for-age centile curves: Statistical Methods and Models Manual

Canadian Paediatric Endocrine Group
Groupe canadien d'endocrinologie pédiatrique

Atul Sharma MD, MSc, FRCP(C)
on behalf of the CPEG/GCEP committee for national growth charts
Jean-Pierre Chanoine MD FRCP(C), University of British Columbia
Elizabeth Cummings MD FRCP(C), Dalhousie University
Sarah Lawrence MD FRCP(C), University of Ottawa
Daniel Metzger MD FRCP(C), University of British Columbia
Mark Palmert MD, PhD, FRCP(C), University of Toronto
Celia Rodd MD, MSc, FRCP(C), McGill University

January 30, 2014

CPEG GCEP

Canadian Pediatric Endocrine Group

Groupe canadien d'endocrinologie pédiatrique

Figure 1: Canadian Paediatric Endocrine Group/Groupe canadien d'endocrinologie pédiatrique: http://cpeg-gcep.net

Contents

1 Summary **3**
- 1.1 Background . . . 3
- 1.2 Presentation . . . 5
- 1.3 Conclusion . . . 7

2 Weight-for-age **9**
- 2.1 Statistical analyses . . . 9
- 2.2 Data sources . . . 9
- 2.3 Exclusion criteria . . . 11
  - 2.3.1 'Outlying' heights-for-age . . . 11
  - 2.3.2 'Unhealthy' weights-for-height . . . 11
- 2.4 Boys: weight-for-height exclusions . . . 13
- 2.5 Girls: weight-for-height exclusions . . . 14
- 2.6 Power transformation of the time axis . . . 16
- 2.7 Optimal smoothing models . . . 18
  - 2.7.1 Optimal smoothing model, boys . . . 19
  - 2.7.2 Optimal smoothing model, girls . . . 20
  - 2.7.3 Applications . . . 21
- 2.8 Fitted model, girls . . . 22
- 2.9 Fitted model, boys . . . 25
- 2.10 Model Diagnostics, girls . . . 29
- 2.11 Model Diagnostics, boys . . . 33
- 2.12 Comparisons with WHO model on NCHS dataset . . . 41

3 Height-for-age **45**
- 3.1 Model identification . . . 46
- 3.2 Fitted model, boys . . . 48
- 3.3 Fitted model, girls . . . 50
- 3.4 Model Diagnostics, boys . . . 52
- 3.5 Model Diagnostics, girls . . . 56

4 BMI-for-age **61**
- 4.1 Model identification . . . 61
- 4.2 Fitted model, boys . . . 63

4.3 Fitted model, girls 66
4.4 Model Diagnostics, boys 69
4.5 Model Diagnostics, girls 73
4.6 Data and methods: short version 79

List of Figures

1 Canadian Paediatric Endocrine Group/Groupe canadien d'endocrinologie pédiatrique: http://cpeg-gcep.net ii

1.1 Representative CPEG/GCEP Growth Curves 6

2.1 Age distribution of N=22,917 subjects (11507 girls, 11410 boys, ages 1-24 years) from NCHS surveys (1963-1975), the so-called 'core data' for estimation of all WHO growth curves for ages 5-19 years. Although constructed with data from 1 to 24 years to minimize edge effects, WHO analysts also merged the core data with cross-sectional data on 18-71 month olds (n $\sim$8000) to smooth the transition between WHO *standard* curves (0-5 years) and WHO *reference* curves (5-19 years). 10
2.2 Exclusions for outlying heights-for-age, A) 14 girls, B) 8 boys . . 12
2.3 BCPE distribution fit to weight-for-height. For boys, df(μ)=13,df(σ)=6, df(ν)=3 and df(τ)=3. 14
2.4 BCPE distribution fit to weight-for-height. For girls, df(μ)=12, df(σ)=4, df(ν)=3 and df(τ=3). 15
2.5 Boys: global deviance as exponent λ of age transformation varied from 1.0 to 1.5, with nadir at λ=1.3 17
2.6 Figure 8: Girls: global deviance as exponent λ of age transformation varied from 1.0 to 1.5, with nadir at λ=1.225 18
2.7 Girls: Time evolution of LMS model parameters for girls. A) Median (μ or M), B) Coefficient of variation σ or S, C) Skew parameter ν or L (Box-Cox exponent) 21
2.8 Girls 2-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year 22
2.9 Girls smoothed centiles (2-19 years) vs WHO centiles (5-10 years). The core NCHS data was used for both analyses, but the WHO added 8000 additional observations with ages 18-71 months to smooth the transition between WHO and NCHS data. 23
2.10 Girls smoothed centiles (2-19 years) vs CDC centiles (2-19 years). 24
2.11 Time evolution of LMS model parameters for boys. A) Median (μ or M), B) Coefficient of variation σ or S, C) Skew parameter ν or L (Box-Cox exponent) 25

2.12 Boys 2-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year 27

2.13 Boys smoothed centiles (2-19 years) vs WHO centiles (5-10 years).The core NCHS data was used for both analyses, but the WHO added 8000 additional observations with ages 18-71 months to smooth the transition between WHO and NCHS data. 27

2.14 Boys smoothed centiles (2-19 years) vs CDC centiles (2-19 years). 28

2.15 Residual plots for assessment of model assumptions. With an adequate fit, residuals should be normally distributed with mean=0, SD=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals (deviations from straight line = deviations from normality) 30

2.16 Worm plot of model residuals. With an adequate model, detrended residuals should lie between the 2 dashed lines (95% confidence interval). The path of the smoothed curve (solid red) can identify specific model violations. Briefly, the best fit line will have an intercept of zero if the median is adequately modeled (μ), and its slope will be zero if the variance (σ) is modeled adequately. Similarly, a U-shaped curve suggests residual skew, and an S-shaped curve speaks to residual kurtosis. 31

2.17 The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, with both worm plots and Z statistics applied separately on each interval. This facilitates localization of so-called "model violations". 32

2.18 Residual plots for assessment of model assumptions (boys). With an adequate fit, residuals should be normally distributed with mean=0, SD=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals (deviations from straight line = deviations from normality) 34

2.19 Influence diagnostics: The 8 outlying observations identified in the residual plot (figure 2.18) were deleted and the model refitted. Solid colored lines represent smoothed centiles calculated with the full dataset; dashed black lines are after deletion. There appears to be no appreciable influence on the fitted model. . . . 35

2.20 Worm plot of model residuals (boys). With an adequate model, de-trended residuals should lie between the 2 dashed lines (95% confidence interval). 36

2.21 Worm plots by age: The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years and worm plots applied separately in each interval. There is some evidence for localized model violations. 37

2.22 Same as figure 2.21 except the 3-parameter BCCG (LMS) model has been replaced by the 4-parameter BCPE model (boys), with the introduction of a kurtosis parameter τ. Worm-plot evidence of kurtosis has largely resolved. 39

2.23 Smoothed centiles from BCCG model (solid colors) and BCPE model (dashed lines) applied to boys weight-for-age data. The 4-parameter BCPE model adjusts for kurtosis (non-normal tail frequencies) in the distribution of weight-for-age. Adjusting for kurtosis has little effect on the fitted model. 40

2.24 Comparison of smoothed centiles from 2 LMS models fitted to the same data: Boys, weight-for-age Solid color lines represent a model with hyperparameters λ=1.4, df(μ)=10, df(σ) = 8, df(ν)=5, and dashed lines represent a model with hyperparameters λ=1.3, df(μ)=13, df(σ)=8, df(ν)=5. In both cases, the models were identified through minimization of appropriate GAIC. Smoothed centiles from the two models are indistinguishable over a range of $\pm$ 2SD. 42

2.25 Comparison of smoothed centiles from 2 LMS models fitted to the same data: Girls, weight-for-age. Solid color lines represent a model with hyperparameters λ=1.3, df(μ)=10, df(σ) = 3, df(ν)=3, and dashed lines represent a model with hyperparameters λ=1.225, df(μ)=14, df(σ)=6, df(ν)=5. Smoothed centiles from the two models are indistinguishable over a range of $\pm$ 2SD. 43

3.1 Global deviance as exponent λ of age transformation varied from 1.0 to 1.5 47

3.2 Boys smoothed centiles (5-19 years). Based on n=11,402 boys aged 1-24 years, NCHS data 49

3.3 Boys 2-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year 49

3.4 Boys smoothed centiles vs WHO centiles (5-19 years). The core NCHS data were used for both analyses, but the WHO added 8306 additional observations (ages 18-71 months) to smooth the transition between WHO and NCHS data. 50

3.5 Girls smoothed centiles (5-19 years). Based on n=11,493 girls aged 1-24 years, NCHS data 51

3.6 Girls 5-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year 51

3.7 Girls smoothed centiles vs WHO centiles (5-19 years). The core NCHS data were used for both analyses, but the WHO added 8306 additional observations (ages 18-71 months) to smooth the transition between WHO and NCHS data. 52

3.8 Residual plots for assessment of model assumptions (boys). With an adequate fit, residuals should be normally distributed with mean=0, sd=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals 53
3.9 Worm plot of model residuals (boys). With an adequate model, de-trended residuals should lie between the 2 dashed lines (95% confidence interval). 54
3.10 Worm plot by age: The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, and worm plots applied separately in each interval. There is no evidence for localized model violations. 55
3.11 Residual plots for assessment of model assumptions (girls). With an adequate fit, residuals should be normally distributed with mean=0, sd=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals 57
3.12 Worm plot of model residuals (girls). With an adequate model, de-trended residuals should lie between the 2 dashed lines (95% confidence interval). The path of the smoothed curve (solid red) can identify specific model violations. 58
3.13 The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, with both worm plots and Z statistics applied separately in each interval. There is no evidence for localized model violations. 59

4.1 A) Girls, B) Boys: Global deviance as exponent λ of age transformation varied from 1.0 to 1.5 62
4.2 Boys smoothed centiles (5-19 years). Based on n=11,106 boys aged 1-24 years, NCHS data 64
4.3 Boys 5-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year 65
4.4 Boys smoothed centiles vs WHO centiles (5-19 years).The core NCHS data were used for both analyses, but the WHO added 8306 additional observations (ages 18-71 months) to smooth the transition between WHO and NCHS data. 65
4.5 Girls smoothed centiles (5-19 years). Based on n=11,193 girls aged 1-24 years, NCHS data 66
4.6 Girls 5-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year 67
4.7 Girls smoothed centiles vs WHO centiles (5-19 years).The core NCHS data were used for both analyses, but the WHO added 8306 additional observations (ages 18-71 months) to smooth the transition between WHO and NCHS data. 68

4.8 Residual plots for assessment of model assumptions (boys). With an adequate fit, residuals should be normally distributed with mean=0, sd=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals 70
4.9 Worm plot of model residuals (boys). With an adequate model, de-trended residuals should lie between the 2 dashed lines (95% confidence interval). The path of the smoothed curve (solid red) can identify specific model violations. 71
4.10 The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, and worm plots applied separately in each interval. Though mild, there is evidence for localized model violations. 72
4.11 Residual plots for assessment of model assumptions (girls). With an adequate fit, residuals should be normally distributed with mean=0, SD=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals 74
4.12 Worm plot of model residuals (girls). With an adequate model, de-trended residuals should lie between the 2 dashed lines (95% confidence interval). The path of the smoothed curve (solid red) can identify specific model violations. 75
4.13 The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, and worm plots applied separately in each interval. 76
4.14 Girls smoothed centiles (5-19 years) from the 4-parameter BCPE model with adjustment for kurtosis and the simpler 3-parameter LMS model without adjustment for kurtosis 79

Table 1: Abbreviations used in text

AIC	Akaiki Information Criteria measures model fit
BMI	Body Mass Index kg/m^2
BCCG	Box-Cox Cole-Green probability model ≡ LMS model
BCPE	Box-Cox Power Exponential ≡ Rigby-Stasinopoulos probability model
CDC	Centers for Disease Control
CPEG	Canadian Pediatric Endocrinology Group
GCEP	Groupe canadien d'endocrinologie pèdiatrique
GAIC(n)	Generalized Akaiki Information Criterion with penalty = n
LMS	3-parameter Box-Cox Cole-Green probability model
MGRS	World Health Organization (WHO) Multicenter Growth Reference Study
NCHS	National Center for Health Statistics
NHANES	National Health and Nutrition Examination Surveys
NHES	National Households Education Surveys
OMS	L'organization mondiale de la Santé
SD	standard deviation
WHO	World Health Organization

Chapter 1

Summary

1.1 Background

De Onis *et al*[1] presented height-for-age, BMI-for-age, and weight-for-age reference curves for *school-aged and adolescent children, aged 5-19 years.* Although still needed by clinicians who wish to follow height and weight concurrently, the weight-for-age curves did not extend beyond 10 years of age, a policy decision intended to emphasize BMI norms for this age group.

The 'core data' for this analysis was provided by the National Center for Health Statistics (NCHS), representing 22,917 children (11507 girls, 11410 boys) pooled from 3 sources: NHES Cycle II (6-11 years, 1963-65), NHES Cycle III (12-17 years, 1966-70), and NHANES Cycle I (1-24 years, 1971-75)[1]. To derive smoothed WHO curves for 5-19 years of age, these data were merged with cross-sectional data from from the WHO Multicenter Growth Reference Study (MGRS, n∼8000, ages 18-71 months). The addition of younger children from 6 countries was intended to smooth the transition between the two datasets at age 5 years [2, 3]. The resulting curves are described by their developers as *reference* curves, distinct from the *standard* curves for 0-5 years of age based on MGRS data collected prospectively in the 1990s.

For school-age and adolescent subjects, the new WHO reference curves differ substantially from those promulgated by the Centers for Disease Control (2000)[4]. Since both groups applied similar statistical methodologies to overlapping datasets, these differences are in large part attributable to the different exclusion criteria used to define the reference populations, with the WHO dropping approximately 3% of subjects with "unhealthy" weights-for-height before fitting their smoothed centiles [3].

In the following manual, we outline our efforts to apply the WHO criteria to the publicly available data from NCHS, the 22,917 subjects between 1-24 years of age, which were provided to CPEG by Dr. Mercedes de Onis and the WHO. In this, we are guided by the statistical methodology described in de

[1]full list of abbreviations is found in table 1

Onis *et al*[1, 2, 5] and in the WHO Methods and Development Technical Report (2006)[3]. These reports generated smoothed centiles based on a 4-parameter probability model, the Box Cox Power Exponential or BCPE distribution, whose parameters can be used to calculate centiles or z-scores for any age[6]. Fitting this model is an iterative process that proceeds stepwise:

- The x-axis may first require a power transform to spread out the time axis and better capture periods of rapid growth. The optimal power λ is determined by minimization of global deviance.
- At each point on the time axis, a probability distribution is identified, characterized by the 4 BCPE parameters, namely μ (median), σ (coefficient of variation), ν (skew), and τ (a measure of kurtosis). A simpler 3 parameter LMS model omits τ in the absence of significant kurtosis.
- The time-evolution of each parameter is then smoothed using cubic splines with smoothing parameters (degrees of freedom df) dictated by the need to balance accurate prediction of sample centiles with a smooth representation. Selection of the optimal degrees of freedom for each parameter is determined by minimization of the Generalized Akaiki Information Criterion (GAIC) with an adjustable penalty term.
- In practice, the hyperparameters λ, df(μ), df(σ), and df(ν) define the smoothing model applied to the reference population to calculate best-fit LMS parameters (ν, μ, σ) by age[7]. The LMS parameters are in turn used to calculate percentiles or z-scores at each age (see section 2.7.3)

For weight, height, and BMI for age curves, it was noted that τ could be fixed at 2 for all curves ([1, 2, 3]). That is to say that kurtosis could be ignored, with the BCPE model now equivalent to the simpler 3-parameter Box-Cox Cole-Green (BCCG) or LMS model[7]. The latter model is defined by 3 parameters μ (median), σ (coefficient of variation), and ν (skew). For weight-for-age curves in school-aged children and adolescents, optimal WHO model parameters were λ=1.4, df(μ)=10, df(σ)=8, df(ν)=5 for boys and λ=1.3, df(μ)=10, df(σ)=3, and df(ν)=3 for girls.

In the following report, we apply both the WHO exclusion criteria and modeling procedure to the core data from NCHS to extend weight-for-age norms to ages 10-19 years. In each case, the optimal WHO models were used as a *starting point* for identification of the model best suited to the available data. Although we expect that our final models will be close to the WHO optima, differences in the dataset can be expected to yield slightly different results. For example, the optimal power transform of the time axis was identified through minimizing the global deviance[3]: This yielded λ=1.3 for boys and $\lambda = 1.225$ for girls, slightly lower than the WHO optima. Optimal smoothing parameters were determined sequentially through minimization of GAIC(2) and GAIC(3)[3]. Final models saw df(μ)=13, df(σ)=8, df(ν)=5 for boys and df(μ)=14, df(σ)=6, and df(ν)=5 for girls. Model identification is important, since 'underfitting' can smooth away important feature (like growth spurts), and 'overfitting' may interpret random fluctuations in sample quantiles as spurious trends.

The "CPEG Statistical Methods and Models manual" outlines the procedure step-by-step. This includes application of the WHO exclusion criteria, identification of optimal weight-for-age models, and model fitting. It also includes post-fit validation through specialized diagnostic procedures (Q statistics, Z statistics, worm plots, examination of model residuals). Final validation involves comparison with sample quantiles and existing WHO and CDC norms.

1.2 Presentation

These results can be presented as weight-for-age curves or tabulated as LMS parameters and smoothed centiles suitable for plotting. Both are available at the CPEG/GCEP website[2]. In figure 1.1, we see the results of merging the current WHO weight-for-age standards (2-10 years) with the complementary "CPEG curves" (10-19 years) for boys. The discontinuity at 10 years is restricted to the more extreme percentiles and more pronounced on the girls curves (figure 2.9). In large part, this reflects the difference in reference populations, since we had access to only the core NCHS data (n=22,917, ages 1-24 years), and the younger children (n∼8000, ages 18-71 months) still influence the curve-fit between 5-10 years of age. A similar discontinuity at 5 years reflects the transition from MGRS to NCHS data in the original WHO analysis.

[2] http://cpeg-gcep.net

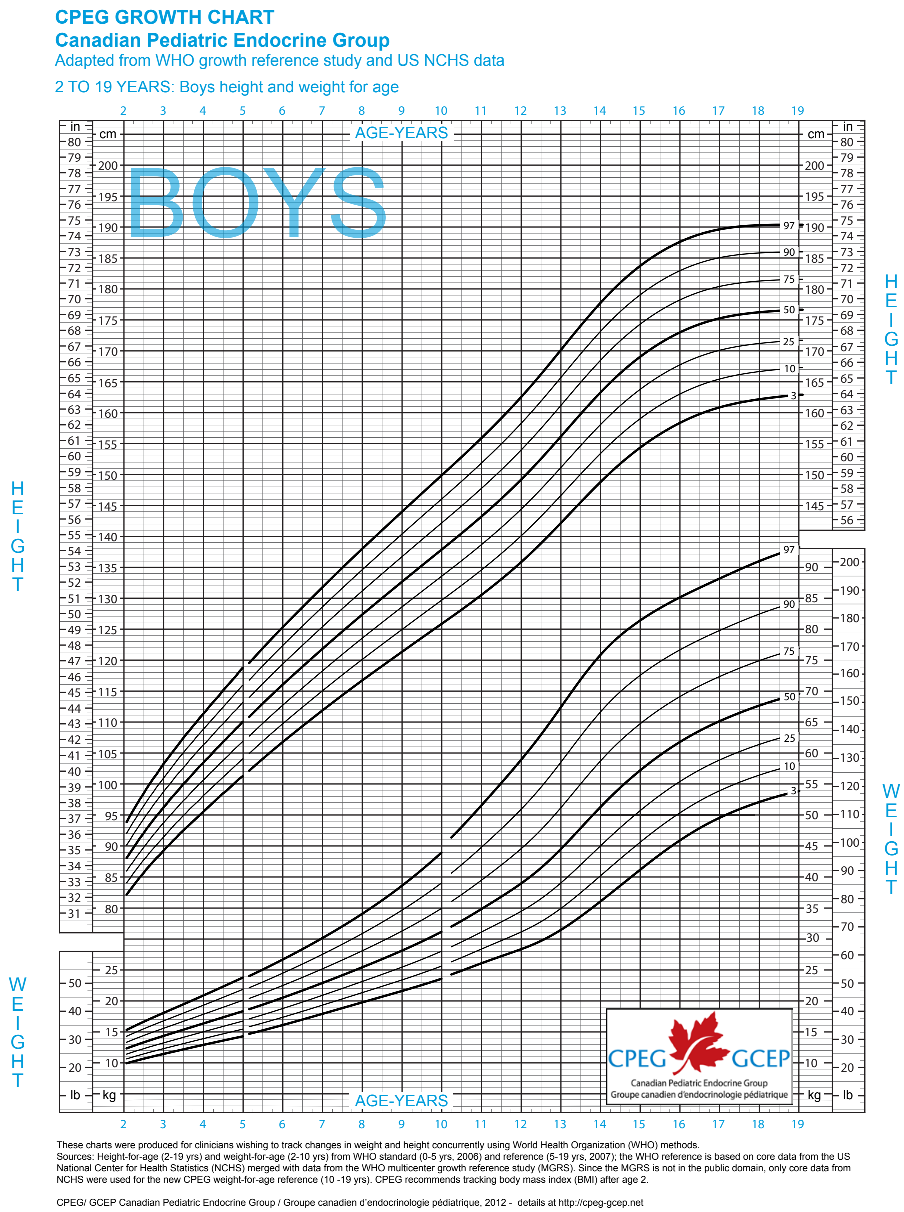


Figure 1.1: Representative CPEG/GCEP Growth Curves

1.3 Conclusion

The discontinuity between WHO and CPEG weight-for-age reference centiles at 10 years is small (figures 2.9, 2.13). Moreover, it is clear from figures 2.8 and 2.12 that our modeling strategy accurately captures sample (empiric) centiles. Nevertheless, the principle obstacle to acceptance of the new weight-for-age reference (10-19 years) will be the uncertain impact of the missing 8000 observations (ages 18-71 mo) from the WHO MGRS, which influence the model fit for older children [1]. Having have carefully applied both the WHO exclusion criteria and curve-fitting methodology in creating our own reference curves, the discontinuity at 10 years will largely reflect the differences in reference populations, in this case the missing 8000. It is less clear how far this discrepancy extends along the interval from 10-19 years.

