STATISTICS IN MEDICINE
*Statist. Med.* 2006; **25**:247–265
Published online 5 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2227

# Construction of the World Health Organization child growth standards: selection of methods for attained growth curves

E. Borghi[1], M. de Onis[1,*,†], C. Garza[2], J. Van den Broeck[3], E. A. Frongillo[4], L. Grummer-Strawn[5], S. Van Buuren[6], H. Pan[7], L. Molinari[8], R. Martorell[9], A. W. Onyango[1] and J. C. Martines[10]

*for the WHO Multicentre Growth Reference Study Group*[‡]

[1]*Department of Nutrition, WHO, Geneva, Switzerland*
[2]*Food & Nutrition Program, United Nations University, Ithaca, NY, U.S.A.*
[3]*Africa Centre for Health & Population Studies, KwaZulu-Natal, South Africa*
[4]*Division of Nutritional Sciences, Cornell University, Ithaca, NY, U.S.A.*
[5]*Maternal and Child Nutrition Branch, Centers for Disease Control and Prevention, Atlanta, GA, U.S.A.*
[6]*TNO Prevention and Health, Leiden, The Netherlands*
[7]*Paediatric Epidemiology and Biostatistics, Institute of Child Health, London, U.K.*
[8]*Abteilung für Wachstum und Entwicklung, Kinderspital, Zurich, Switzerland*
[9]*Department of International Health, The Rollins School of Public Health, Emory University, Atlanta, GA, U.S.A.*
[10]*Department of Child and Adolescent Health and Development, WHO, Geneva, Switzerland*

*Correspondence to: Mercedes de Onis, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland.
†E-mail: deonism@who.int

‡Members of the WHO Multicentre Growth Reference Study Group
*Co-ordinating Team* Mercedes de Onis [Study Co-ordinator], Adelheid Onyango, Alain Pinol, Elaine Borghi, Amani Siyam (Department of Nutrition, World Health Organization).
*Executive Committee* Cutberto Garza [Chair], Mercedes de Onis, Jose Martines, Reynaldo Martorell, Cesar G. Victora (up to October 2002), Maharaj K. Bhan (from November 2002).
*Steering Committee*
*Coordinating Centre* (*WHO*, *Geneva*): Mercedes de Onis, Jose Martines, Adelheid Onyango, Alain Pinol.
*Investigators* (*by country*): Cesar G. Victora and Cora Luiza Araújo (Brazil), Anna Lartey and William B. Owusu (Ghana), Maharaj K. Bhan and Nita Bhandari (India), Kaare R. Norum and Gunn-Elin Aa. Bjoerneboe (Norway), Ali Jaffer Mohamed (Oman), Kathryn G. Dewey (U.S.A.).
*Representatives United Nations Agencies*: Cutberto Garza (UNU), Krishna Belbase (UNICEF).
*Advisory Group*
Cameron Chumlea, Tim Cole, Edward Frongillo, Laurence Grummer-Strawn, Reynaldo Martorell, Roger Shrimpton, Jan Van den Broeck.
*Participating countries and investigators*

*Brazil*: Cora Luiza Araújo, Cesar G. Victora, Elaine Albernaz, Elaine Tomasi, Rita de Cássia Fossati da Silveira, Gisele Nader (Departamento de Nutrição and Departamento de Medicina Social, Universidade Federal de Pelotas; and Núcleo de Pediatria and Escola de Psicologia, Universidade Católica de Pelotas).
*Ghana*: Anna Lartey, William B. Owusu, Isabella Sagoe-Moses, Veronica Gomez, Charles Sagoe-Moses (Department of Nutrition and Food Science, University of Ghana; and Ghana Health Service).
*India*: Nita Bhandari, Maharaj K. Bhan, Sunita Taneja, Temsunaro Rongsen, Jyotsna Chetia, Pooja Sharma, Rajiv Bahl (All India Institute of Medical Sciences).
*Norway*: Gunn-Elin Aa. Bjoerneboe, Anne Baerug, Elisabeth Tufte, Kaare R. Norum (Directorate of Health and Social Affairs; National Breastfeeding Centre, Rikshospitalet University Hospital; and Institute for Nutrition Research, University of Oslo).
*Oman*: Ali Jaffer Mohamed, Deena Alasfoor, Nitya S. Prakash, Ruth M. Mabry, Hanadi Jamaan Al Rajab, Sahar Abdou Helmi (Ministry of Health).
*U.S.A.*: Kathryn G. Dewey, Laurie A. Nommsen-Rivers, Roberta J. Cohen, M. Jane Heinig (University of California, Davis).