To address this question, chapters 3 and 4 outline the results of refitting the height-for-age and BMI-for-age curves using just the NCHS data (n=22,917). Formal comparison with the WHO curves over the full age-range (5-19 years) highlights the influence of the missing children, since we have adhered closely to the WHO exclusion criteria and modeling principles in other respects. While not intended to supplant existing WHO norms, the working group felt this would be important as internal validation, to reassure readers as to the validity of the re-analyzed weight-for-age curves. Comparison with WHO centiles may be found in the appropriate chapters, specifically figures 3.4 (boys, height-for-age), 3.7 (girls, height-for-age), 4.4 (boys, BMI-for-age) and 4.7 (girls, BMI-for-age). Height-for-age curves are virtually indistinguishable from WHO norms, while the BMI-for-age curves show small discrepancies at the highest percentiles. We conclude that the missing cohort has a small impact on the fitted curves.

Chapter 2

Weight-for-age

2.1 Statistical analyses

All analyses were performed with the same SPlus/R software library used by the WHO, namely GAMLSS (Generalized Additive Models for Location, Shape, and Scale by Rigby and Stasinopoulos [6, 8])[1]. Unedited output from R-GAMLSS is displayed in `Courier` font to distinguish it from text. All values are means ± standard deviations (SD) unless indicated. As in the WHO and CDC technical manuals, the Methods and Models portion of this text fully documents all analyses carried out in the course of these studies.

2.2 Data sources

De Onis *et al*[1], provide an outline for the preparation of data prior to fitting smoothed percentiles for weight-for-age, height-for-age, and BMI-for-age in school aged children and adolescents (aged 5-19 years). Additional description is found in the WHO Methods and Development Technical report (2006)[3]. For each curve, the 'core data' refers to the 22,917 children (11507 girls, 11410 boys) pooled from 3 sources: NHES Cycle II (6-11 years, 1963-65), NHES Cycle III (12-17 years, 1966-70), and NHANES Cycle I (1-24 years, 1971-75). In addition "a smooth transition from the WHO child growth *standards* ($0-5$ years) to the *reference* curves beyond 5 years was provided by merging data from the cross-sectional portion of the WHO Multicenter Growth Reference Study (MGRS, $18-71$ months) with the NCHS samples before fitting the new growth curves"[1]. The MGRS involved longitudinal follow-up on 1737 children aged $0-24$ months and approximately 8000 cross-sectional observations on 6697 children aged $18-71$ months - i.e. 3450 boys, 3219 girls[2]. The growth curves for

[1] The WHO also recommends Tim Cole's LMS Pro software, which offers an easy-to-use and interactive user interface for the LMS method[3, 7, 9]

[2] Since the precise number of observations ranges from 7778-8667 depending on which anthropometric measure is being considered, we will frequently refer to this cohort as the

ages 5 to 19 years were thus constructed using data from 1 to 24 years to minimize edge effects (see figure 2.1). Only the core data are publicly available as we seek to extend weight-for-age norms to the older age range; fortunately, the impact of the younger children (18-71 mo) on the fitted curves at 10-19 years is small (see chapters 3, 4).

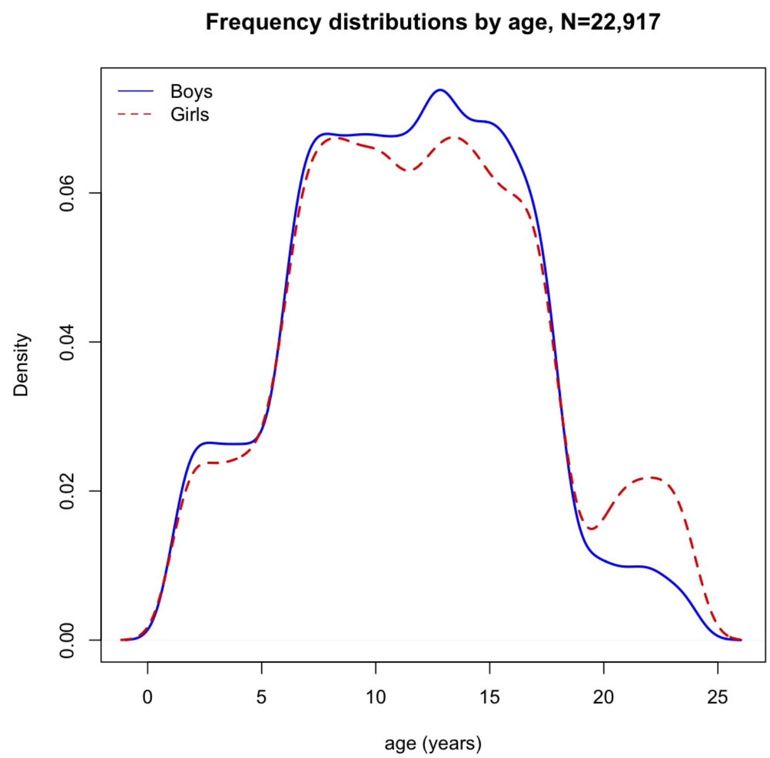


Figure 2.1: Age distribution of N=22,917 subjects (11507 girls, 11410 boys, ages 1-24 years) from NCHS surveys (1963-1975), the so-called 'core data' for estimation of all WHO growth curves for ages 5-19 years. Although constructed with data from 1 to 24 years to minimize edge effects, WHO analysts also merged the core data with cross-sectional data on 18-71 month olds (n ∼8000) to smooth the transition between WHO *standard* curves (0-5 years) and WHO *referenvce* curves (5-19 years).

'missing 8000'. They contributed to both the WHO *standard* (2006) and WHO *reference* (2007) curves, for convenience identified as 'WHO 2006' in figure legends.

2.3 Exclusion criteria

According to the published methodology[1, 3], data preparation invoked multiple exclusion criteria: First, 14 girls and 8 boys were dropped with outlying heights-for-age. For the weight based measures, an additional 596 subjects - 300 girls, 296 boys - were excluded with "unhealthy" weights-for-height. The latter exclusion was intended to define a "non-obese sample with expected height" [1] and based on weight-for-height indices less than the 0.135^{th} percentile (-3SD) or greater than the 97.7^{th} percentile (+2SD). A further 4 boys and 1 girl were dropped for weight-for-age observations deemed 'influential on the final fitted curves, although we do not know whether they arose from the NCHS or MGRS datasets. These exclusions merit attention, since differences between the CDC and WHO norms are "largely a reflection of differences in the populations on which the two sets of curves are based" [3].

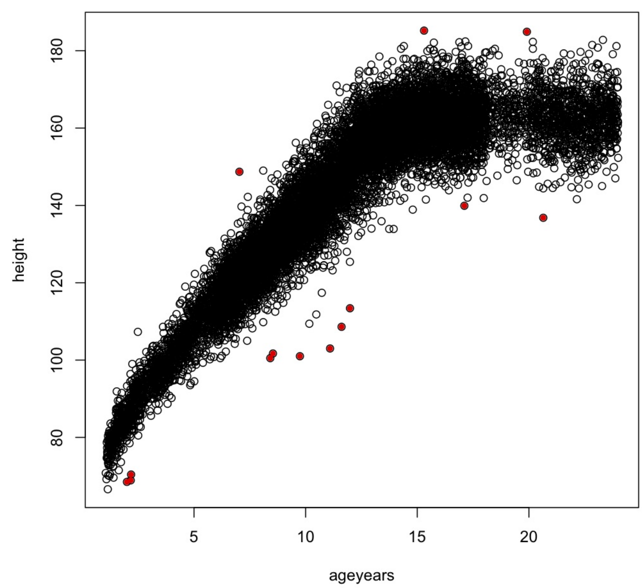
2.3.1 'Outlying' heights-for-age

Core data for the N=22,917 subjects (11507 girls, 11410 boys) from the NCHS surveys was kindly provided by Dr. de Onis. Plotting each gender cohort (figure 2.2) permitted manual exclusion of outlying heights-for-age, 14 girls and 8 males. For girls, this left 11396 after deleting observations: 266, 409, 613,647, 683, 685, 2987, 5936, 7724, 8042, 9477, 9623, 9701, 10870. For boys, this left 11402 after deleting observations: 729, 741, 759, 3745, 8102, 8154, 9882, 9945. The deleted observations are marked in red in figure 2.2.

2.3.2 'Unhealthy' weights-for-height

To exclude those with "unhealthy" weights-for-height outside of percentiles [0.135-97.7], a direct approach was adopted, fitting the full data range with a flexible 4-parameter Box Cox Power Exponential (BCPE) distribution, which can accommodate both skew and kurtosis[8]. For any value on the x-axis, the four parameters fitted by the model specify the median (μ), coefficient of variation (σ), skew (ν) and kurtosis (τ). Their time evolution is subsequently smoothed with cubic splines, with degrees of freedom (df) chosen to balance between parsimony (smoothness) and accurate representation of the sample centiles.

Readers who have used computer graphics programs are already familiar with cubic splines, which use cubic polynomial to draw smooth curves as the user manipulates control points with the mouse. The term"degrees of freedom" (df) refers to the number of control points along the curve. When there are as many control points as data points (df = n), the splines interpolate the data *exactly*. With fewer control points (df), the curves pass through fewer data points, becoming smoother in the process.



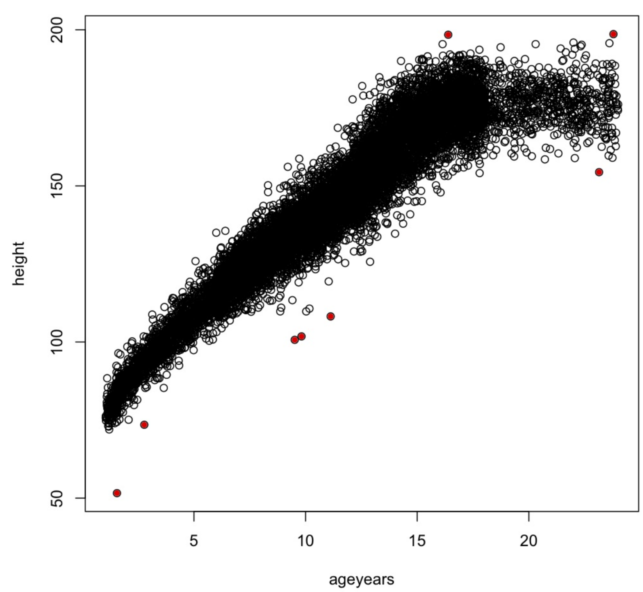


Figure 2.2: Exclusions for outlying heights-for-age, A) 14 girls, B) 8 boys

2.4 Boys: weight-for-height exclusions

For boys (N=11402), degrees of freedom were 13, 6, 3, 3 for μ, σ, ν, and τ, respectively, which represents a balance between smoothness and accurate representation of the sample centiles- see figure 2.3. Moreover, the model seems to reproduce the sample quantiles across the full range of heights:

```
% of cases below  0.135 centile is  0.1403
% of cases below  5 centile is  4.561
% of cases below  25 centile is  25.15
% of cases below  50 centile is  50.59
% of cases below  75 centile is  74.6
% of cases below  95 centile is  94.89
% of cases below  97.7 centile is  97.56
```

Performance is comparable on either side of the midline, a concern to WHO investigators.

	72 to 145.3	145.3 to 195.9
0.135	0.1406	0.1401
5	4.5686	4.5526
25	24.8814	25.4246
50	50.3426	50.8317
75	74.6793	74.5229
95	95.1327	94.6419
97.7	97.5751	97.5486

This model was used to identify 296 “unhealthy” height-for-weight values, the bulk (280) greater than percentile 97.7. Total exclusions thusfar (304) represent 2.7% of n=11410 boys. This compares to 321 exclusions reported by the WHO with the same criteria[1].

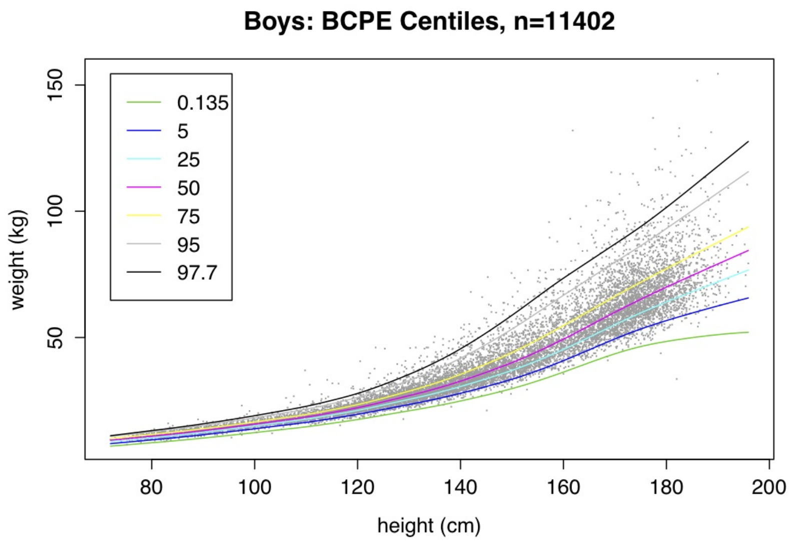


Figure 2.3: BCPE distribution fit to weight-for-height. For boys, df(μ)=13,df(σ)=6, df(ν)=3 and df(τ)=3.

2.5 Girls: weight-for-height exclusions

Degrees of freedom for girls were 12, 4, 3, 3 for μ, σ, ν, and τ - see figure 2.4. For a screening procedure, the model captures the sample centiles reasonably well:

```
% of cases below  0.135 centile is  0.174
% of cases below  5 centile is  4.742
% of cases below  25 centile is  25.01
% of cases below  50 centile is  50.7
% of cases below  75 centile is  75.04
% of cases below  95 centile is  94.67
% of cases below  97.7 centile is  97.56
```

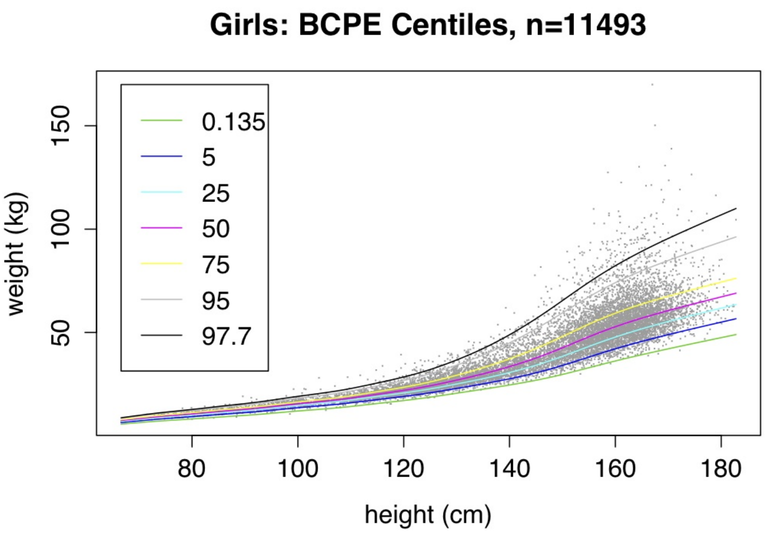


Figure 2.4: BCPE distribution fit to weight-for-height. For girls, df(μ)=12, df(σ)=4, df(ν)=3 and df(τ=3).

Despite the differing variability in the two half-plots, agreement with sample centiles is also comparable on each side of the midline:

	66.6 to 148.7	148.7 to 182.8
0.135	0.2268	0.1215
5	4.5018	4.9809
25	25.3533	24.6616
50	51.3174	50.0868
75	75.6936	74.3839
95	94.6606	94.6720
97.7	97.7840	97.3447

This model lead to exclusion of 300 observations , with most (280) greater than percentile 97.7. Total exclusions for both outlying heights-for-age and "unhealthy" weights-for-height (314) represent 2.7% of N=11507 girls. This compares to 356 exclusions by the WHO using the same criteria[1].

2.6 Power transformation of the time axis

In the initial publication of new growth standards [2, 3], the WHO adopted the 4-parameter Box-Cox Power Exponential or BCPE distribution for *all* growth measures. This model includes parameters describing the behavior of the median (μ), coefficient of variation (σ), skew (ν, the Box-Cox exponent for transformation to normality), and kurtosis (τ)[6, 8]. However, in all cases τ was fixed at a value of 2, denoting an absence of significant kurtosis. As a result, the BCPE model may be simplified to the 3-parameter Box-Cox Cole-Green (BCCG or LMS) distribution, a simpler model which omits the kurtosis measure τ (formal correspondence between models is given by $\mu \equiv$ M, $\sigma \equiv$ S and $\nu \equiv$ L, τ=2)[7, 9]. A variant of the LMS model was also applied by the CDC in their 2000 revision of growth norms, making it the common standard for both analyses[3, 4]. The parameter values at each point on the x-axis were subsequently smoothed with cubic splines, whose smoothing parameters (degrees of freedom df) must be user-specified.

As noted, many anthropometric measures require a preliminary power transform (exponent λ) of the x-axis to "spread out" time and better capture periods of rapid change. The optimal power transform in the WHO analysis was determined by sensitivity analysis using an arbitrary model to minimize global deviance, which lead to λ=1.4 in boys and λ=1.3 in girls. Before the LMS parameters can be estimated, the optimal smoothing model must also be identified. Appropriate model identification is a critical step: Too simple a model leads to underfitting, which can smooth away important feature (like growth spurts). Conversely, overfitting random fluctuations leads to spurious trends. Consequently, the WHO technical report outlines a stepwise approach to identification of model hyperparameters df(μ), df(σ), df(ν) and df(τ) based on sequential optimization of model fit.

Given that the NCHS core data represents more than 2/3 of the data used to construct the WHO norms and includes all the data in the target age range (10-19 years), the impact of the missing 18-71 month olds on the fitted curves at 10-19 years is presumably small, an assumption to which we shall return shortly. Hence, our general strategy consisted of *initiating* our model identification procedure with the 'optimal' WHO model i.e. for boys, the optimal hyperparameters were df(μ)=10, df(σ)= 8, and df(ν) = 5. For girls, hyperparameters were df(μ)= 10, df(σ)= 3, and df(ν) = 3. Refinement of model then proceeded in a stepwise fashion.

These basic models were used to first estimate the optimal value for transformation of the time axis by minimizing the global deviance, systematically varying λ from 1.0 to 1.5 in increments of 0.025 (for boys, see figure 2.5).

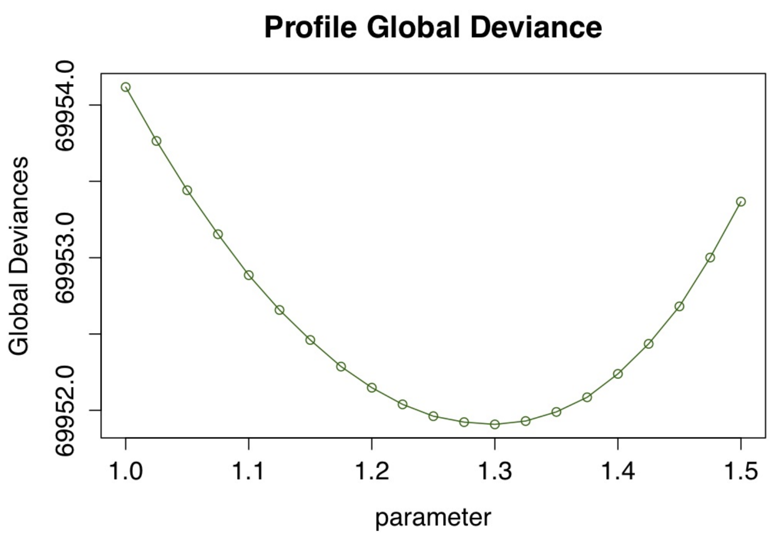


Figure 2.5: Boys: global deviance as exponent λ of age transformation varied from 1.0 to 1.5, with nadir at λ=1.3

```
Best estimate of the fixed parameter is 1.3
with a Global Deviance equal to 69952  }
```

Similarly, for girls, the exponent λ of the age transformation was systematically varied from 1.0 to 1.5 (see figure 2.6).

```
Best estimate of the fixed parameter is  1.225
with a Global Deviance equal to  71428
A  95 % Confidence interval is: ( 1.001 , 1.425 ) }
```

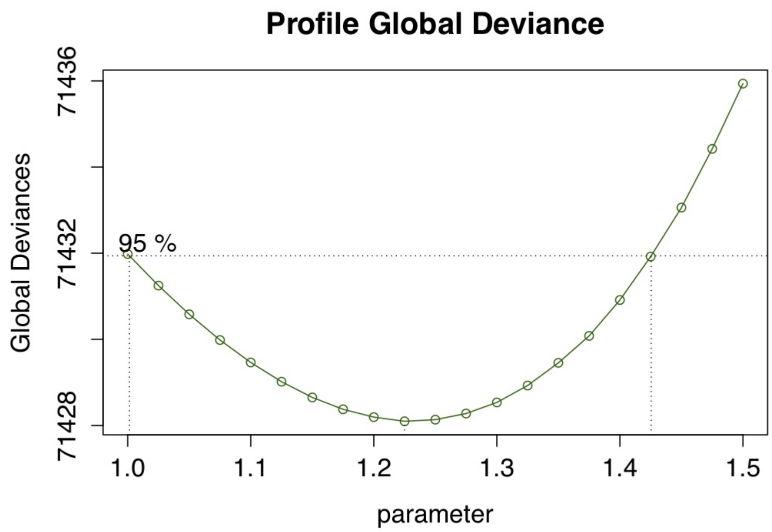


Figure 2.6: Figure 8: Girls: global deviance as exponent λ of age transformation varied from 1.0 to 1.5, with nadir at λ=1.225

2.7 Optimal smoothing models

Model fit - i.e. agreement with sample quantiles - is measured using either global deviance or the Generalized Akaiki Information Criterion (GAIC)[6, 7, 8, 9]. The latter allows a balance between local fit and smoothness through an adjustable penalty term for 'roughness': A penalty = 2 reduces to the familiar Akaiki Information Criterion (AIC) and favors local fit, while a penalty = 3 favors smoother curves. These two measures were used in parallel by the WHO investigators to determine df(μ) and df(σ), seeking a consensus judgment when possible. Otherwise, the minimum AIC value was used to select the smoothing model for df(μ), and GAIC(3) was minimized to determine the optimal value for df(σ). Only GAIC(3) was used to identify df(ν) (skew) and df(τ) (kurtosis). In rare instances, diagnostic worm plots, Q-tests, and comparisons between predicted and sample centiles also guided model selection. In what follows, we retain this sequential approach to fixing df(μ), df(σ), df(ν) and - when needed - df(τ). Model suitability is then confirmed by appropriate diagnostic testing with worm plots, Q-statistics and comparison with sample centiles [3, 10, 11]. This sequential approach finds theoretical justification in the relative importance of the model parameters to overall fit, first forcing good agreement with the sample median through the GAIC(2) criterion before relaxing the penalty to favor smoothness via GAIC(3). It is also helpful that the degrees of freedom can generally be fixed independently of each other [7, 10]

Having first identified $\lambda = 1.3$ for transforming the time axis (boys), we have an advantage on the WHO analysts, in that we know know their optimal smoothing model and expect ours to be close (in a sense to be formalized shortly), since the two datasets overlap significantly and the exclusion criteria are the same. Consequently, the WHO model can be used to initiate the optimization engine in the GAMLSS function `find.hyper()`, which systematically searches parameter space to find the optimal degrees of freedom that minimize the GAIC. Once initiated, this iterative search relies on the Broyden-Fletcher-Goldfarb-Shannon steepest-descent algorithm (BFGS) to find the minimum point on the 'GAIC surface' (for obvious reasons, these optimization procedures are also known as 'hill-climbing' algorithms; users should be forewarned that the algorithm runs S-L-O-W-L-Y)[6, 8]. As noted by Cole and Green[7], this task is made simpler by recognizing that the 3 degrees of freedom can be optimized independently of each other.

2.7.1 Optimal smoothing model, boys

From any starting point, we invoke the GAMLSS function `find.hyper()` to determine the optimal hyperparameter values (degrees of freedom). To avoid convergence problems, the initial model should be close to the desired optimum. The first output below refers to the hyperparameter values (*par*) at the minimal GAIC(2), with GAIC(2) *value*=70008. The second output is the result of minimizing GAIC(3). Here, we initiate our search at the WHO optima and systematically explore parameter space for minimizing values of df(μ), df(σ), and df(ν).

```
# penalty = 2.0
$par [1] 13.100  8.000  6.268
$value [1] 70008

# penalty=3.0
$par [1] 10.55  8.00  5.0
$value [1] 70039
```

The first search yields 13 , 8, and 6 at the AIC minimum. The second yields 11, 8 and 5 with GAIC(3). As expected, a more parsimonious model is selected by GAIC(3), which favors smoothness over local fit. Proceeding sequentially, a third search was undertaken to minimize GAIC(3) with λ=1.3, df(μ)=13 and df(σ)=8.

```
# penalty=3.0
$par [1] 5
$value [1] 71125
```

This identified the smoothing model 13, 8, and 5 as the minimizing values for df(μ), df(σ) and df(ν), given λ=1.3. As seen below, there is no evidence of over-fitting (i.e. 'roughness') in either the parameter curves or smoothed centiles.

2.7.2 Optimal smoothing model, girls

Our initial smoothing model had degrees of freedom 10, 3 and 3. From this starting point, we invoke the GAMLSS function `find.hyper()` to search for those degrees of freedom minimizing GAIC(2) and/or GAIC(3).

```
# penalty=2.0
$par [1] 13.643  8.434  5.779
$value [1] 71436

# penalty=3.0
$par [1] 11.489  6.442  4.914
$value [1] 71467
```

Minimum AIC identified model df(μ)=14, df(σ)=8, df(ν)=5 as optimal. Similarly, the GAIC(3) criterion identifies df(μ)=11, df(σ)=6, df(ν)=5. A third search was therefore run to minimize GAIC(3) with λ=1.225,and df(μ)=14.

```
# penalty=3.0
$par [1] 6.476 4.927
$value [1] 71468
```

This third application of the `find.hyper()` function identified df(μ)=14, df(σ)=6, df(ν)=5 as the optimal model, with the GAIC(3) nadir at `71468`, given λ=1.225. In the final step, this model is applied to the data and returns the LMS parameters by age (figure 2.7)[3].

[3]LMS parameters by age are also available in spreadsheet form at the CPEG website

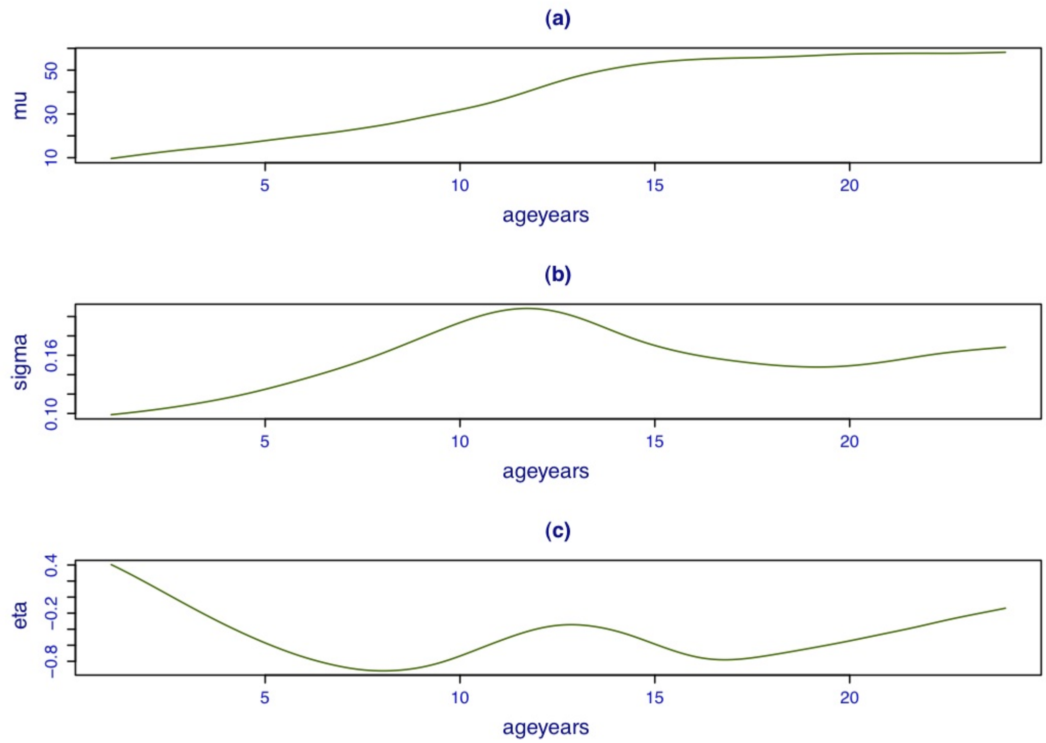


Figure 2.7: Girls: Time evolution of LMS model parameters for girls. A) Median (μ or M), B) Coefficient of variation σ or S, C) Skew parameter ν or L (Box-Cox exponent)

2.7.3 Applications

For completeness sake, we cite the relevant conversion formulae here. Given a measurement of interest y, the corresponding z-score is given by the following equation, where the Box-Cox power transformation via ν first normalizes the skew distribution[7, 8]:

$$z_\alpha = \frac{(y/\mu)^\nu - 1}{\sigma\ \nu}$$

Given a standardized quantile z_α (in SD units), the corresponding 100 α^{th} percentile value is then given by:

$$\begin{aligned} y_{100\alpha} &= \mu \cdot (1 + \sigma \cdot \nu \cdot z_\alpha)^{\frac{1}{\nu}}, \quad \nu \neq 0 \\ &= \mu \cdot exp(\sigma \cdot z_\alpha), \quad \nu = 0 \end{aligned}$$

These formula are strictly applicable on -3 $\leq z_\alpha <$ +3, since estimation is difficult in the extreme tails of a skew distribution. Althought the WHO offered an *ad hoc* adjustment for z-scores outside this range, ±3SD covers percentiles 0.135 to 99.9, which is sufficient for most clinical applications. For reference, the CDC identified the principle curves of interest as those for z-scores of - 1.881, -1.645, -1.282, -0.674, 0, 0.674, 1.036, 1.282, 1.645, and 1.881; corresponding to the 3rd, 5th, 10th, 25th, 50th, 75th, 85th, 90th, 95th, and 97th percentiles, respectively[4].

2.8 Fitted model, girls

At each age, the LMS parameters are used to generate smoothed centiles, which can be compared with sample centiles (girls, figure 2.8).

```
% of cases below 3 centile is 2.868
% of cases below 25 centile is 25.1
% of cases below 50 centile is 50.09
% of cases below 75 centile is 75.5
% of cases below 97 centile is 96.73
```

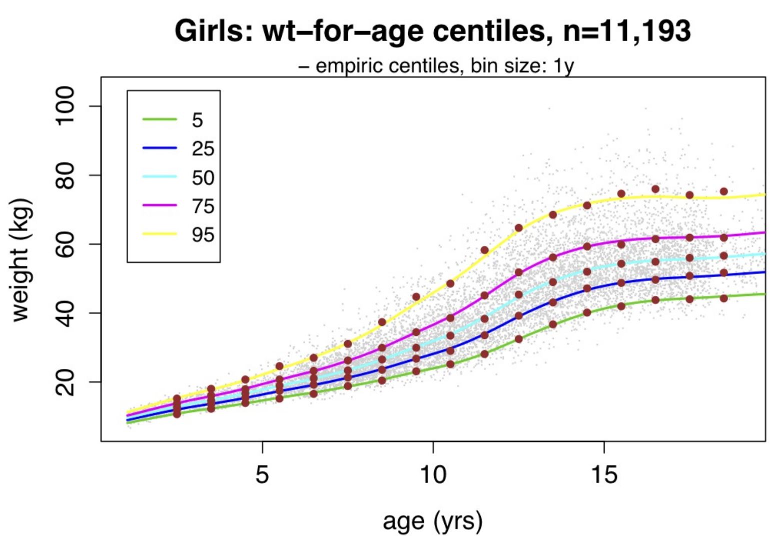


Figure 2.8: Girls 2-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year

Comparisons with WHO norms is also informative, since we seek a practical extension of their weight-for-age curves along the age axis. From 5-10 years, WHO weight-for-age norms are available for comparison, and agreement is good except at the upper centiles (figure 2.9). Although the core data and exclusion criteria are the same, the missing 8000 observations between ages 18-71 months may account for a small discrepancy even in more remote segments of the fitted curve.

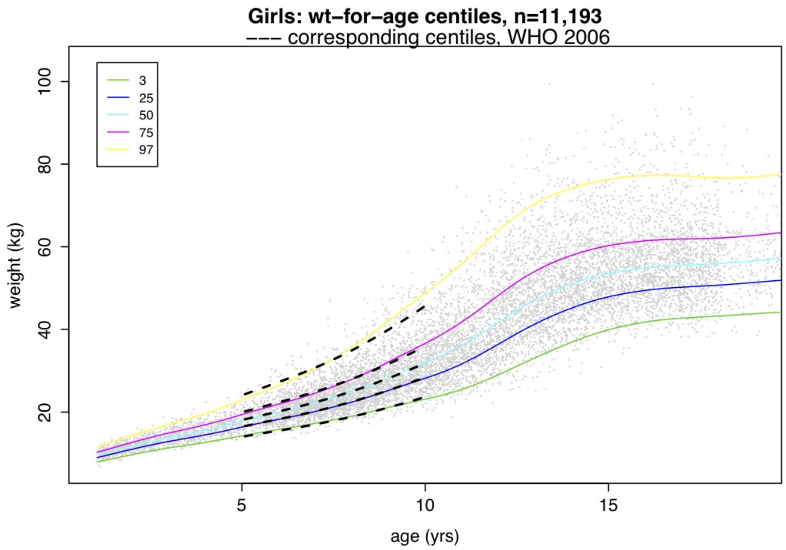


Figure 2.9: Girls smoothed centiles (2-19 years) vs WHO centiles (5-10 years). The core NCHS data was used for both analyses, but the WHO added 8000 additional observations with ages 18-71 months to smooth the transition between WHO and NCHS data.

In numeric terms, the divergence from WHO curves can be quantified as the mean absolute deviation (MAD) ± SD (averaged over monthly measurements from ages 5-10). As seen in the graph, the discrepancy is maximal at the higher percentiles.

```
97th percentile: 3.304 +- 2.018 %
50th percentile: 0.748 +- 0.397 %
3rd percentile:  1.158 +- 0.604 %
```

These monthly comparisons can also be expressed as the mean absolute deviation in kg:

```
97th percentile:  1.184 +- 0.9098 kg
50th percentile:  0.167 +- 0.0686 kg
3rd percentile:   0.204 +- 0.1101 kg
```

In contrast, we expect less agreement with the CDC norms (figure 2.10), since the reference populations are not the same by intention (the WHO excluded a significant number of children with "unhealthy" weights-for-height). To facilitate comparison with the previous result, the mean absolute deviation vs CDC norms was averaged over monthly measurements on the same interval, ages 5-10 years:

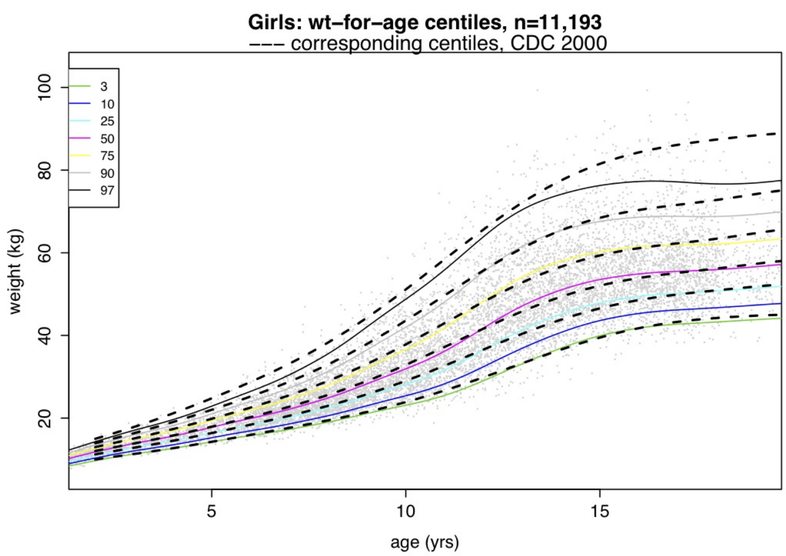


Figure 2.10: Girls smoothed centiles (2-19 years) vs CDC centiles (2-19 years).

```
97th percentile: 6.815 +- 1.376 %
50th percentile  2.219 +- 0.6872 %
3rd percentile:  1.747 +- 0.7863 %
```

The mean absolute deviation can also be expressed in kg:

```
97th percentile: 2.404 +- 0.2623 kg
50th percentile: 0.570 +- 0.2307 kg
3rd percentile:  0.346 +- 0.1812 kg
```

2.9 Fitted model, boys

The optimal model for boys was 13, 8, and 5 for df(μ), df(σ) and df(ν). Fitting this model returns LMS parameters by age, which plotted in figure 2.11[4]. These age-specific parameters yield smoothed centiles, which can be compared to sample quantiles in figure 2.12.

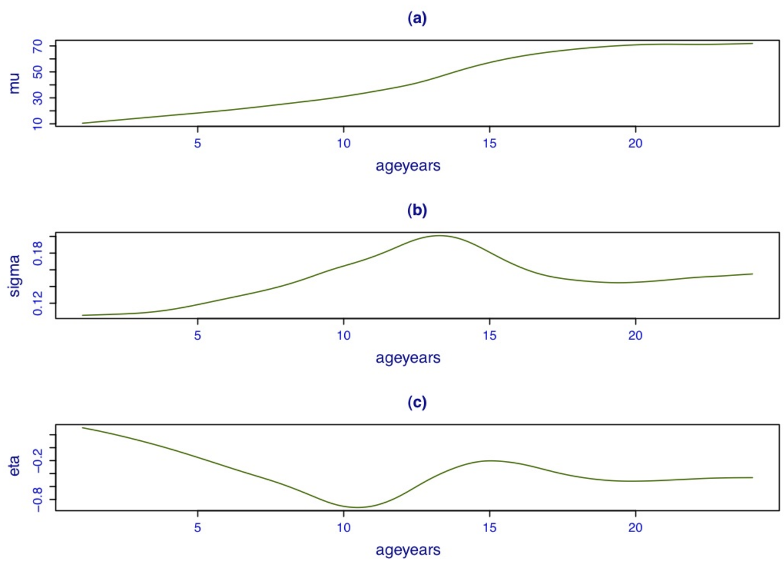


Figure 2.11: Time evolution of LMS model parameters for boys. A) Median (μ or M), B) Coefficient of variation σ or S, C) Skew parameter ν or L (Box-Cox exponent)

```
% of cases below  3 centile is  2.908
% of cases below  25 centile is  24.47
% of cases below  50 centile is  50.31
% of cases below  75 centile is  75.55
% of cases below  97 centile is  96.78
```

In this case, the predicted centile curves appear identical to the WHO norms for ages 5-10 years. While small, the mean absolute deviation averaged over

[4]LMS parameters by age are also available in spreadsheet form at the CPEG website

monthly measurements from ages 5-10 years is proportionally larger for the "outer" percentiles, with mean absolute discrepancy (vs WHO curves over the interval 5-10 years):

```
97th percentile: 1.182 +- 0.6731 %
50th percentile: 0.104 +- 0.0736 %
3rd percentile:  0.690 +- 0.5211 %
```

These monthly comparison can also be expressed as the mean absolute deviation in kg:

```
97th percentile:  0.3707 +- 0.2061 kg
50th percentile:  0.0249 +- 0.0171 kg
3rd percentile:   0.1189 +- 0.0732 kg
```

Although we do not expect agreement with the CDC norms, we note that the worst discrepancies are again localized to the upper percentiles. To permit comparison with the previous result, the mean absolute deviation vs CDC was calculated from monthly measurements on the interval 5-10 years:

```
97th percentile: 10.58 +- 1.596
50th percentile: 1.871 +- 1.530
3rd percentile:  1.701 +- 0.846
```

The monthly comparisons (MAD) can also be expressed in kg:

```
97th percentile: 4.000 +-  1.391 kg
50th percentile: 0.509 +-  0.504 kg
3rd percentile:  0.298 +-  0.126 kg
```

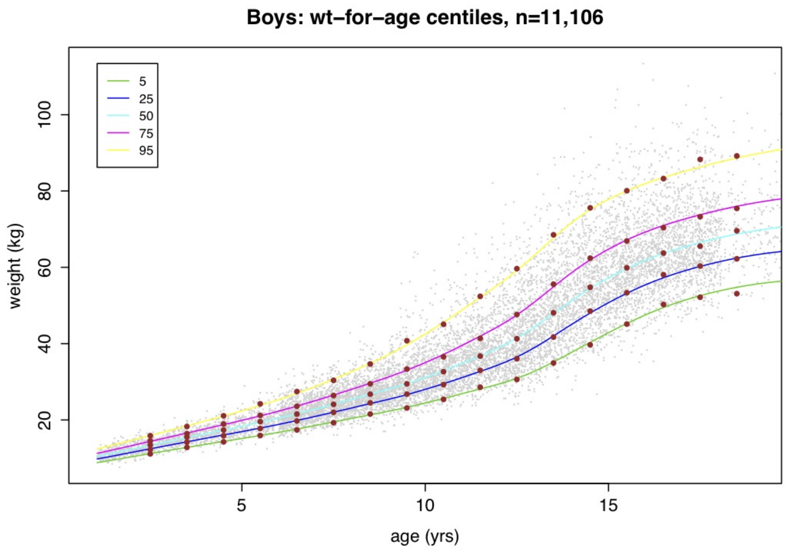


Figure 2.12: Boys 2-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year

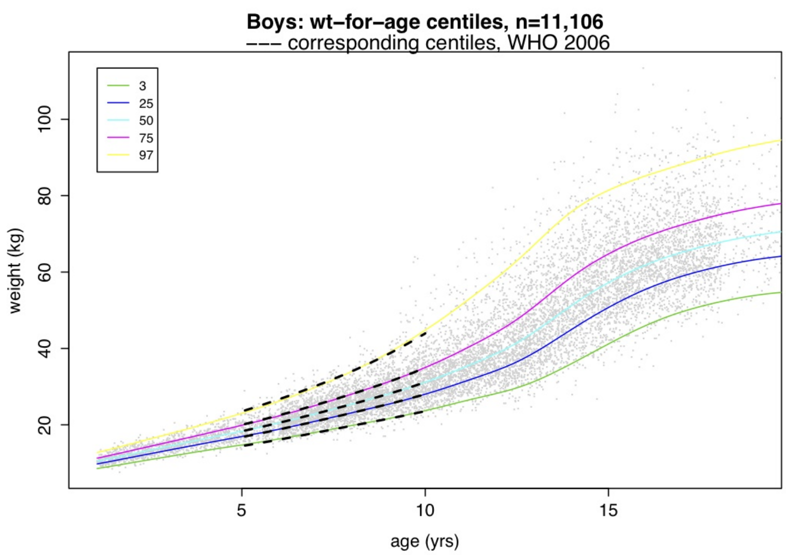


Figure 2.13: Boys smoothed centiles (2-19 years) vs WHO centiles (5-10 years).The core NCHS data was used for both analyses, but the WHO added 8000 additional observations with ages 18-71 months to smooth the transition between WHO and NCHS data.

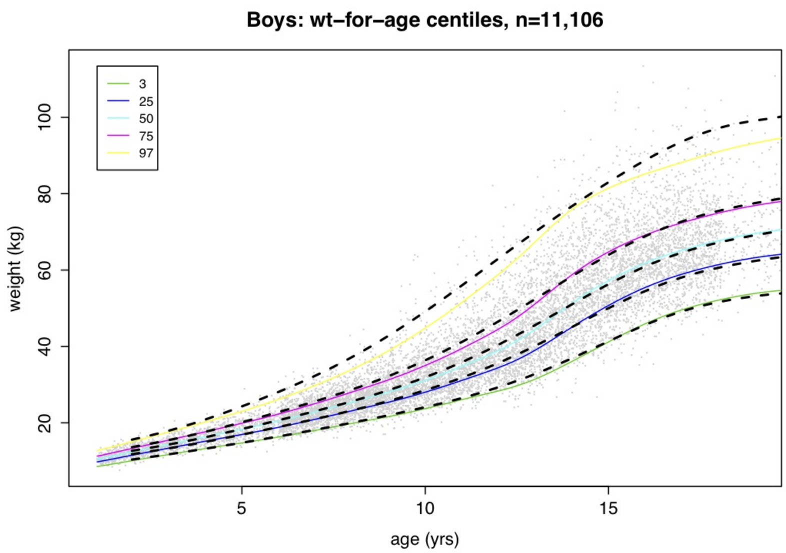


Figure 2.14: Boys smoothed centiles (2-19 years) vs CDC centiles (2-19 years).

2.10 Model Diagnostics, girls

Since diagnostic procedures for model adequacy will be essential to what follows, we will spend a few moments discussing their interpretation [3, 10, 11]. Model adequacy is first assessed by examining the normality of the residuals. As shown here, mean = 0, SD =1, and there is no evidence of skew or kurtosis (for a normal distribution, the coefficient of skew = 0 and coefficient of kurtosis = 3, not to be confused with the parameters ν and τ in the BCPE model[8]).

```
      Summary of the Quantile Residuals
                           mean   =  -5.644e-05
                       variance   =  1
               coef. of skewness  =  0.002827
               coef. of kurtosis  =  2.969
Filliben correlation coefficient  =  0.9998
```

Residual plots, including frequency histograms and QQ-plots, confirm the normality of the residuals (figure 2.15). They are also useful for identifying potentially influential outliers. Recall that with 10,000 observations, we may reasonably expect 1-2 observations at ±4 SD (99.99^{th} percentile). The absence of more egregious outliers is therefore reassuring. When outliers were identified in residual plots, they were deleted from the dataset and the model was refitted. This serves to assess their influence on the final results. Fortunately, there were no instances where outliers were deemed to have undue influence on the smoothed centiles (*vide infra*)

For diagnosing specific "model deviations" or failure to adequately model specific parameters, the WHO relied primarily on worm plots. Q and Z statistics also complemented these graphical diagnostics, particularly in ambiguous cases[3, 10, 11].