*Received 19 May 2004*
*Accepted 18 February 2005*

SUMMARY

The World Health Organization (WHO), in collaboration with a number of research institutions worldwide, is developing new child growth standards. As part of a broad consultative process for selecting the best statistical methods, WHO convened a group of statisticians and child growth experts to review available methods, develop a strategy for assessing their strengths and weaknesses, and discuss methodological issues likely to be faced in the process of constructing the new growth curves. To select the method(s) to be used, the group proposed a two-stage decision-making process. First, to select a few relevant methods based on a list of set criteria and, second, to compare the methods using available tests or other established procedures. The group reviewed 30 methods for attained growth curves. Using the pre-defined criteria, a few were selected combining five distributions and two smoothing techniques. Because the number of selected methods was considered too large to be fully tested, a preliminary study was recommended to evaluate goodness of fit of the five distributions. Methods based on distributions with poor performance will be eliminated and the remaining methods fully tested and compared.


KEY WORDS: growth curves; child growth; standards; skewness; kurtosis; smoothing

## 1. INTRODUCTION

Child growth charts are among the most commonly used tools for assessing the general well-being of infants and children and the communities in which they live [1]. They are very useful determining the degree to which physiological needs for growth and development are being met during the fetal and childhood periods. Recognizing the shortcomings of the current National Center for Health Statistics/World Health Organization (NCHS/WHO) international growth reference, WHO in 1994 began planning for new references based on how children *should* grow in all countries rather than merely describing how they grew at a particular time and place [2, 3]. This approach explicitly recognizes that growth references are often used as standards, that is, as tools that enable value judgements.

The WHO Multicentre Growth Reference Study (MGRS) (1997–2003), constituting the second phase of the growth standards project, collected primary growth data and related information from about 8500 children from widely differing ethnic backgrounds and cultural settings (Brazil, Ghana, India, Norway, Oman and U.S.A.). The MGRS aimed to describe the growth of children whose care has followed recommended health practices and behaviours associated with healthy outcomes [4]. The design combines a longitudinal study from birth to 24 months with a cross-sectional study of children aged 18–71 months. In the longitudinal study, mothers and newborns were screened and enrolled at birth and visited at home a total of 21 times on weeks 1, 2, 4 and 6; monthly from 2 to 12 months; and bimonthly in the second year. A detailed description of the study protocol and its implementation can be found elsewhere [5]. Rigorous quality control measures were applied to ensure high quality data [6, 7].

Attained growth references are being constructed for weight-for-age, length/height-for-age, weight-for-length/height, head circumference-for-age, mid-upper arm circumference (MUAC)-for-age, body mass index (BMI)-for-age, triceps skinfold-for-age and subscapular skinfold-for-age. Numerous methods for constructing growth references are described in the literature. Some of these present similar analytical approaches but differ in specific methodological aspects (e.g. distributional assumptions or smoothing techniques). An important step in constructing the new growth curves was to review the strengths and weaknesses of the most relevant methods with respect to the MGRS data set's characteristics (e.g. frequency of measurements and study designs specifics) and to select a subset for further evaluation.