Worm plots are detrended normal Q-Q plots of residuals, which highlight departures from normality as points *outside* the 95% confidence intervals (dotted lines). Different patterns also serve to identify distinct types of model inadequacy, which are detailed with examples in the WHO technical report (page 10, [3] or in [10]). Typically, the detrended residuals are smoothed, and the path of the smoothed curve (the solid red line in figure 2.16) can identify specific model violations. For example, a best-fit line will have an intercept of zero if the median μ is adequately modeled , and its slope will be zero if the variance σ is modeled adequately. Similarly, a parabolic or U-shaped curve speaks to residual skew or an inadequate model for ν i.e. an upward U-shape suggests a leftward skew, while an inverted U suggests a rightward skew. And an S-shape speaks to residual kurtosis or errors in the model for τ; if the S bends up on the left, the tails are too heavy. If it bends down on the left, the tails are too light [3, 10].

Here, all residuals fall within the 95% confidence interval (figure 2.16).

When localizing deviations from normality, worm plots by age interval add diagnostic precision[10]. Here all residuals fall within the confidence limits for

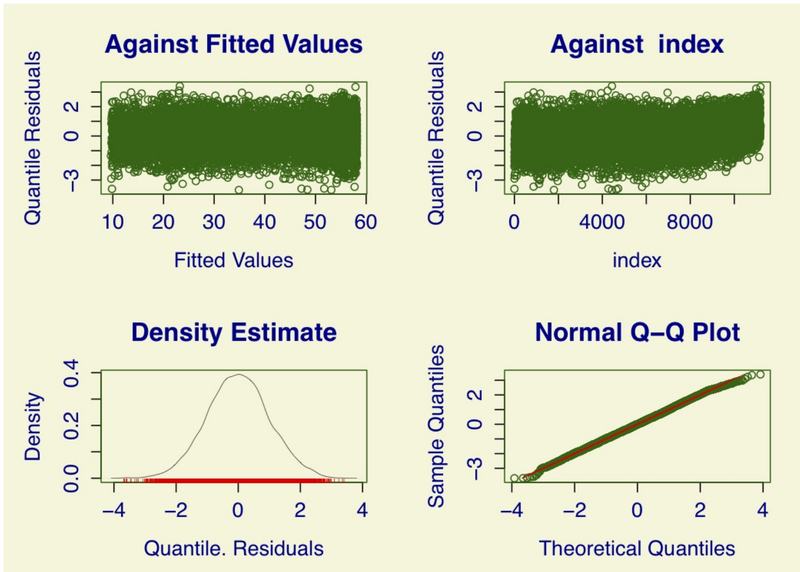


Figure 2.15: Residual plots for assessment of model assumptions. With an adequate fit, residuals should be normally distributed with mean=0, SD=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals (deviations from straight line = deviations from normality)

each individual age interval, so there appear to be no localized regions of model inadequacy (figure 2.17).

In theory, both Z_i and Q_i (i=1,2,3,4) evaluate the adequacy of the fit (respectively) for mean parameter (μ), variance parameter (σ), skew parameter (ν = 1 in the absence of skew), and kurtosis parameter τ (= 2 in the absence of kurtosis)[11]. Under the null hypothesis with an adequate fit, Z_i is gaussian $\sim$ Normal[0,1] and in general, $Z_i>2$ indicates model inadequacy for a specific interval on the age-axis. Recall that with 20 age intervals, one $Z_i>2$ is expected (p=0.05), and caution must therefore be shown in the interpretation of mild deviations from normality (e.g. a small number of Z_i between 2-3).

After inspecting the individual Z_i, the sum of squared Z_i over all age intervals yields the overall Q_i statistic, which is assessed by comparison with an appropriate chi-squared distribution. Significant Q_1, Q_2, Q_3, Q_4 statistics indicate possible inadequacies in the model for parameters μ, σ, ν and τ respectively, which may be overcome by increasing their degrees of freedom. However, the Q statistics are sensitive to mild deviations from normality, and care must be taken in their interpretation. In this case, model fit appears adequate: in the specific age groups, all $Z_i<2$, and p=NS for overall Q statistics.

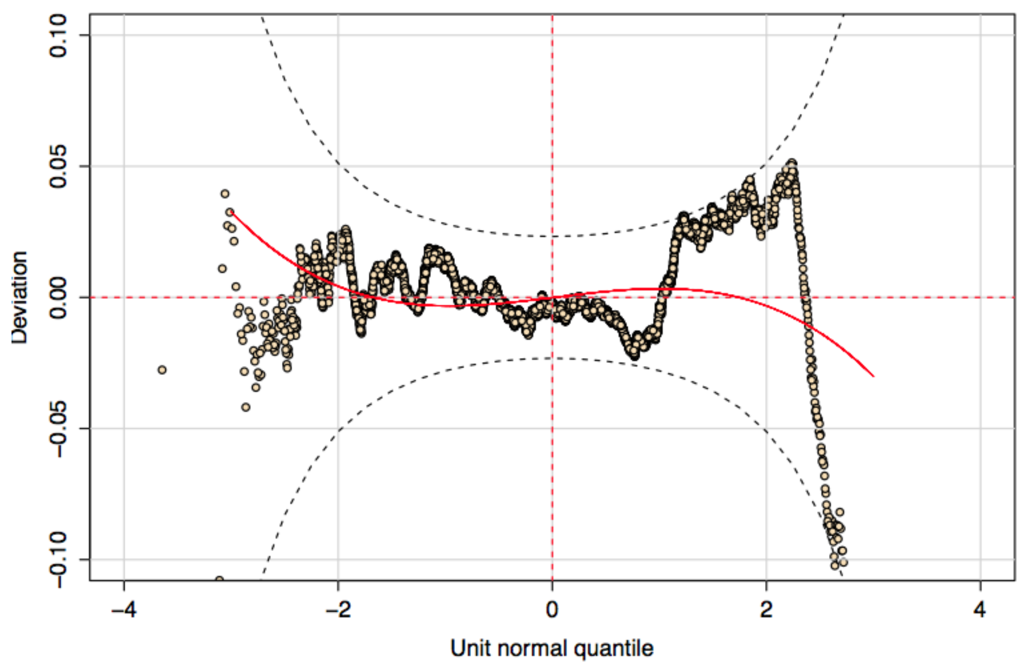


Figure 2.16: Worm plot of model residuals. With an adequate model, de-trended residuals should lie between the 2 dashed lines (95% confidence interval). The path of the smoothed curve (solid red) can identify specific model violations. Briefly, the best fit line will have an intercept of zero if the median is adequately modeled (μ), and its slope will be zero if the variance (σ) is modeled adequately. Similarly, a U-shaped curve suggests residual skew, and an S-shaped curve speaks to residual kurtosis.

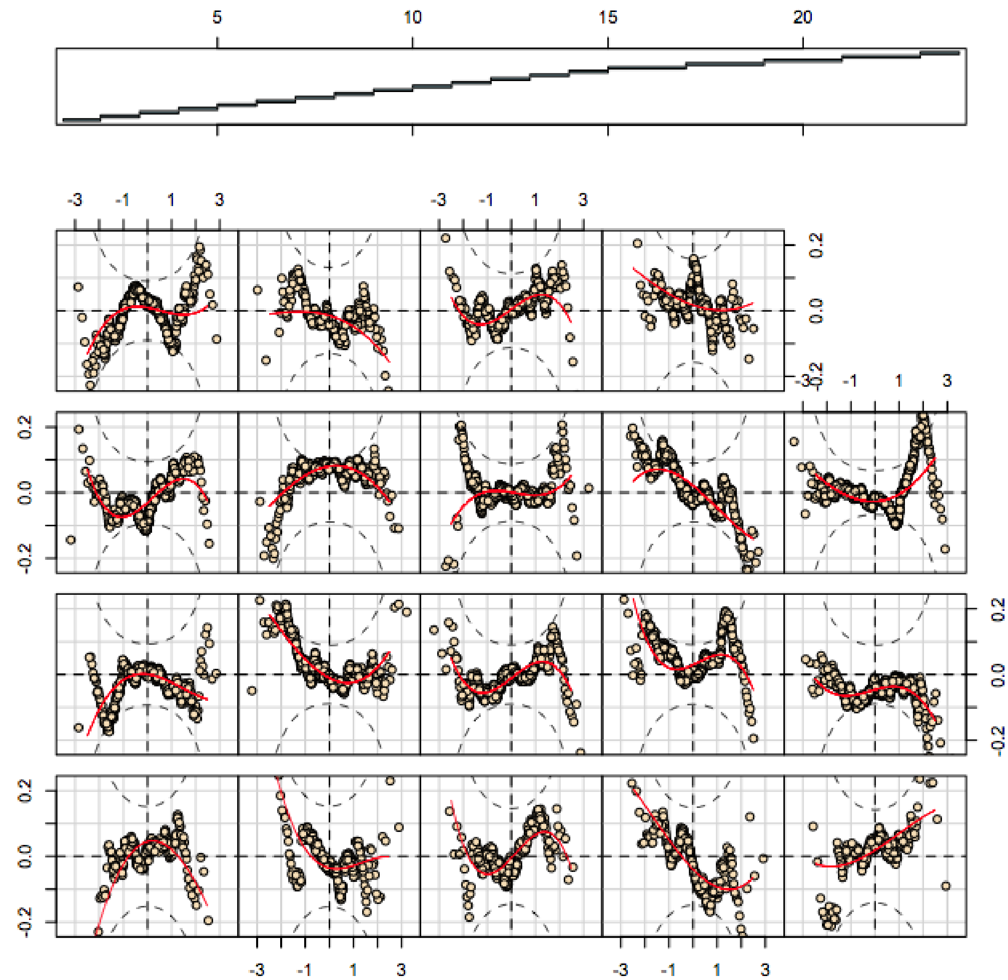


Figure 2.17: The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, with both worm plots and Z statistics applied separately on each interval. This facilitates localization of so-called “model violations”.

```
> Q.stats(nchs1, xvar=ageyears,xcut.points=xc)
                        Z1       Z2       Z3       Z4          N
1.04859 to 2      -0.02335  0.69311 -1.79139  0.75215        263
2 to 3            -0.04496 -0.89280  1.18631 -0.19270        259
3 to 4             0.09686  0.56188  0.43208 -1.45522        286
4 to 5            -0.35698 -1.52096  0.77165  0.34612        274
5 to 6             0.43673  0.99341  0.22304  0.04850        301
6 to 7            -0.54063  0.25789 -1.39256  0.68693        699
7 to 8             0.31416 -1.17609  1.52057  0.40435        761
8 to 9            -0.26372  0.80965  0.16497 -1.52545        746
9 to 10            1.06910 -0.28218  0.77049 -1.74186        732
10 to 11          -1.41599 -0.05611 -0.33443 -0.86995        759
11 to 12          -0.57689  0.97541  0.58732 -1.59685        682
12 to 13           1.67933  0.28198 -1.23357  0.03571        748
13 to 14          -0.07135  0.34583 -0.36798  1.15511        780
14 to 15           0.24804 -1.83000 -0.98251  1.09648        748
15 to 17          -0.35158  0.40102  1.70978  0.38813        1343
17 to 19          -0.05802  0.32992 -0.73213  0.88424        744
19 to 21          -0.48912 -0.63076 -0.52269  0.19324        350
21 to 23           0.07530  0.66590  0.02386 -0.90412        475
23 to 23.9972      0.41068 -0.49222  0.41175  0.24490        243
TOTAL Q stats      7.69875 12.93406 17.30240 16.53096      11193
df for Q stats     3.00253 14.50068 11.99923 19.00000          0
p-val for Q stats  0.05277  0.57016  0.13854  0.62161          0
```

2.11 Model Diagnostics, boys

For the boys, adequacy is also assessed by examining the normality of the residuals. As shown here, mean = 0, SD =1, and there is no evidence of skew or kurtosis.

```
      Summary of the Quantile Residuals
                           mean   =  0.0001588
                       variance   =  1
               coef. of skewness  =  -0.002909
               coef. of kurtosis  =  3.163
Filliben correlation coefficient  =  0.9997
```

This impression is confirmed by examination of model residuals - including frequency histograms and Q-Q plots (figure 2.18). Some outliers are evident in this plot, with absolute values ≈ 3.5-4. While some outliers are inevitable with 11,000 observations, more extreme values raise concerns about potential influence on the fitted model. When this is a concern, the more extreme obsevations are identified by index number, in this case observations: 12, 808, 1228, 1285, 1338, 2811, 4949, 6517. Potential influence can be assessed by deletion of these observations, which had no effect on the fitted centiles (figure 2.19).

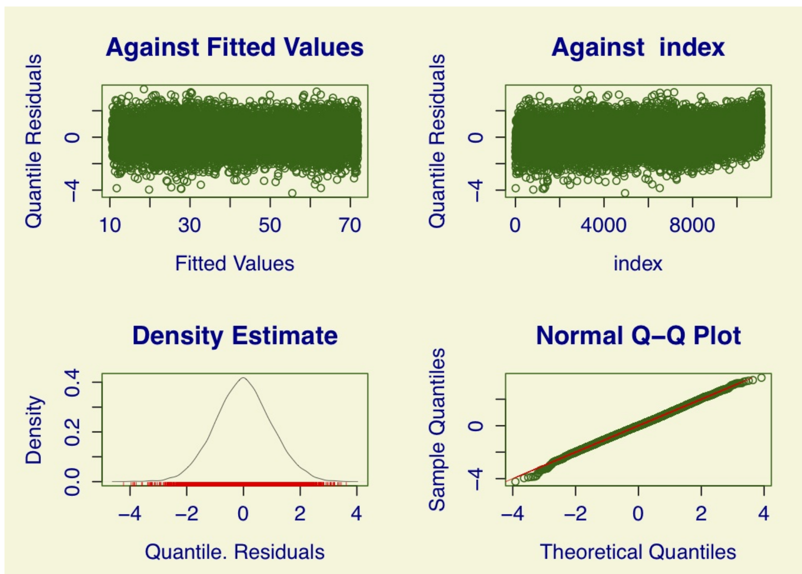


Figure 2.18: Residual plots for assessment of model assumptions (boys). With an adequate fit, residuals should be normally distributed with mean=0, SD=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals (deviations from straight line = deviations from normality)

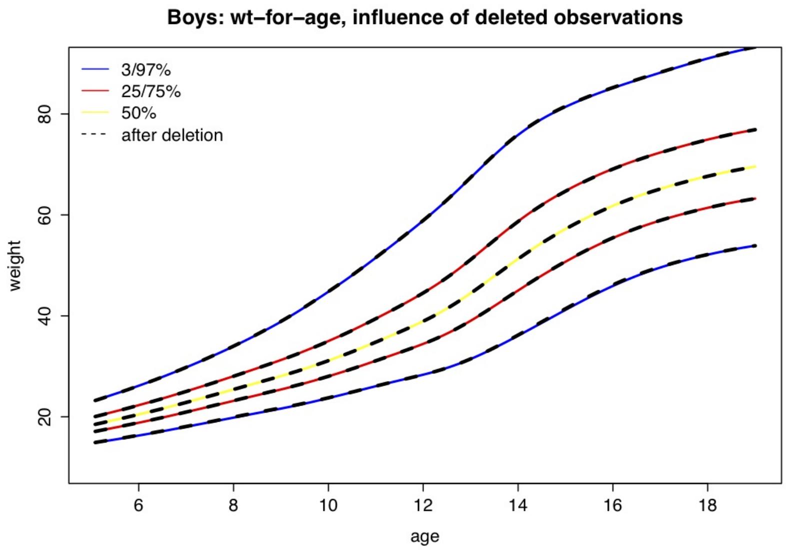


Figure 2.19: Influence diagnostics: The 8 outlying observations identified in the residual plot (figure 2.18) were deleted and the model re-fitted. Solid colored lines represent smoothed centiles calculated with the full dataset; dashed black lines are after deletion. There appears to be no appreciable influence on the fitted model.

Again, we examine the detrended Q-Q (worm) plots (figure 2.20). Here, a small number of observations fall outside the 95% CI expected under the null hypothesis, and the S-shaped pattern suggests residual kurtosis. The subplots for different age ranges localize the worst violations in plots corresponding to ages 8-10 and 11-12 (figure 2.21).

When worm plots suggest model inadequacy, further assessment may involve both Q and Z statistics, calculated for the same age intervals examined by worm plot[3, 11]. With an adequate model, Z_4 should be distributed as Normal[0,1]; given 20 distinct age groups, one aberrant value would be expected by random chance alone (p=0.05). Nevertheless for Z_4, 3 outliers are at ages 5-6, 8-9, and 15-17. All are between 2-3 SD. While this fails the Q_4 test for adequacy (p=0.026), the kurtosis is mild.

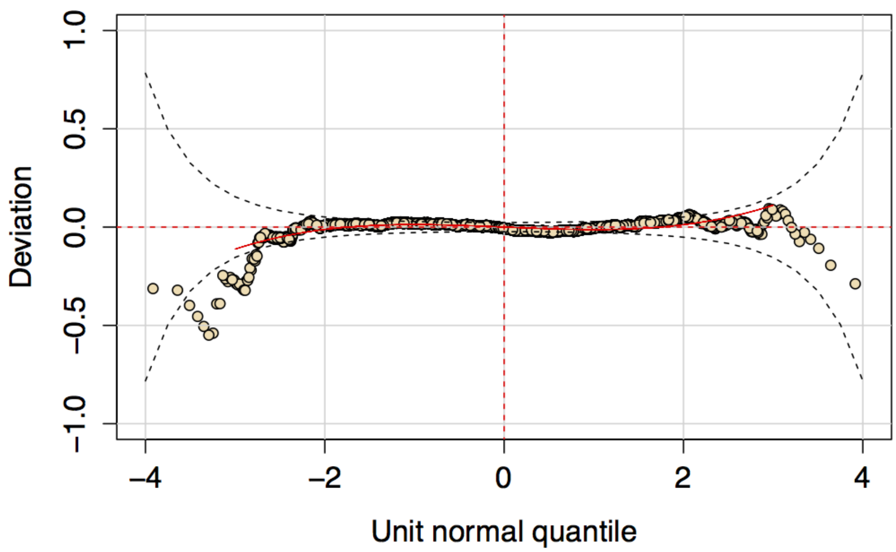


Figure 2.20: Worm plot of model residuals (boys). With an adequate model, detrended residuals should lie between the 2 dashed lines (95% confidence interval).

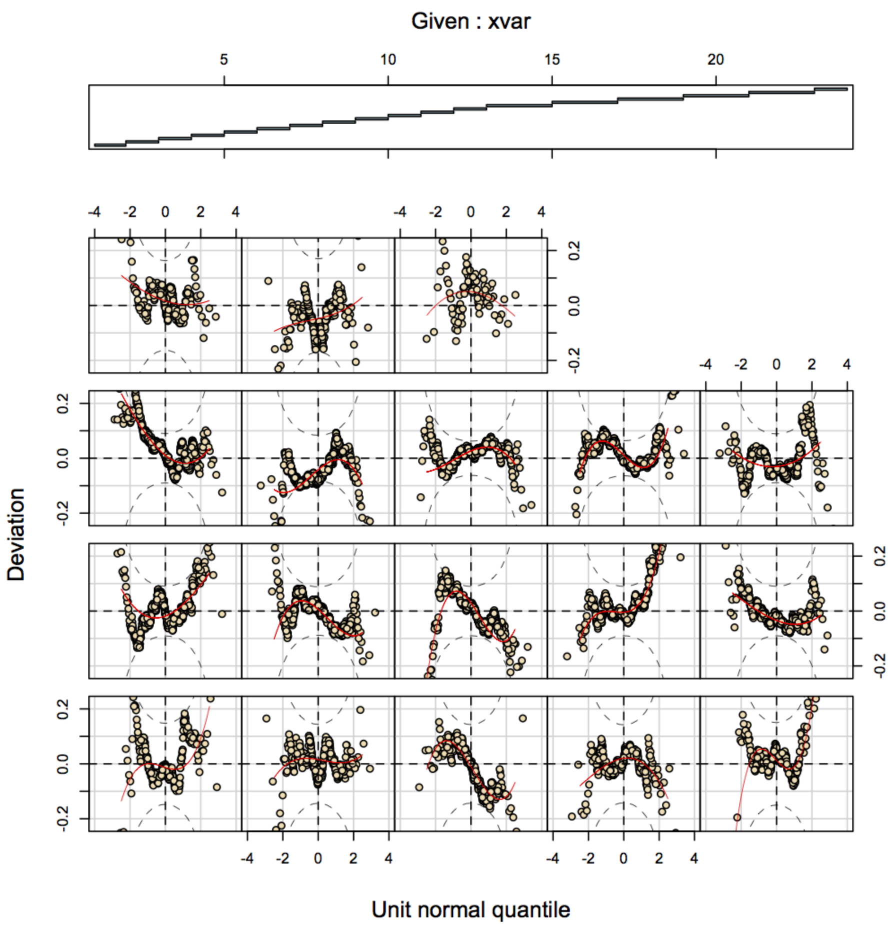


Figure 2.21: Worm plots by age: The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years and worm plots applied separately in each interval. There is some evidence for localized model violations.

```
> Q.stats(nchs1, xvar=d$ageyears,xcut.points=xc)

                           Z1        Z2        Z3        Z4     N
1.05133 to 2      -0.088531  0.36017  0.08609  1.37101    277
2 to 3             0.201546  0.04832 -0.20281  0.32746    292
3 to 4            -0.271896 -1.28881 -0.22886  1.12207    295
4 to 5             0.013548  0.13635 -0.81586 -0.16004    299
5 to 6             0.338305  1.02395  0.14495  2.49995    268
6 to 7             0.056381  0.70427  1.23421  0.08104    736
7 to 8            -0.254224 -0.97746 -1.25534  1.54249    775
8 to 9            -0.006025 -1.01716 -2.49230  2.95458    747
9 to 10            0.502672  1.88267  1.44590  1.75488    746
10 to 11          -0.644643 -0.57621  0.62523  0.31937    744
11 to 12           0.849042 -1.53210  1.38689  0.23539    746
12 to 13          -1.559644  0.98342 -0.63126 -0.83043    862
13 to 15           0.618951  0.41149 -1.03139 -0.24480    1555
15 to 17           0.543905 -0.43855  0.18508  2.61096    1480
17 to 19          -0.494821  0.42414  0.83176  0.09289     765
19 to 21           0.351003 -0.45722  0.21375  0.16072     227
21 to 23          -0.664408  0.27440  0.08876  0.11382     209
23 to 24           0.322081 -0.07804 -0.33627  0.01458      83
TOTAL Q stats      5.718627 13.36910 16.85824 31.48794    11106
df for Q stats     2.998787 12.49911 11.00092 18.00000    0
p-val for Q stats  0.126031  0.38087  0.11217  0.02526    0
```