As part of a broad consultative process for selecting the best methods, in January 2003 WHO convened a group of statisticians and child growth experts. This paper summarizes the discussions and recommendations from this statistical advisory group. Section 2 of the paper discusses methodological issues relevant to the construction of the WHO growth curves (including distributional and smoothing aspects, the merging of longitudinal and cross-sectional data, and the handling of the edge effect); Section 3 reviews available statistical methods for constructing growth curves building up on an earlier review [8]; Section 4 defines criteria for selecting the most appropriate methods and, based on these criteria, selects a few methods that merit further consideration; and Section 5 describes diagnostic tools to be used for assessing the goodness of fit.

## 2. METHODOLOGICAL CONSIDERATIONS ON THE CONSTRUCTION OF THE WHO CHILD GROWTH STANDARDS

To construct attained growth curves, the distributional properties of the anthropometric measurements must be studied and centile estimates derived within and across ages. For the most common anthropometric indices knowledge from existing references can be used as a starting point [9–14]. However, original features associated with the longitudinal design of the MGRS, its prescriptive approach, and collection of additional anthropometric measurements are sufficiently novel to raise new challenges [5]. For example, little is known about longitudinal patterns of subscapular and triceps skinfold thicknesses in early childhood.

### *2.1. Distributional aspects*

Methods based on distributional assumptions have been used widely for their ability to produce $z$-scores and estimate extreme centiles more accurately. This requires appropriate agreement between the data's distribution and the selected method's distributional assumptions. Distributions usually are characterized by summary statistics related to three moments: mean, standard deviation (or coefficient of variation (CV)) and skewness. However, the effect of the distribution's fourth moment, i.e. the kurtosis, is increasingly seen as possibly important in estimating extreme centiles.

We illustrate in Figure 1, the case where kurtosis is significantly larger than in the normal distribution (i.e. data distribution presents heavier tails than the normal distribution), using data of subscapular skinfold thickness for boys at 4 months. The histogram in Figure 1(a) shows the fit (line) of the Box–Cox normal distribution [15] to the data. The detrended quantile–quantile ($Q$–$Q$) plot for the fitted $z$-scores in Figure 1(b) indicates how data differ from the assumed underlying distribution. The vertical axis of the detrended $Q$–$Q$ plot represents, for each observation, the difference between empirical and (theoretical distribution) expected $z$-scores. The points form a worm which, when flat, indicates good fit of the assumed distribution [16]. Each shape of worm describes a different aspect of the model fit. For example, the $S$-curve shape is associated with the kurtosis (fourth moment of the distribution). The D'Agostino tests [17] contain a test for the skewness, another for the kurtosis and an omnibus test

 

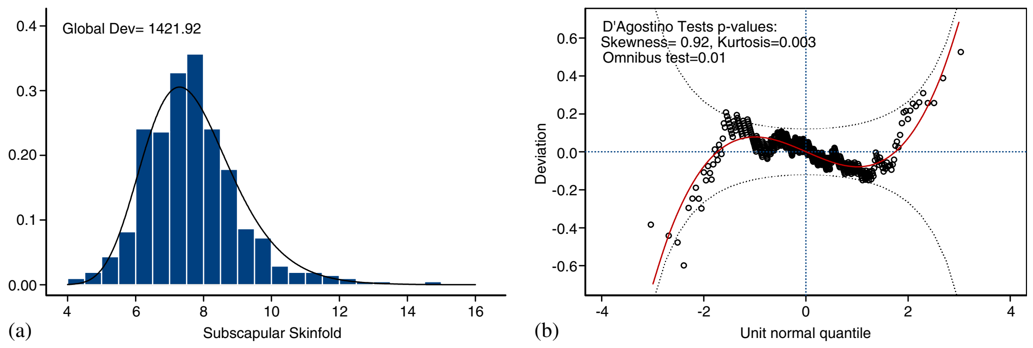


Figure 1. (a) Distribution of subscapular skinfold thickness data for boys at 4 months; and (b) Goodness of fit for the Box–Cox normal distribution (global deviance and D'Agostino tests $p$-values).

combining skewness and kurtosis. In this example, the D'Agostino test for the kurtosis applied to the $z$-scores produced by the Box–Cox normal distribution model indicates residual kurtosis ($p$-value equal to 0.003), agreeing with the diagnostic using the detrended $Q$–$Q$ plot.

The WHO statistical advisory group suggested that the behaviours of the sample median, standard deviation, skewness and kurtosis be examined for all measurements across age. Attention was called to the difficulties of effectively modelling kurtosis through extrapolation given the sparse information usually available at the tails. Thus, the group recommended that methods adjusting for kurtosis be compared with methods that adjust only for the first three moments. A more complex model should be used only if there is significant improvement.

### *2.2. Smoothing aspects*

Selecting the degree of smoothness necessary for adequate centile estimations is not a straightforward task, and it can be a rather subjective undertaking. Lack of smoothing leads to very irregular growth curves, even with large sample sizes. This is due to sampling variability across ages and to the effect of remaining outliers and measurement variability. Smoothing is an attempt to remove part of these effects without disturbing the true underlying growth pattern. The over-smoothing of growth curves flattens biological growth patterns' peaks and valleys, resulting in biased fitted growth curves. The choice of the smoothing function is as important as the choice of the distributional assumptions. While the distribution assumption is age-specific, i.e. it defines the probability with which measurements occur at a specific age, the smoothing function deals with across-age relations.

For the age-smoothing component, flexible functions are needed, especially at ages when growth velocity is faster. In some cases, the use of age transformation prior to smoothing is necessary to facilitate dealing with peaks and valleys at early ages, i.e. close to birth or close to the first measurement record. For example, the normal gain in fat in the first four months of infancy, followed by a drop in per cent body fat, generates a peak in the triceps skinfold thickness pattern between 4 and 6 months. In considering this same anthropometric indicator, adiposity rebound starting at approximately 2 years of age in relatively fatter children is associated with rapid increases in the absolute values of upper centiles, thereby affecting skewness after that age. The smoothing technique combined with the estimated distribution should be able to reflect specific aspects of such measurement pattern. For some curve construction methods, worm plots [16] can be used to help ascertain the appropriate degree of smoothing.

### 2.3. Merging longitudinal and cross-sectional data

The new curves being constructed use longitudinal (birth to 24 months) and cross-sectional data (18–71 months), thus providing overlap between 18 and 24 months. One important advantage of the MGRS is that both longitudinal and cross-sectional samples were taken from the same populations [5]. Nevertheless, differences in the means and variances between the two samples may occur due to sampling errors and their different designs. In the longitudinal study, the measurements are concentrated around pre-specified targeted ages (with a 10 per cent tolerance) and correlated among ages, whereas in the cross-sectional study, observations are independent and evenly spread over the age range.

Figure 2 shows how the MRGS longitudinal and cross-sectional data merge for weight and length/height in terms of empirical quantiles against age. The cross-sectional data were adjusted to the midpoint of the age group using a linear interpolation for each 3-month age interval. To the cross-sectional height data, 0.7 cm were added, the difference between length and height estimated by the mean of the differences between length and height for children whose both measurements were taken during the overlapping period. For these two measures,