Given a mild degree of 'model violation', WHO analysts would opt for the simpler 3-parameter LMS model, noting that the introduction of a kurtosis parameter τ rarely affected the predicted centiles significantly and rarely justified the 4-parameter BCPE distribution. The validity of this assumption can be demonstrated easily enough by the introducing kurtosis parameter τ and using `find.hyper()` and GAIC(3) to optimize its smoothing parameter df(τ). The modified model with λ=1.3, df(μ)=13, df(σ) = 8, df(ν)=5, and df(τ)=3 appears to remedy the kurtosis noted earlier (see figure 2.22). The aberrant Z4 and Q4 statistics have also resolved:

```
Q.stats(nchs1, xvar=d$ageyears,xcut.points=xc)

                           Z1        Z2       Z3        Z4     N
1.05133 to 2      -0.001696  0.32086  0.1415  0.70607    277
2 to 3             0.255689  0.05454 -0.2305 -0.33601    292
3 to 4            -0.294651 -1.25505 -0.2233  0.45061    295
4 to 5            -0.020853  0.18218 -0.7168 -1.02534    299
5 to 6             0.306511  0.90352  0.1771  1.63126    268
6 to 7             0.062806  0.70616  1.0322 -1.08255    736
7 to 8            -0.285306 -0.91364 -1.0403  0.19982    775
8 to 9            -0.048112 -0.99025 -2.0509  1.71717    747
```

```
9 to 10             0.509261  1.81723  1.3633  0.73807    746
10 to 11           -0.588170 -0.51304  0.6797 -0.45667    744
11 to 12            0.897826 -1.50146  1.3512 -0.26044    746
12 to 13           -1.534699  0.94430 -0.5185 -1.26389    862
13 to 15            0.622502  0.39521 -0.9602 -1.00054   1555
15 to 17            0.597121 -0.41064  0.1864  1.48294   1480
17 to 19           -0.447930  0.41796  0.8320 -0.74464    765
19 to 21            0.365691 -0.44766  0.1981 -0.20442    227
21 to 23           -0.654173  0.26623  0.1155 -0.01634    209
23 to 24            0.321355 -0.07213 -0.3412  0.03684    083
TOTAL Q stats       5.710568 12.33222 13.2550 14.90397   11106
df for Q stats      2.998822 12.49912 11.0009 13.53995      0
p-val for Q stats   0.126475  0.46000  0.2771  0.35149      0
```

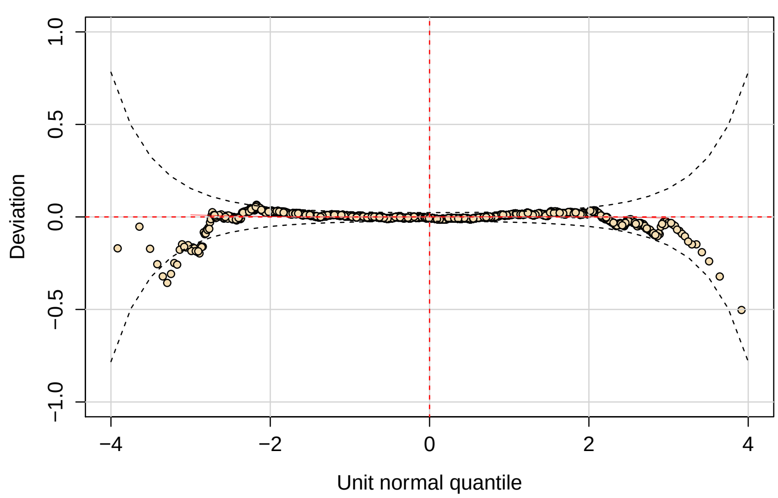


Figure 2.22: Same as figure 2.21 except the 3-parameter BCCG (LMS) model has been replaced by the 4-parameter BCPE model (boys), with the introduction of a kurtosis parameter τ. Worm-plot evidence of kurtosis has largely resolved.

Clearly, introducing τ in the BCPE model remedies the residual kurtosis. Nevertheless, simplicity favors the simpler 3-parameter model if feasible. Inspecting the predicted centiles in figure 2.23 suggests that the addition of τ to the model only affects prediction of the more extreme centiles (e.g. 0.135%, 99.9%), with general agreement between percentiles 3-97. We can also quantify the mean absolute deviation between the two models by comparing monthly predictions. Over ages 2-19 years, the mean discrepancy (%) is negligible except on the outlying percentiles (e.g. 0.135, 99.9%):

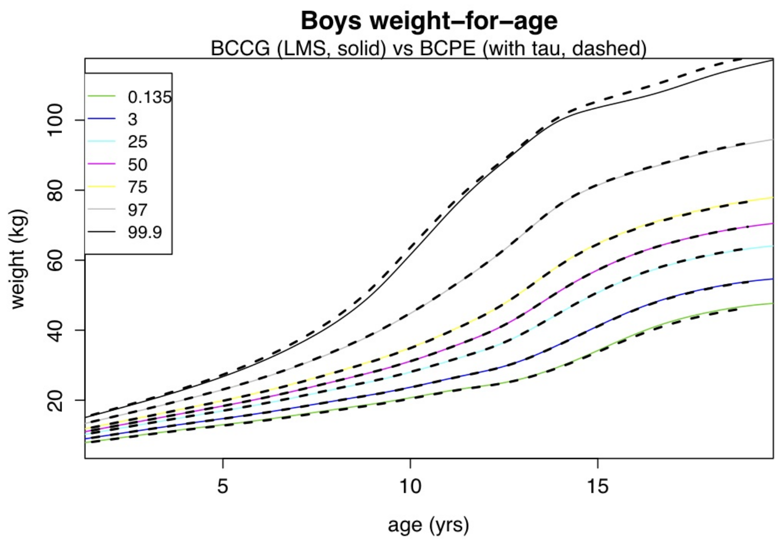


Figure 2.23: Smoothed centiles from BCCG model (solid colors) and BCPE model (dashed lines) applied to boys weight-for-age data. The 4-parameter BCPE model adjusts for kurtosis (non-normal tail frequencies) in the distribution of weight-for-age. Adjusting for kurtosis has little effect on the fitted model.

```
C0.135       C3      C25      C50      C75      C97    C99.9
1.30322  0.21406  0.23802  0.02798  0.27840  0.24497  2.11401
```

In fact, between percentiles 3 to 97, the two models generally differ only in the second place after the decimal. Given this relatively mild deviation from normality, we also decide to retain the simpler LMS model, with smoothing parameters, df(μ)=13, df(σ)=8, df(ν)=5 given λ=1.3 (boys).

2.12 Comparisons with WHO model on NCHS dataset

Given the overlap between datasets, exclusion criteria, and fitting methodologies, it would be reassuring to know that the above models were close to those developed by the WHO analysts using the full data. By "close", we are not referring to the hyperparameters - i.e. the optimal values of λ, df(μ), df(σ), df(ν) - or even to the LMS parameters (ν, μ, σ) specifically, but to model performance i.e. the smoothed centiles when corresponding models are fit to the same dataset. For boys, the hyperparameters for the optimal WHO model were λ=1.4, df(μ) = 10, df(σ)=8, and df(ν)=5[1]. This model can be fitted directly to the NCHS dataset (N=22917) and the smoothed centiles compared with those obtained using λ=1.3, df(μ)=13, df(σ)=8, df(ν)=5 (figure 2.24). The same comparison was done for girls i.e. λ=1.3, df(μ)=10,df(σ) = 3, df(ν)=3 vs hyperparameters λ=1.225, df(μ)=14 df(σ) = 6, df(ν) = 5 (figure 2.25). In both cases, the results are indistinguishable, attesting to the closeness of the two models in performance terms. In the latter case (girls), the mean absolute deviation is 0.44% (over the full range of monthly predictions from 2-19 years for centiles 3-97).

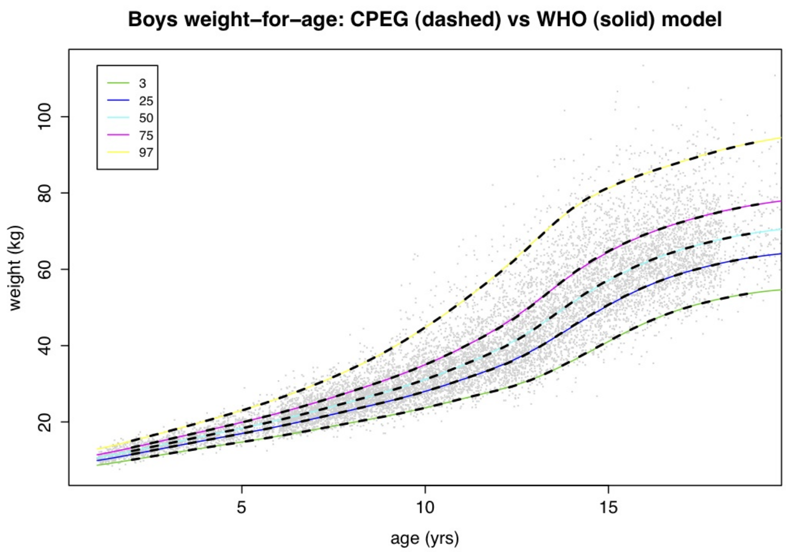


Figure 2.24: Comparison of smoothed centiles from 2 LMS models fitted to the same data: Boys, weight-for-age Solid color lines represent a model with hyperparameters λ=1.4, df(μ)=10, df(σ) = 8, df(ν)=5, and dashed lines represent a model with hyperparameters λ=1.3, df(μ)=13, df(σ)=8, df(ν)=5. In both cases, the models were identified through minimization of appropriate GAIC. Smoothed centiles from the two models are indistinguishable over a range of $\pm$ 2SD.

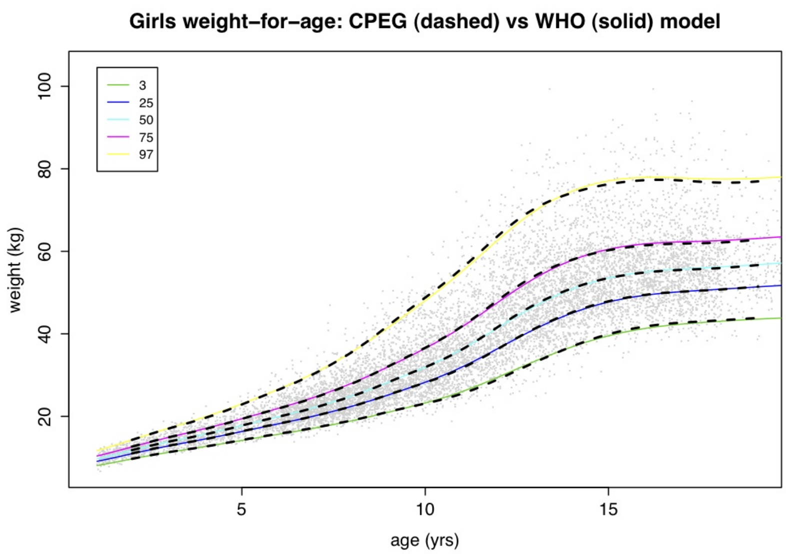


Figure 2.25: Comparison of smoothed centiles from 2 LMS models fitted to the same data: Girls, weight-for-age. Solid color lines represent a model with hyperparameters λ=1.3, df(μ)=10, df(σ) = 3, df(ν)=3, and dashed lines represent a model with hyperparameters λ=1.225, df(μ)=14, df(σ)=6, df(ν)=5. Smoothed centiles from the two models are indistinguishable over a range of $\pm$ 2SD.

Chapter 3

Height-for-age

There is no clinical need to re-analyze height-for-age or BMI-for-age, as the WHO has prepared normative reference curves for school age children and adolescents (5-19 years) based on the NCHS dataset. To smooth the transition between their *standard* curves (0-5 years, based exclusively on MGRS data) and *reference* curves (5-19 years, based largely on NCHS data), they applied 2 preparatory steps:

- Deeming them "unhealthy", they arbitrarily trimmed ≈ 3% of the NCHS population, most from above the 97.7^{th} percentile in weight-for-height. This maneuever - applied only to weight-based measures - can easily be reproduced, as was done in the previous chapter.

- They also added data on more than 8000 18-71 month olds from the cross-sectional phase of the MGRS (8306 observations on 6697 children, 3450 boys and 3219 girls). The fact that the latter data are not in the public domain means that any re-analysis (e.g. to extend the weight-for-age curves to older children) must rely on only the core NCHS data. While we expect the impact of the younger children on the curves for older children to be small, this is an assumption that must be tested given the sometimes unpredictable nature of non-linear curve fitting.

For this reason, we propose re-fitting both height-for-age and BMI-for-age curves using only the NCHS data. By comparing these results with WHO curves based on the full data over the full age-range (5-19 years to avoid edge effects), we hope to better assess the impact of the missing 8306 youngsters. Data preparation and model identification proceed as before. Once again, we remind the reader that too simple a model may lead to underfitting , which can smooth away important feature (like growth spurts). Conversely, overfitting random fluctuations leads to spurious trends.

3.1 Model identification

The optimal power transform in the WHO analysis was determined by sensitivity analysis using arbitrary models to minimize global deviance[1]. Before the LMS parameters could be estimated, optimal smoothing models were also specified, based on minimization of GAIC(2) and GAIC(3) criteria. Final models were described in de Onis *et al*[1]. For height-for-age in boys, λ=1, df(μ)=12, df(σ)=4, ν=1, and τ=2. For girls, λ=0.85, df(μ)=10, df(σ)=4, ν=1, and τ=2. Assuming that differences in the data will lead to small changes in optima, these basic models were applied to the NCHS data (N=22,917, 1-24 years of age) to systematically evaluate the global deviance as λ was varied from 0.5-1.5 in steps of 0.025. Results are summarized below[2]

- Based on minimization of global deviance (girls, height-for-age), the age-axis requires a power transform with $\lambda = 0.725$ (figure 3.1A). The optimal power transform for the boys age-axis was λ=1.05 (figure 3.1B)

```
# Girls, height-for-age
Best estimate of the fixed parameter is  0.725
with a Global Deviance equal to  74050  at position  10

# Boys height-for-age
Best estimate of the fixed parameter is  1.05
with a Global Deviance equal to  74287  at position  23
A  95% Confidence interval is: ( 0.8431 , 1.218 )
```

- Girls, height-for-age: Taking the basic WHO model as starting point, we invoke `find.hyper()` to minimize GAIC(2) and GAIC(3). This function is based on the popular BFGS optimization engine, a 'steepest-descent' or 'hill-climbing' algorithm that systematically seeks out minima on the GAIC surface.

```
# penalty=2.0
$par[1] 17.958 11.782  8.363
$value [1] 74056

# penalty = 3.0
$par[1] 14.901  9.647  3.000
$value [1] 74097
```

- Since degrees of freedom can be optimized independently of each other, a sequential approach is usually taken; first fixing df(μ) followed by df(σ), df(ν) and df(τ)[3, 10, 7, 9]. In this case, the estimates are further refined through minimization of GAIC(3) with λ=0.725, df(μ)=18.

[1] a reasonable measure when model degrees of freedom are fixed

[2] Twenty-two subjects were excluded for 'outlying' heights-for-age. In generating their height-for-age norms, the WHO did not exclude on the basis of weights-for-height. Hence, analysis here is based on $n = 11,402$ boys and $n = 11,493$ girls

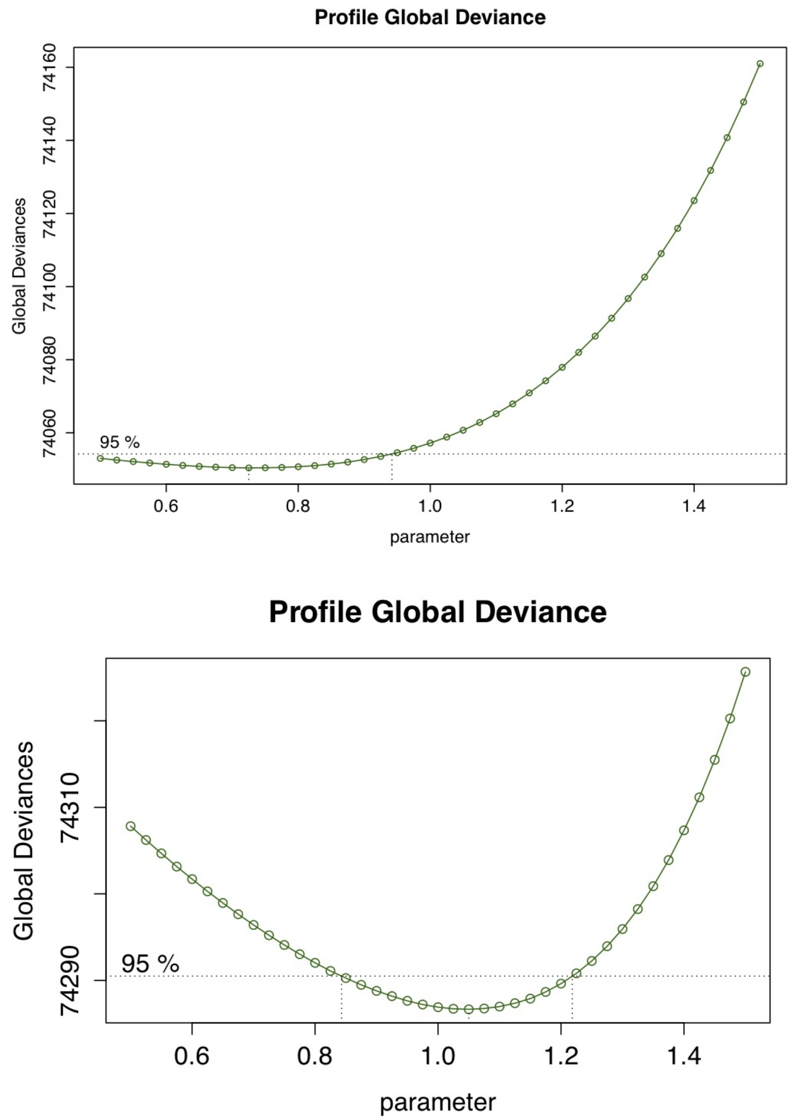


Figure 3.1: Global deviance as exponent λ of age transformation varied from 1.0 to 1.5

```
# penalty=3.0
$par[1] 9.684 3.000
$value [1] 74099
```

- Boys, height-for-age: Taking the basic WHO model as starting point, we invoke `find.hyper()` to minimize GAIC(2) and GAIC(3):

```
# penalty=2.0
$par [1] 14.293 11.035  4.308
$value [1] 74309

# penalty=3.0
$par [1] 12.157  9.304  3.000
$value [1] 74342
```

- These estimates are further refined through minimization of GAIC(3) with λ=1.05, df(μ)=14.

```
# penalty=3.0
$par[1] 9.282 3.000
$value[1] 74343
```

Sequential model optimization therefore yields λ=0.725, df(μ)=18, df(σ)=10, df(ν)=3 (girls, height-for-age) and λ=1.05, df(μ)=14, df(σ)=9, df(ν)=3 (boys, height-for-age).

3.2 Fitted model, boys

Evaluation of the fitted model follows the outline in the previous chapter (weight-for-age)

- Smoothed height-for-age centile for boys aged 5-19 are shown in figure 3.2.

- Comparison with sample centiles indicates that they are being accurately captured (figure 3.3):

```
% of cases below  3 centile is  2.903
% of cases below  25 centile is  24.88
% of cases below  50 centile is  49.67
% of cases below  75 centile is  75.37
% of cases below  97 centile is  96.91
```

- Even at the “outer” percentiles, the smoothed curves are very close to the WHO norms (figure 3.4). In fact, the mean absolute discrepancy in monthly measurements between 5-19 years of age is negligible.

```
97th percentile 0.4529 +- 0.3185 %
50th percentile 0.1328 +- 0.1424 %
3rd percentile 0.5095  +- 0.3678 %
```

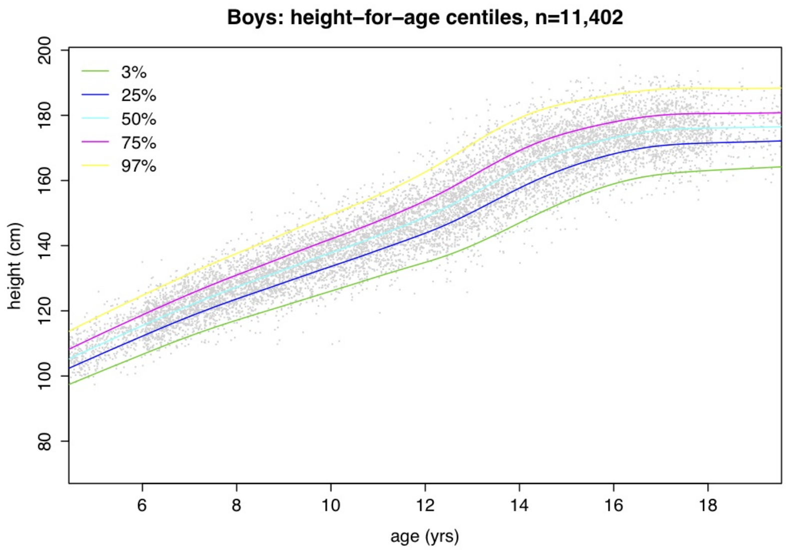


Figure 3.2: Boys smoothed centiles (5-19 years). Based on n=11,402 boys aged 1-24 years, NCHS data

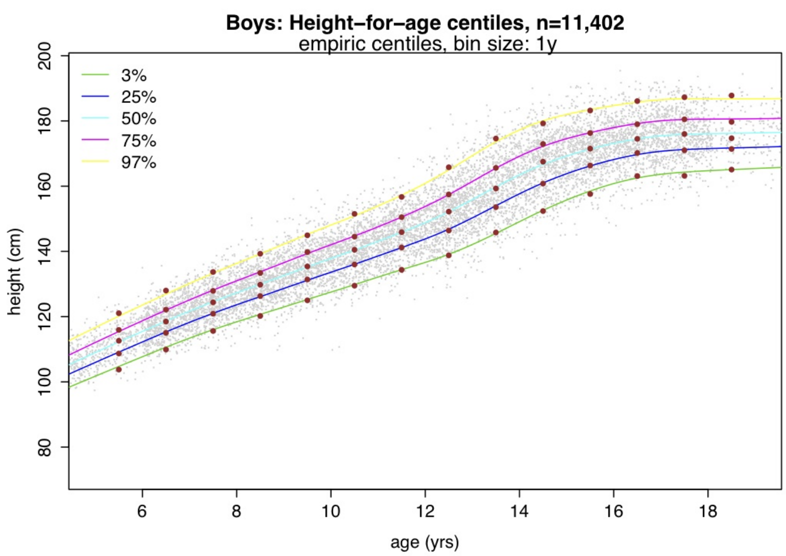


Figure 3.3: Boys 2-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year

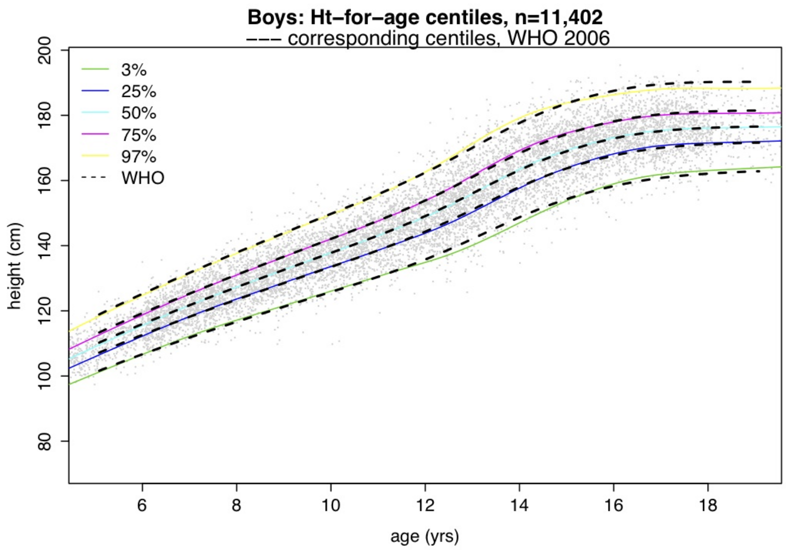


Figure 3.4: Boys smoothed centiles vs WHO centiles (5-19 years). The core NCHS data were used for both analyses, but the WHO added 8306 additional observations (ages 18-71 months) to smooth the transition between WHO and NCHS data.

3.3 Fitted model, girls

Evaluation of the fitted model follows the same outline:

- Smoothed height-for-age centile for girls aged 5-19 are shown in figure 3.5.

- Comparison with sample centiles indicates that they are being accurately captured (figure 3.6):

```
% of cases below  3 centile is  3.002
% of cases below  25 centile is  24.55
% of cases below  50 centile is  50.11
% of cases below  75 centile is  75.63
% of cases below  97 centile is  97.07
```

- Even at the "outer" percentiles, the smoothed curves are very close to the WHO norms (figure 3.7). In fact, the mean absolute discrepancy in monthly measurments between 5-19 years of age is negligible.

```
97th percentile 0.2671 +- 0.1610 %
50th percentile 0.1936 +- 0.1612 %
3rd percentile 0.4126  +- 0.3898 %
```

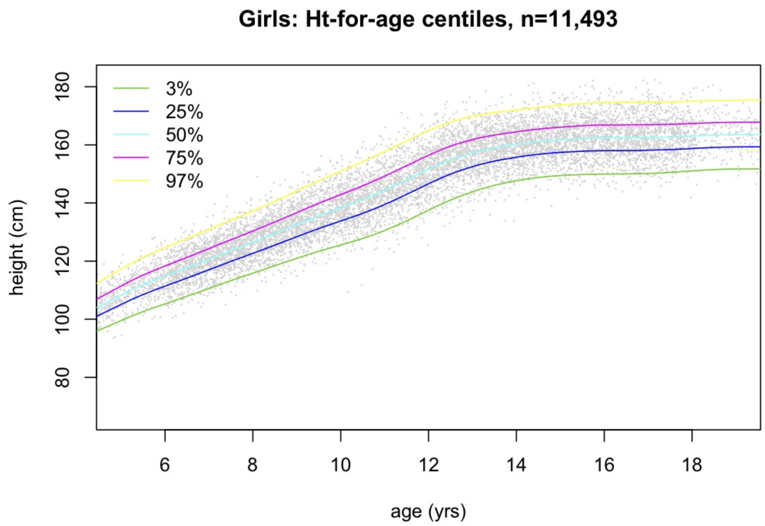


Figure 3.5: Girls smoothed centiles (5-19 years). Based on n=11,493 girls aged 1-24 years, NCHS data

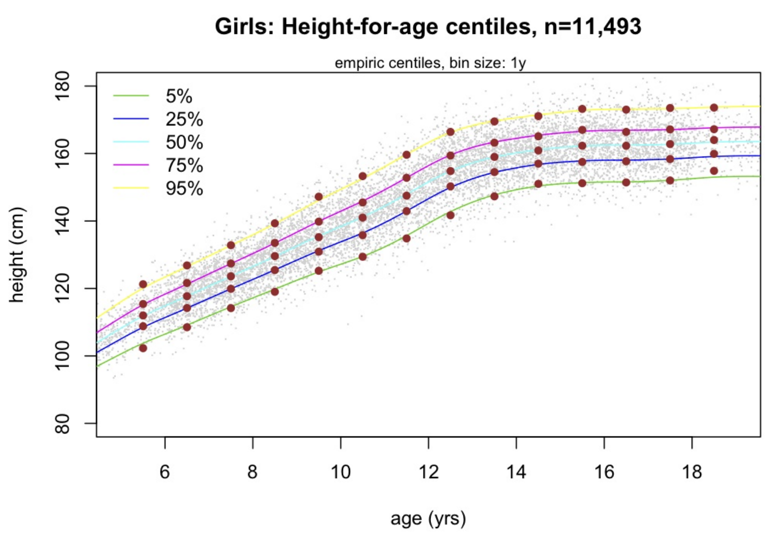


Figure 3.6: Girls 5-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year

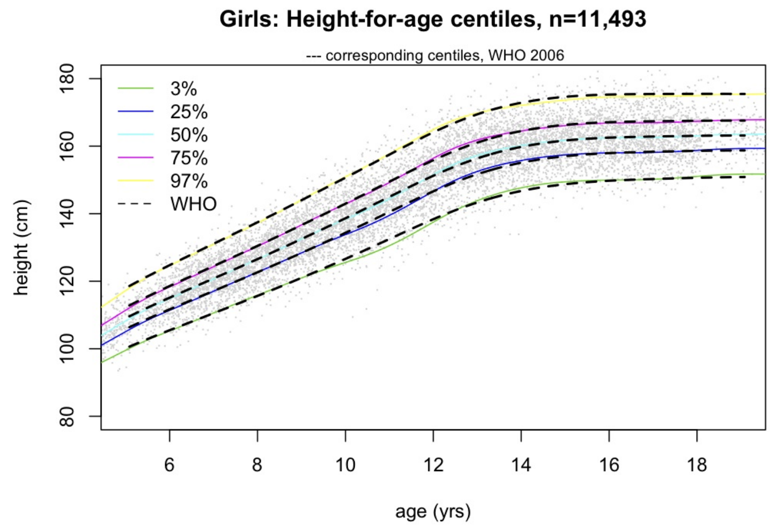


Figure 3.7: Girls smoothed centiles vs WHO centiles (5-19 years). The core NCHS data were used for both analyses, but the WHO added 8306 additional observations (ages 18-71 months) to smooth the transition between WHO and NCHS data.

3.4 Model Diagnostics, boys

This assessment, too, follows on the one outlined in the previous chapter (weight-for-age):

- Formally assessing normality of the residuals does not indicate any problems with model fit (figure 3.8) and outliers signal no cause for alarm:

```
Summary of the Quantile Residuals
                           mean   =  7.241e-05
                       variance   =  1
               coef. of skewness  =  -0.0004472
               coef. of kurtosis  =  3.194
Filliben correlation coefficient  =  0.9997
```

- Worm plot shows virtually all de-trended residuals fall within the 95% confidence interval (figure 3.9). Curiously, application of the WHO optimal model to the NCHS data showed evidence for model inadequacy, which resolved with model optimization (not shown). The WHO model was based on the larger dataset with an additional 8306 younger children.
- Worm plots of specific age intervals show no evidence for localized model violations (figure 3.10).

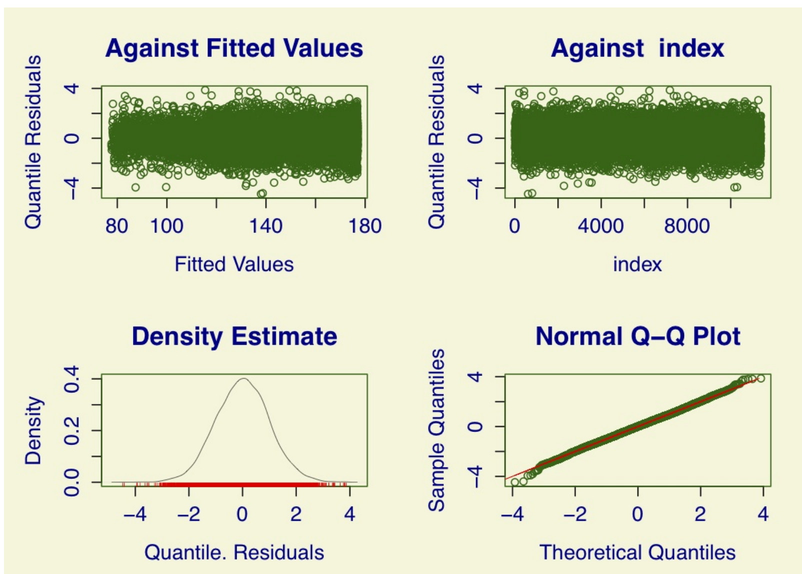


Figure 3.8: Residual plots for assessment of model assumptions (boys). With an adequate fit, residuals should be normally distributed with mean=0, sd=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals

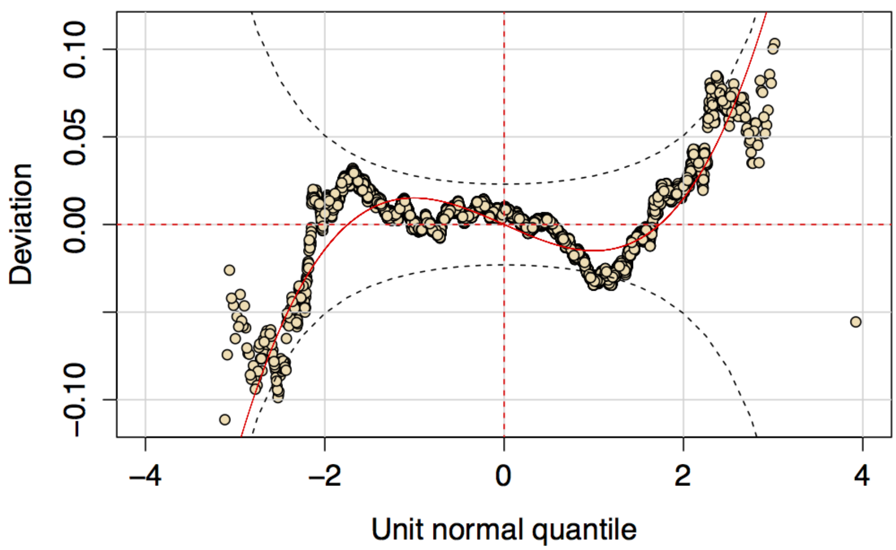


Figure 3.9: Worm plot of model residuals (boys). With an adequate model, detrended residuals should lie between the 2 dashed lines (95% confidence interval).

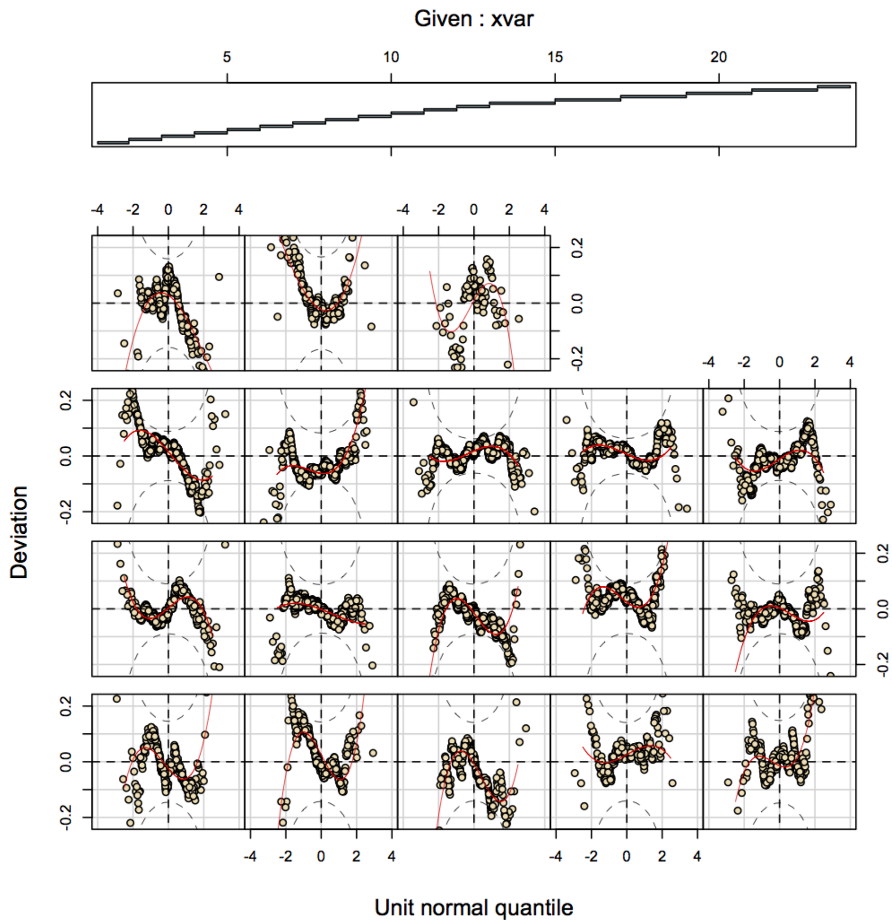


Figure 3.10: Worm plot by age: The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, and worm plots applied separately in each interval. There is no evidence for localized model violations.

- Z and Q statistics are confirmatory. All Z_1, Z_2 and $Z_3 < 2$ speak to model adequacy for μ, σ and ν. There is some evidence of mild kurtosis with $Z_4 \approx$ 2-3, but the worm plot is reassuring. Moreover, it is easy to demonstrate that mild kurtosis has little impact on smoothed centiles (not shown). This is an example of why Q statistics have to be interpreted cautiously, since there is a marginal p-value associated with Q_3 (p=0.038), which suggests residual skew. In light of the worm plot and all $Z_3 < 2$, this is a spurious result.

```
Q.stats(nchs1, xvar=d$ageyears,xcut.points=xc)
                          Z1      Z2       Z3          Z4      N
1.05133 to 2      -0.14271 -0.1237  0.84553  1.740e+00    286
2 to 3             0.36780 -0.1985 -0.54440  2.734e+00    297
3 to 4            -0.77246 -0.2522 -1.79260  2.374e+00    304
4 to 5             0.46648  0.3648  0.06151 -4.286e-01    306
5 to 6             0.28855  1.1755  1.43996  1.878e+00    276
6 to 7             0.11254  0.3301  0.05211 -2.078e+00    751
7 to 8            -0.15583 -0.7272 -0.27373  4.073e-01    798
8 to 9            -0.70074 -0.4956 -0.80903  3.097e+00    767
9 to 10            1.20901  0.3096  1.53325  2.347e+00    772
10 to 11          -0.44910  0.3424 -1.79681  2.621e+00    760
11 to 12           0.30999 -1.9980 -0.27847  1.245e+00    769
12 to 13          -1.14042  1.1231  1.83528  1.748e+00    880
13 to 15           0.48781  0.6653 -0.56189 -1.080e+00   1592
15 to 17           0.35285 -0.7258  0.30211  9.457e-01   1510
17 to 19          -0.40063  0.7035 -0.19118 -1.012e+00    786
19 to 21          -0.27577 -0.3657 -1.78659  6.959e-01    239
21 to 23           0.37894 -0.1307  1.82936  5.978e-01    220
23 to 24          -0.12063  0.3089 -0.49735 -1.374e+00    89
TOTAL Q stats      5.39840  9.6934 23.28868  5.672e+01  11402
df for Q stats     2.00245 11.9991 13.00061  1.800e+01      0
p-val for Q stats  0.06741  0.6428  0.03835  6.858e-06      0
```

3.5 Model Diagnostics, girls

By now, this should be a familiar routine.

- Formal assessment of model residuals confirms their normality (figure 3.11) and outliers signal no cause for alarm:

```
                        mean  =  -7.377e-05
                    variance  =  1
           coef. of skewness  =  0.003411
           coef. of kurtosis  =  3.164
Filliben correlation coefficient  =  0.9998
```

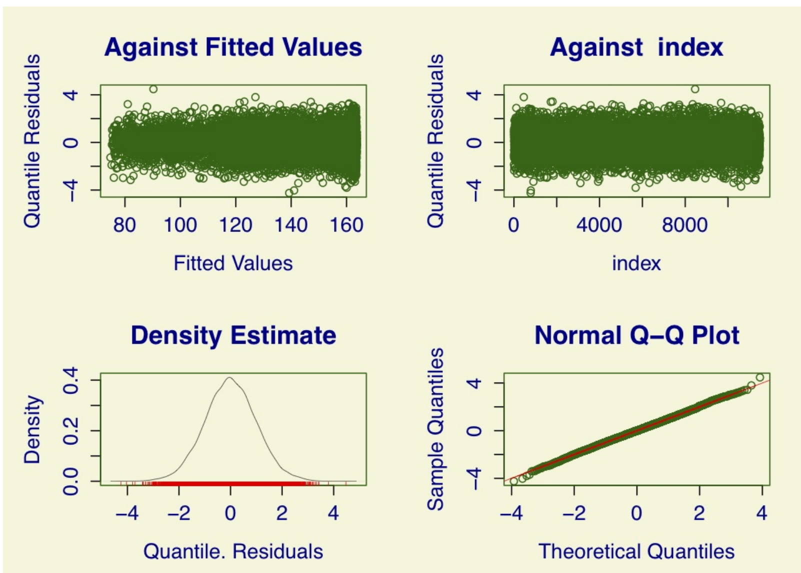


Figure 3.11: Residual plots for assessment of model assumptions (girls). With an adequate fit, residuals should be normally distributed with mean=0, sd=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals

- Worm plot shows virtually all de-trended residuals fall within the 95% confidence interval (figure 3.12).

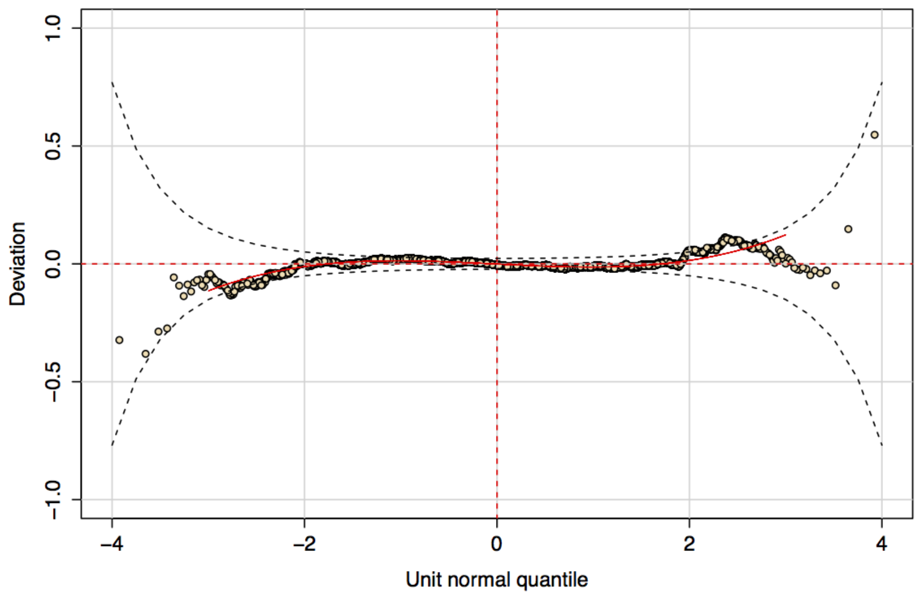


Figure 3.12: Worm plot of model residuals (girls). With an adequate model, de-trended residuals should lie between the 2 dashed lines (95% confidence interval). The path of the smoothed curve (solid red) can identify specific model violations.

- Worm plots of specific age intervals show no evidence for localized model violations (figure 3.13).

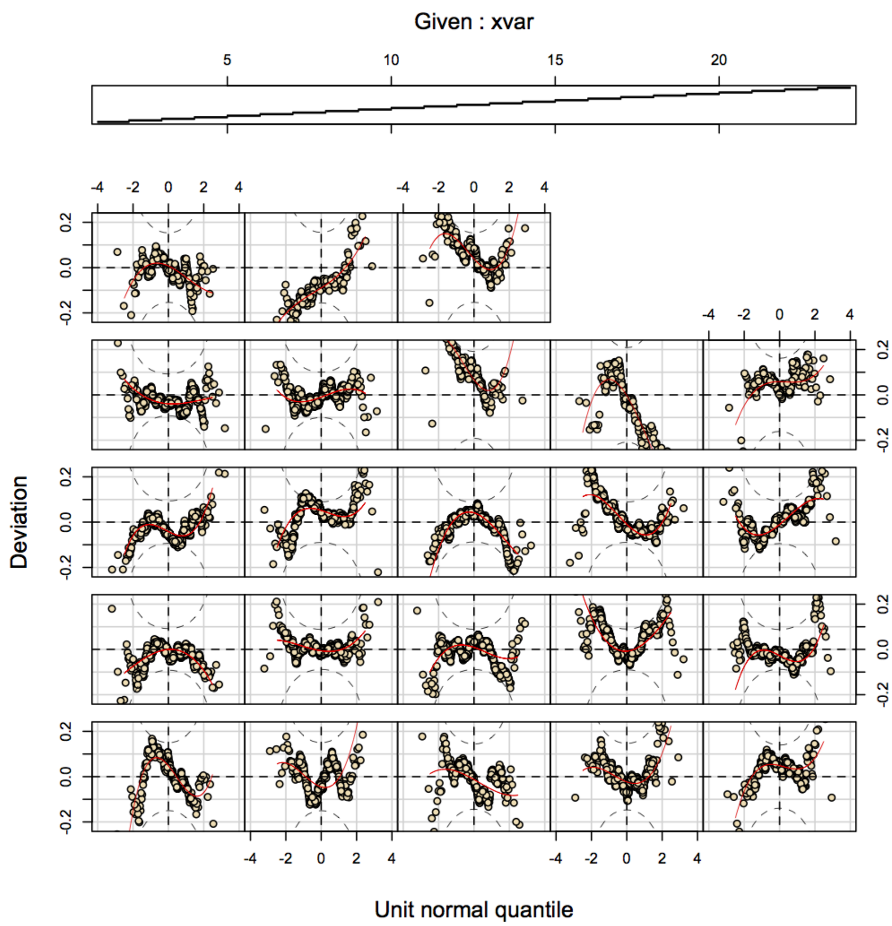


Figure 3.13: The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, with both worm plots and Z statistics applied separately in each interval. There is no evidence for localized model violations.

Chapter 4

BMI-for-age

As discussed previously, we have undertaken to re-fit both height-for-age and BMI-for-age curves using only the NCHS data. By comparing these results with WHO curves based on the full data over the full age-range (5-19 years), we hope to assess the impact of the missing 8306 youngsters.

4.1 Model identification

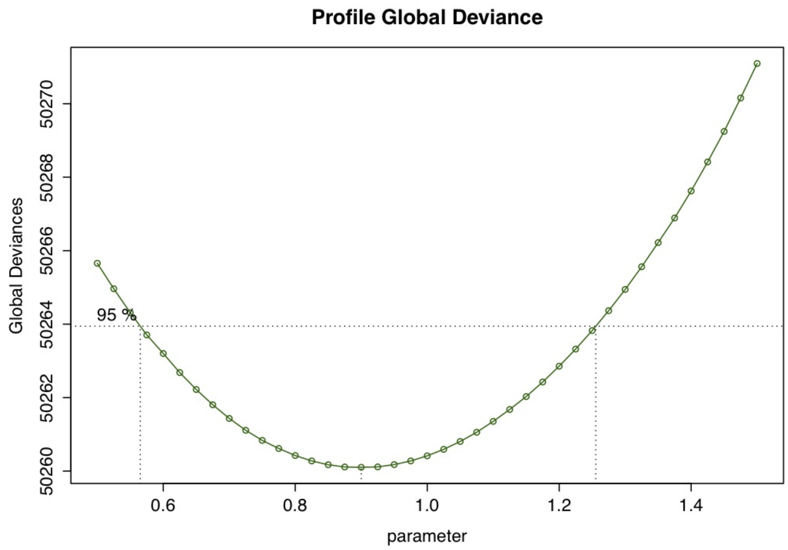
The optimal power transform in the WHO analysis was determined by sensitivity analysis using arbitrary models to minimize global deviance. Before the LMS parameters could be estimated, optimal smoothing models were also specified, based on minimization of GAIC(2) and GAIC(3) criteria. Final models were described in de Onis *et al*[1]. For BMI-for-age, the final model for boys was λ=0.8, df(μ)=8, df(σ)=4, df(ν)=4, and τ=2. For girls, λ=1, df(μ)=8, df(σ)=3, df(ν)=4, and τ=2. Here, analysis here is based on n=11,106 boys and n=11,193 girls

- For the "cleaned" BMI-for-age data, the optimal values were $\lambda = 0.9$ (girls) and λ=0.6 boys (figure 4.1).

```
# Girls: BMI-for-age
Best estimate of the fixed parameter is  0.9
with a Global Deviance equal to  50260  at position  17
A  95 % Confidence interval is: ( 0.5651 , 1.256 )

# Boys: BMI-for-age
Best estimate of the fixed parameter is  0.6
with a Global Deviance equal to  46311  at position  5
```

- Girls, BMI-for-age: Taking the basic WHO model as starting point, we invoke `find.hyper()` to minimize GAIC(2) and GAIC(3). This function



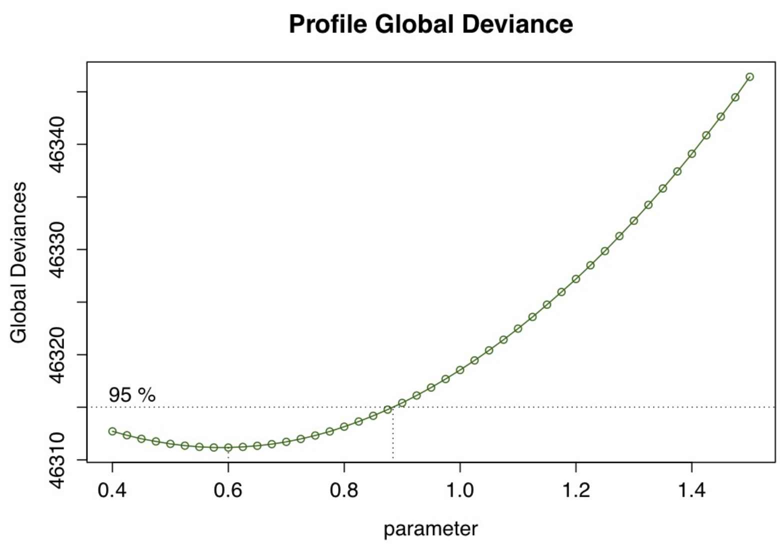


Figure 4.1: A) Girls, B) Boys: Global deviance as exponent λ of age transformation varied from 1.0 to 1.5

is based on the BFGS optimization algorithm that systematically seeks out minima on the GAIC surface.

```
# penalty=2.0
$par [1] 9.770 6.013 5.016
```

```
$value [1] 50272


# penalty = 3.0
$par [1] 7.884 5.518 4.101
$value [1] 50297
```

- Since degrees of freedom can be optimized independently of each other, a sequential approach is usually taken[3, 10, 7, 9]. In this case, the estimates are further refined through minimization of GAIC(3) with λ=0.9, df(μ)=10.

```
# penalty=3.0
$par [1] 6.790 4.762
$value [1] 71469
```

- Boys, BMI-for-age: Taking the basic WHO model as starting point, we invoke `find.hyper()` to minimize GAIC(2) and GAIC(3):

```
# penalty=2.0
$par[1] 10.021  6.073  4.595
$value  [1] 46351


# penalty=3.0
$par [1] 7.727 5.197 3.395
$value [1] 46375
```

- These estimates are further refined through minimization of GAIC(3) with λ=0.6, df(μ)=10.

```
# penalty=3.0
$par[1] 5.199 3.267
$value [1] 46376
```

Sequential model optimization therefore yields λ=0.9, df(μ)=10, df(σ)=7, df(ν)=5 (girls, BMI-for-age) and λ=0.6, df(μ)=10, df(σ)=5, df(ν)=3 (boys, BMI-for-age).

4.2 Fitted model, boys

Evaluation of the fitted model follows the outline in the previous chapters

- Smoothed BMI-for-age centile for boys aged 5-19 are shown in figure 4.2.

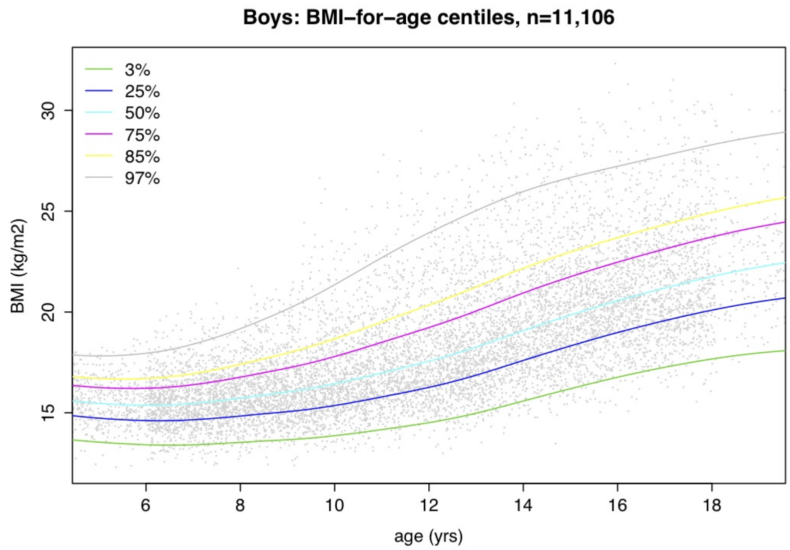


Figure 4.2: Boys smoothed centiles (5-19 years). Based on n=11,106 boys aged 1-24 years, NCHS data

- Comparison with sample centiles indicates that they are being accurately captured (figure 4.4):

```
% of cases below  3 centile is  3.025
% of cases below  25 centile is  24.82
% of cases below  50 centile is  51.31
% of cases below  75 centile is  75.18
% of cases below  97 centile is  96.65
```

- For the $3^{rd}$to 85^{th} percentiles, smoothed centiles are close to the WHO norms throughout the 5-19 year age range (figure 4.4), but there is a visible discrepancy in the 97^{th} percentile. It peaks at 12-13 years of age and dissipates after 14 years. Its mean absolute value is 1.8 ± 1.1 %, measured at monthly intervals over the range 5-19 years. This is considerably larger than the corresponding 50th percentile value of 0.2 ± 0.2 %. Nevertheless, the discrepancy is small and short-lived. Agreement is good for the $3^{rd}$to 85^{th} percentiles

```
97th percentile:  1.78 +- 1.106 %
50th percentile   0.232 +- 0.1846%
3rd percentile  0.7072 +- 0.587%
```

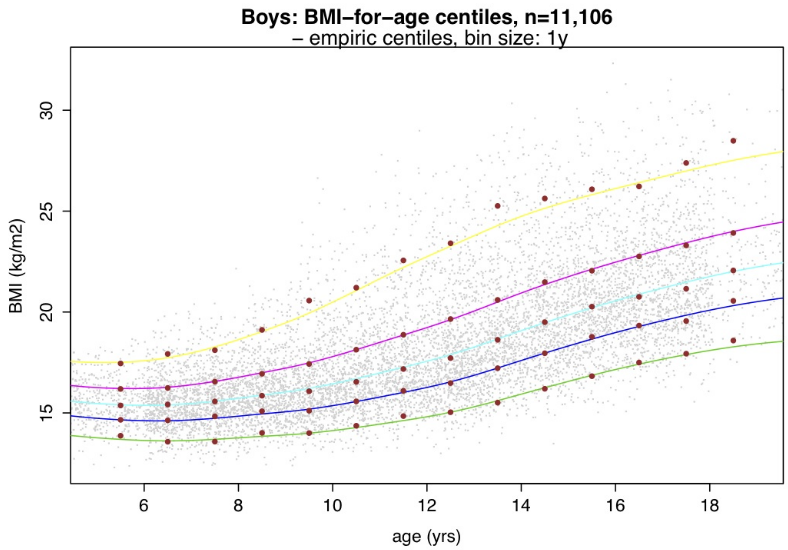


Figure 4.3: Boys 5-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year

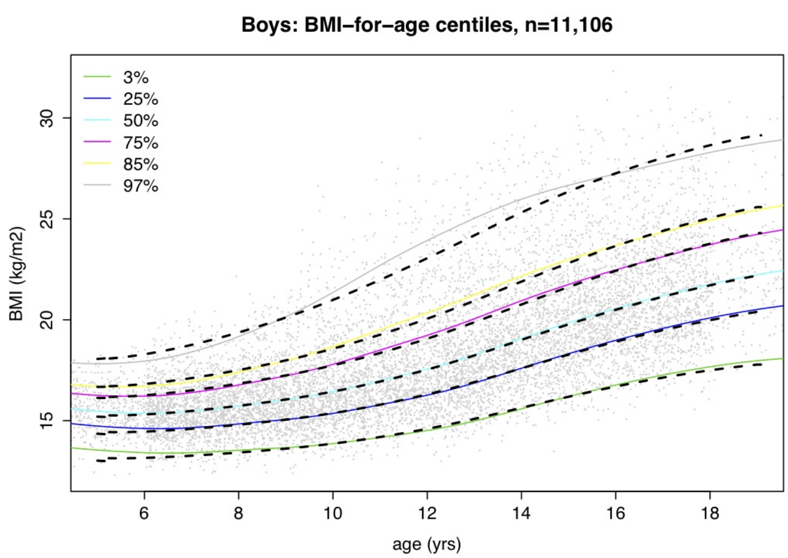


Figure 4.4: Boys smoothed centiles vs WHO centiles (5-19 years).The core NCHS data were used for both analyses, but the WHO added 8306 additional observations (ages 18-71 months) to smooth the transition between WHO and NCHS data.

4.3 Fitted model, girls

Evaluation of the fitted model follows the same outline as before:

- Smoothed BMI-for-age centile for girls aged 5-19 are shown in figure 4.5.

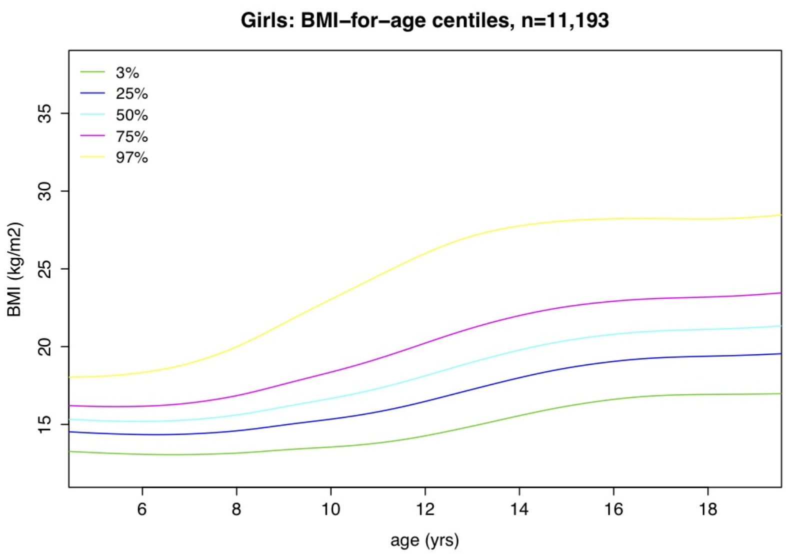


Figure 4.5: Girls smoothed centiles (5-19 years). Based on n=11,193 girls aged 1-24 years, NCHS data

- Comparison with sample centiles indicates that they are being accurately captured (figure 3.6):

```
% of cases below  3 centile is  2.796
% of cases below  25 centile is  25.2
% of cases below  50 centile is  50.88
% of cases below  75 centile is  75.03
% of cases below  97 centile is  96.63
```

- For percentiles 3-75, the smoothed curves are close to the WHO norms (figure 4.7). Greater divergence is seen in the upper percentiles (e.g.> 75^{th}). This impression is confirmed by the mean absolute discrepancies in monthly measurements between 5-19 years of age :

```
97th percentile: 3.184 +- 1.971%
50th percentile: 0.5972 +- 0.3211%
3rd percentile: 1.374 +- 0.5918%
```

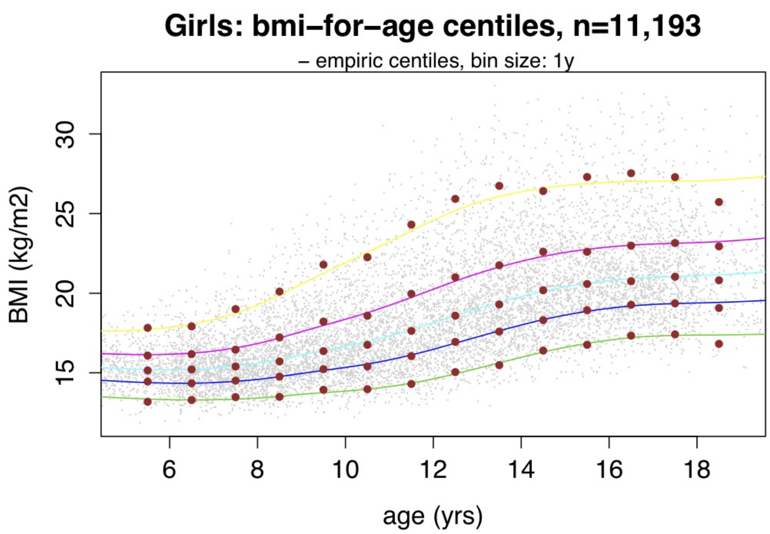


Figure 4.6: Girls 5-19 years, smoothed (lines) vs sample (dots) centiles, the latter calculated for a bin size of 1 year

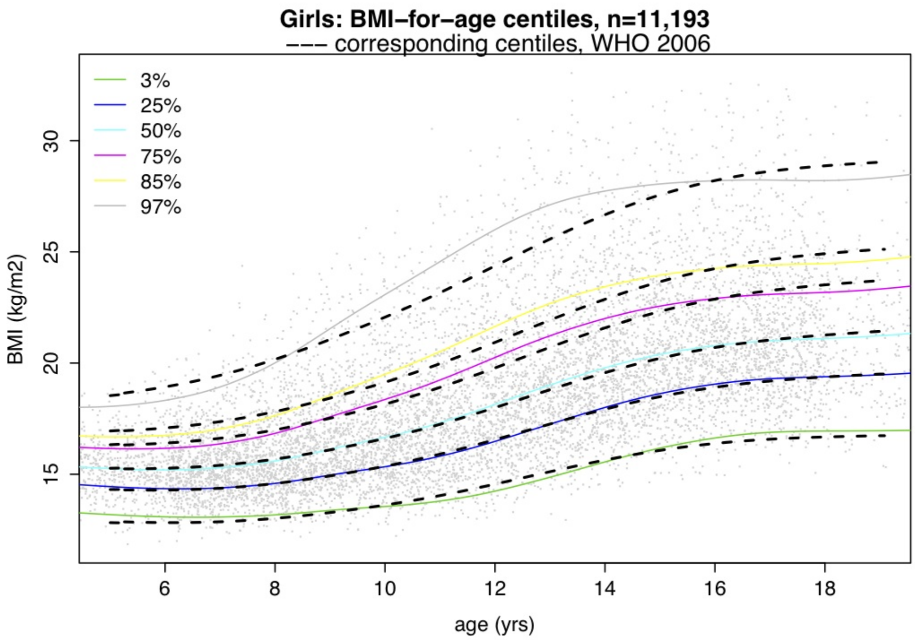


Figure 4.7: Girls smoothed centiles vs WHO centiles (5-19 years).The core NCHS data were used for both analyses, but the WHO added 8306 additional observations (ages 18-71 months) to smooth the transition between WHO and NCHS data.

4.4 Model Diagnostics, boys

By now, this should be a numbingly familiar routine.

- Formal assessment of model residuals confirms their normality (figure 4.8) and outliers signal no cause for alarm:

```
Summary of the Quantile Residuals
                           mean   =  0.0004669
                       variance   =  1
               coef. of skewness  =  0.007037
               coef. of kurtosis  =  2.874
Filliben correlation coefficient  =  0.9992
```

- The worm plot shows evidence of mild kurtosis, with an S-shaped pattern and a points outside the 95% confidence limits (figure 4.9).

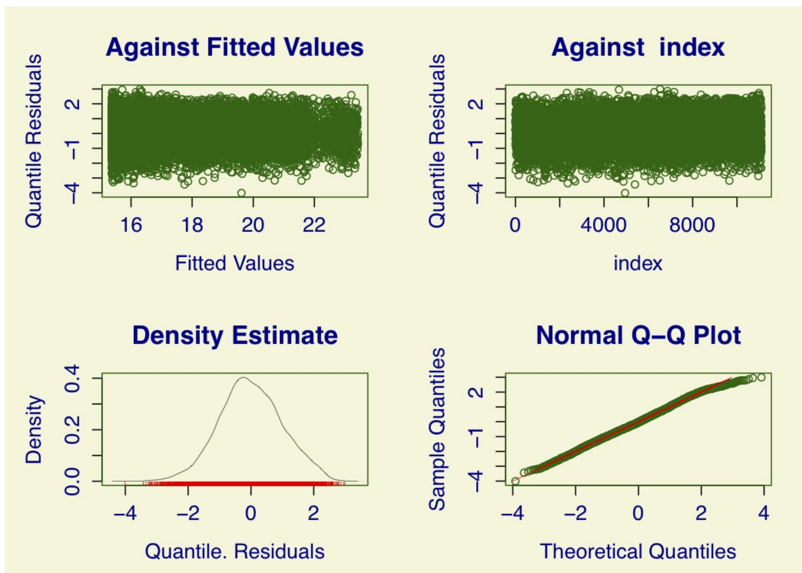


Figure 4.8: Residual plots for assessment of model assumptions (boys). With an adequate fit, residuals should be normally distributed with mean=0, sd=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals

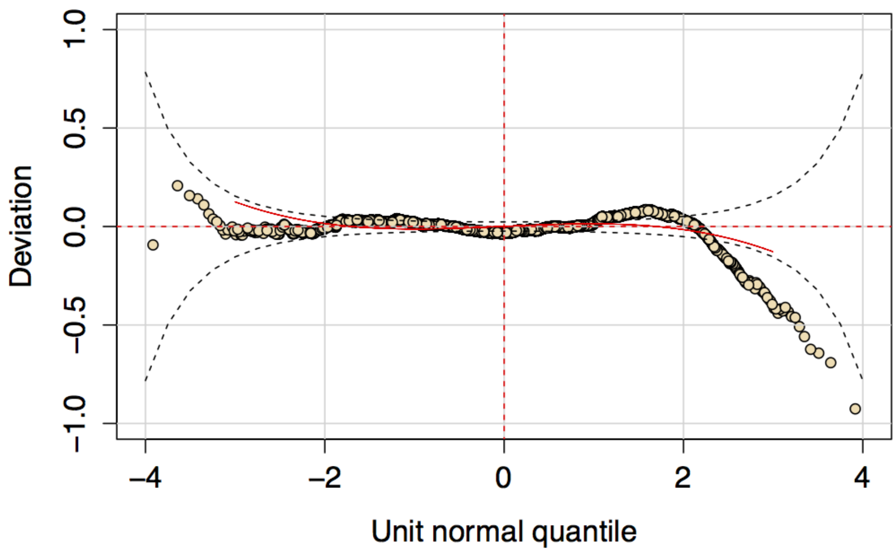


Figure 4.9: Worm plot of model residuals (boys). With an adequate model, detrended residuals should lie between the 2 dashed lines (95% confidence interval). The path of the smoothed curve (solid red) can identify specific model violations.

- Worm plots on specific age intervals also show evidence for localized model violations at 9-10 and 11-12 years (figure 4.10). However, the excursions outside the 95% confidence limits are few and mild.

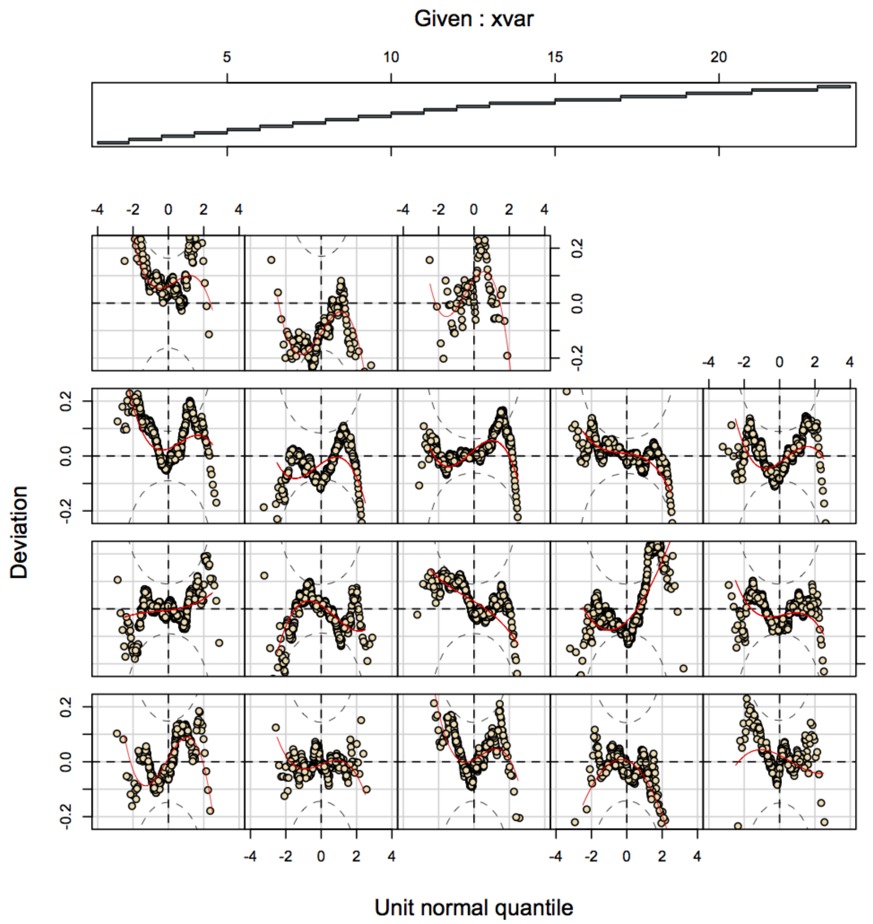


Figure 4.10: The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, and worm plots applied separately in each interval. Though mild, there is evidence for localized model violations.

- All Z_i<2 and p>0.05 for all Q statistics. The model is deemed adequate.

```
> Q.stats(nchs1, xvar=d$ageyears,xcut.points=xc)
                        Z1        Z2       Z3      Z4         N
1.05133 to 2       0.07100  0.825081 -0.18337 -1.8769       277
2 to 3            -0.18707 -0.130950 -0.03975 -0.7080       292
3 to 4             0.50845 -0.490358  0.94195 -1.5862       295
4 to 5            -0.56642 -0.561062 -1.59183  0.4616       299
5 to 6             0.24807 -0.410709 -0.45465  0.7868       268
6 to 7            -0.05671  0.515232  0.20745  0.3652       736
7 to 8            -0.23106 -0.506459 -1.52720  1.2714       775
8 to 9             0.65309 -1.761523 -0.09962 -0.1095       747
9 to 10           -0.47711  2.221167  1.81953 -0.1458       746
10 to 11          -0.55111 -0.409843  0.37449 -0.9341       744
11 to 12           1.40080 -0.461959  1.74313 -1.6224       746
12 to 13          -1.15159  0.520876 -0.52895 -1.0854       862
13 to 15           0.52362  1.086538 -0.48863 -1.6856      1555
15 to 17           0.15358 -1.368450 -0.36813 -0.9125      1480
17 to 19          -0.29058  0.452614  1.02183 -1.3397       765
19 to 21           1.27151 -0.448146  0.93265 -1.0889       227
21 to 23          -1.57283  0.565613 -0.04023 -1.8334       209
23 to 24           0.32900 -0.006126 -0.95314 -1.2027        83
TOTAL Q stats      9.56480 14.411899 16.01596 25.3002     11106
df for Q stats     6.00180 13.999471 13.00066 18.0000         0
p-val for Q stats  0.14433  0.419457  0.24831  0.1169         0
```

4.5 Model Diagnostics, girls

This assessment, too, follows on the one outlined in the previous chapters:

- Formal assessment of the normality of the residuals does not indicate any problems with model fit (figure 4.11):

```
Summary of the Quantile Residuals
                           mean   =  0.0002221
                       variance   =  1
               coef. of skewness  =  0.01472
               coef. of kurtosis  =  2.805
Filliben correlation coefficient  =  0.9992
```

- Worm plot with many de-trended residuals falling outside the 95% confidence interval (figure 4.12). The S-shape suggests kurtosis with heavy tail distributions and inadequate modeling of parameter τ.

- Worm plots on specific age intervals (figure 4.13) show further evidence for localized model violations between 9-10 and 11-13 years.

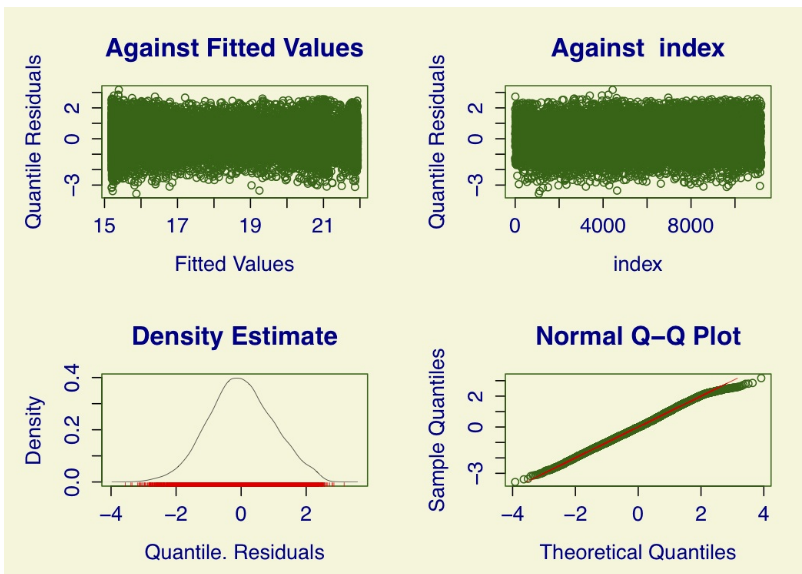


Figure 4.11: Residual plots for assessment of model assumptions (girls). With an adequate fit, residuals should be normally distributed with mean=0, SD=1. A) Quantile residuals vs predicted centiles, B) Index plot of quantile residuals, 3) Frequency histogram of quantile residuals, and D) Q-Q plot of quantile residuals

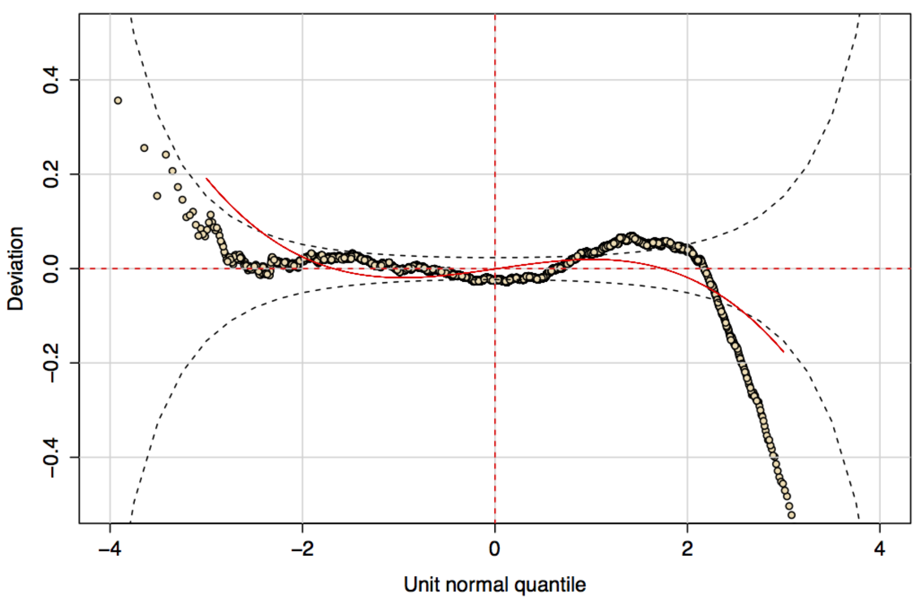


Figure 4.12: Worm plot of model residuals (girls). With an adequate model, detrended residuals should lie between the 2 dashed lines (95% confidence interval). The path of the smoothed curve (solid red) can identify specific model violations.

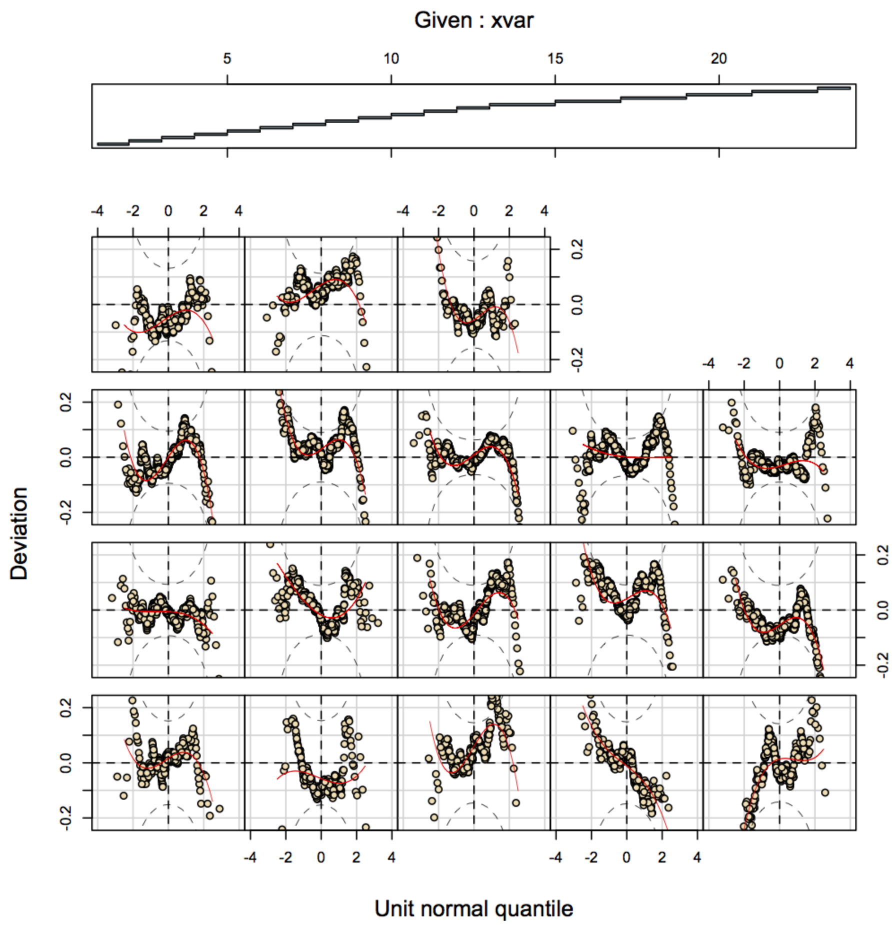


Figure 4.13: The age axis was divided at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, and 23 years, and worm plots applied separately in each interval.

- Z_1,Z_2, and Z_3 are all <2, but there is evidence of mild kurtosis in Z_4 in age intervals 8-9 and 10-15 years with values $\approx$ 2-3. This agrees qualitatively with the impression gleaned from the worm plots above.

```
> Q.stats(nchs1, xvar=d$ageyears,xcut.points=xc)
                        Z1       Z2       Z3         Z4     N
1.04859 to 2       0.1516  0.04113 -0.25879 -0.9807687   263
2 to 3            -0.8642 -0.13669  0.05702  0.8456867   259
3 to 4             0.9175  0.88566 -0.06416 -1.6825399   286
4 to 5            -0.2546 -1.70717 -0.16444 -0.5810743   274
5 to 6            -0.2932  1.30270 -1.26872  1.0312904   301
6 to 7            -0.2928 -0.38978 -0.32725 -0.2742910   699
7 to 8             0.1735 -0.87508  1.72003  0.4588972   761
8 to 9            -0.1503  1.06831  0.68537 -2.0344634   746
9 to 10            1.3481 -0.36734  0.71250 -1.9252913   732
10 to 11          -1.4854 -0.28551  0.16828 -2.4205477   759
11 to 12          -0.2245  0.82069 -0.46922 -3.0513612   682
12 to 13           1.0367 -0.55917  0.70495 -2.6530250   748
13 to 15           0.1718  0.25914 -0.25561 -2.7270067  1528
15 to 17           0.1708 -0.31429  0.44903  0.3046826  1343
17 to 19          -0.7203 -0.01769  0.50849 -0.6345599   744
19 to 21          -1.0202  0.49339 -0.30658 -0.4089101   350
21 to 23           1.2566  0.42955 -0.58965 -0.5915835   475
23 to 23.9972     -0.4045 -0.73098  1.04233 -1.6946305   243
TOTAL Q stats     10.4108  9.80641  8.55313 47.5702630 11193
df for Q stats     6.0022 12.99942 10.99902 18.0000000     0
p-val for Q stats  0.1085  0.70964  0.66298  0.0001744     0
```

- Faced with diagnostic evidence of kurtosis, we introduce the the parameter τ in the 4-parameter BCPE model and again apply the `find.hyper()` function sequentially to minimize GAIC(3). This procedure identified df(μ)=10,df(σ)=7, df(ν)=5, and df(τ)=3 as optimal, and the augmented model was re-fitted. The Q and Z statistics confirm that the kurtosis has been remedied.

```
> Q.stats(nchs2, xvar=d$ageyears,xcut.points=xc)
                         Z1       Z2         Z3      Z4        N
1.04859 to 2        0.09816  0.04631 -0.2561393 -0.4157      263
2 to 3             -0.88187 -0.11353  0.0004577  1.1640      259
3 to 4              0.90938  0.87877 -0.0511373 -1.6083      286
4 to 5             -0.26187 -1.69918 -0.1693116 -0.6538      274
5 to 6             -0.29516  1.28170 -1.2449556  0.9160      301
6 to 7             -0.30244 -0.39927 -0.3477463 -0.3262      699
7 to 8              0.12370 -0.89580  1.7231932  0.7146      761
8 to 9             -0.26241  1.09303  0.6345480 -1.2232      746
9 to 10             1.20550 -0.36153  0.6838789 -0.7172      732
10 to 11           -1.63819 -0.29234 -0.0897526 -0.8303      759
11 to 12           -0.32492  0.83740 -0.7024707 -1.4369      682
12 to 13            0.98492 -0.62218  0.7682578 -1.3567      748
13 to 15            0.07246  0.24771 -0.4552145 -1.1041     1528
15 to 17            0.13518 -0.31144  0.4034719  0.7772     1343
17 to 19           -0.74210 -0.03537  0.5245421 -0.4528      744
19 to 21           -1.04128  0.49203 -0.3700506 -0.1068      350
21 to 23            1.21025  0.48728 -0.6853160  0.2085      475
23 to 23.9972      -0.43582 -0.78592  1.1572313 -0.9401      243
TOTAL Q stats      10.47128 10.03516  9.2901420 15.5698    11193
df for Q stats      6.00220 12.99942 10.9990333 13.0007        0
p-val for Q stats   0.10627  0.69101  0.5950459  0.2732        0
```

Although the BCPE model is significantly better than the BCCG (LMS) model ($\pm$ adjustment for kurtosis), figure 4.14) compares the smoothed centiles from the two models. Clearly, adjustment for kurtosis has a negligible effect except at the extremes i.e. the 99.9^{th} percentile. We concur with the WHO analysts and opt to retain the simpler 3-parameter LMS model for practical applications.

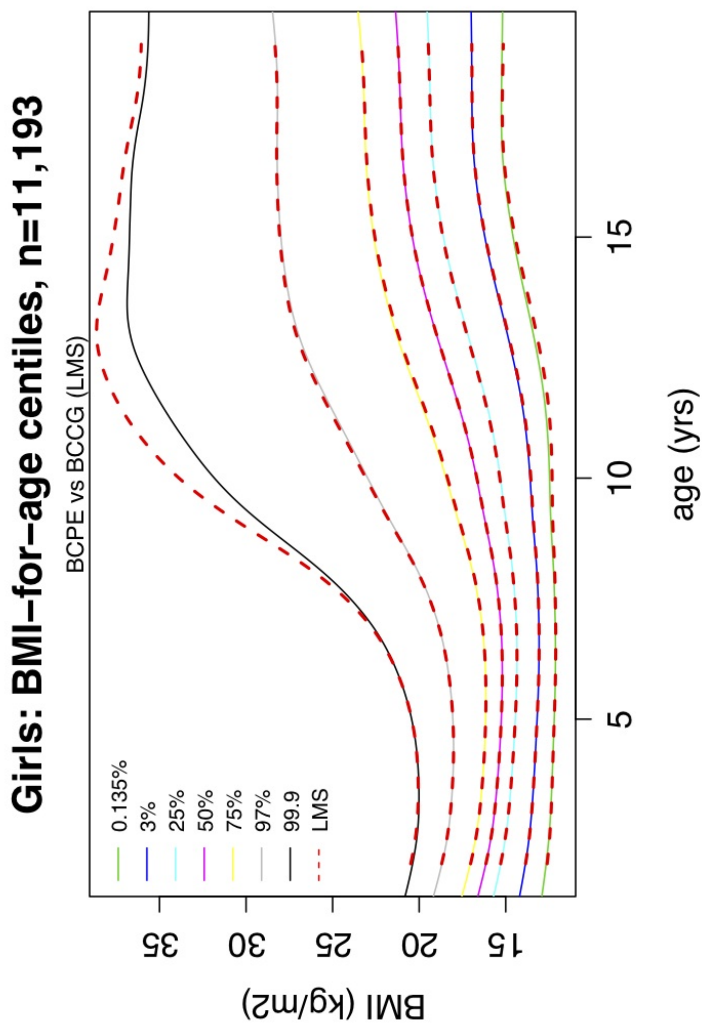


Figure 4.14: Girls smoothed centiles (5-19 years) from the 4-parameter BCPE model with adjustment for kurtosis and the simpler 3-parameter LMS model without adjustment for kurtosis

4.6 Data and methods: short version

The WHO has issued height-for-age, BMI-for-age, and weight-for-age *reference* curves for *school-aged and adolescent children, aged 5-19 years*[1, 3]. Recognizing the importance of BMI in older children, their weight-for-age curves did not extend beyond 10 years of age. Fortunately, the 'core data' for this analysis is publicly available from the National Center for Health Statistics (NCHS), comprising data collected from 1963-

1975 on 22,917 US subjects aged 1-24 years (11507 girls, 11410 boys). In addition, this core data was supplemented by an additional 8306 observations on younger children (ages 18-71 months) from the WHO multicenter growth reference study (MGRS) to smooth the transition between the two datasets, the latter not yet in the public domain.

In preparing the NCHS data for analysis, great care was taken to apply the same exclusion criteria and curve-fitting methods used by the WHO and outlined in their published reports[1, 2, 3], which generated smoothed centiles based on the Box Cox Power Exponential (BCPE) model that explicitly models the time-evolution of 4 parameters i.e. μ (median), σ (coefficient of variation), ν (skew) and τ (kurtosis). For a detailed review of the exclusion process and modelling procedure, please consult the CPEG Statistical Methods and Models manual at the CPEG website http://www.cpeg-gcep.net. A brief summary follows:

As in the WHO reports, there were exclusions for both 'outlying' heights-for-age (14 girls, 8 boys) and 'unhealthy' weights-for-height (300 girls, 304 boys), the latter defined by the WHO as weights-for-height $<0.135^{th}$ or $>97.7^{th}$ percentiles. This compares with a total of 356 girls (3.0%) and 321 boys (2.8%) in the WHO series.

Before the BCPE model can be applied, the time axis may require a power transformation (exponent λ) to better capture periods of rapid change, with the appropriate power transformation identified through minimization of global deviance. Then, the optimal smoothing model must also be specified as degrees of freedom (df) for each model parameter. For each anthropometric measure, smoothing degrees of freedom were identified through sequential minimization of the Generalized Akaike Information Criterion (GAIC), with an adjustable penalty term to balance accurate representation of sample centiles and overall smoothness. All fitted models were subsequently confirmed through appropriate diagnostic studies − including worm plots, Q-statistics, residual plots, and comparisons with sample/published centiles.

- Weight-for-age:
  - For girls, λ =1.225 , df(μ) = 14, df(σ) = 6, df(ν) = 5, and τ=2.
  - For boys, λ =1.3 , df(μ) = 13, df(σ) = 8, df(ν) = 5, and τ=2.
- Height-for-age:
  - For girls, λ = 0.725 , df(μ) = 18, df(σ) = 10, df(ν) = 3, and τ=2.
  - For boys, λ = 1.05 , df(μ) = 14, df(σ) = 9, df(ν) = 3, and τ=2.
- BMI-for-age:
  - For girls, λ = 0.9 , df(μ) = 10, df(σ) = 7, df(ν) = 5, and τ=2.
  - For boys, λ = 0.6 , df(μ) = 10, df(σ) = 5, df(ν) = 3, and τ=2.

Bibliography

[1] M. de Onis, A. W Onyango, E. Borghi, A. Siyam, C. Nishida, and J. Siekmann. Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization*, 85:660–667, 2007.

[2] WHO Multicenter Growth Reference Study Group. WHO child growth standards based on length/height, weight and age. *Acta Paediatrica*, Suppl 450:76–85, 2006.

[3] WHO Multicenter Growth Reference Study Group. *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for- height and body mass index-for-age: Methods and development.* World Health Organization Press, Geneva, 2006.

[4] R.J. Kuczmarski, C.L. Ogden CL, and S.S. Guo *et al*. *2000 CDC Growth Charts for the United States: Methods and Development*, volume 11. National Center for Health Statistics Vital Health Stat, 2002.

[5] Mercedes de Onis, Cutberto Garza, A. W. Onyango, and Elaine Borghi. Comparison of the WHO child growth standards and the CDC 2000 growth charts. *Journal of Nutrition*, pages 144—148, 2007.

[6] D.M. Stasinopoulos, R.A. Rigby, and C. Akantziliotou. *Instructions on how to use the GAMLSS package in R, Second Edition, http://www.gamlss.org/*, 2008.

[7] T.J. Cole and P.J. Green. Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine*, 11:1305–1319, 1992.

[8] R.A. Rigby and D. M. Stasinopoulos. Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. *Statistics in Medicine*, 2004(23):3053—3076.

[9] T. Cole and H. Pan. *LMSChartmaker Pro v. 2.54: A program for calculating age-related reference centiles using the LMS method.* Medical Research Council UK, 2.54 edition, July 2011.

[10] S. van Buuren and M. Fredriks. A simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, 20:1259—1277., 2002.

[11] P. Royston and E.M. Wright. Goodness-of-fit statistics for age-specific reference intervals. *Statistics in Medicine*, 19:2943—2962, 2000.