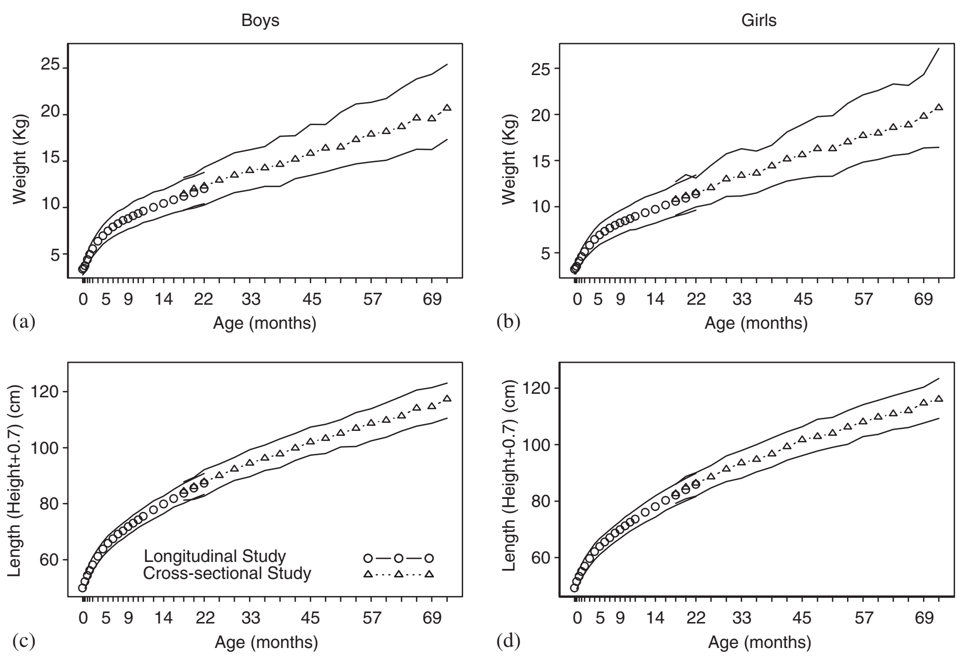


Figure 2. Empirical medians (symbols) and 10th and 90th centiles (lines) for weight for (a) boys and (b) girls and length for the longitudinal component and height +0.7 for the cross-sectional component for (c) boys and (d) girls.

small differences in medians are observed during the overlapping period. For length, differences are negligible for the median, 10th and 90th centile curves. For weight, differences in medians are small, bigger for boys (between 250 and 400g from 20 to 24 months). Differences between the 90th empirical centiles of the longitudinal and cross-sectional samples for weight are also small (between 200 and 550 g 20 to 24 months), approximately 4 per cent of the mean value of the 90th centiles at these age groups, which is around 13 250 g combining boys and girls.

The statistical advisory group suggested that empirical estimates of the longitudinal and cross-sectional samples means and variances be compared for measurements with distribution close to normal, and medians, 10th and 90th centiles be compared for measurements that depart from the normal distribution. If no significant differences are detected, merging of the two data sets should be done simply by pooling the two samples. However, if differences are noted, especially in the overlapping period, it is unlikely that available methods will be able to smooth out artificial pattern deviations without applying a weighting, or other procedure, to correct for the varying amounts of information contributed by each of the two samples.

### *2.4. Handling edge effects*

Edge or boundary effects in the context of growth curves are related to the fact that the precision of the estimation is smaller at the extremes of the age range than near the mean. The size of this effect on the final estimation can be assessed by means of the leverage of the individual points on the curve, which depends on the shape of the curve, the choice of the smoother, the pattern of ages when measurements are made, and the number of measurements at each age. A simulation study was carried out to assess the leverage of observations following a similar design presented by the MGRS, based on the minimum samples sizes requested (400 children by age in the longitudinal component and 400 children by 3-month age group in the cross-sectional). The study indicated that over-sampling at birth (4 times the original sample of 400 children/age) and extending the cross-sectional component from 5 to 6 years would stabilize the leverage function within the target age reference interval from birth to 5 years (document available upon request from the authors).

Thus, although the new growth curves will extend only to 60 months, data were collected until 71 months of age and all the data will be used to construct the curves to minimize the 'right-edge effect'. To minimize the 'left-edge effect' for weight, head circumference and length, the number of data points for the lower limit, i.e. the at-birth sample, was enlarged by using newborn data collected during screening. As a result, the birth sample is 1737. For MUAC and skinfolds, data collection started at 3 months of age and the sample size was not increased at this age to handle the 'left-edge effect'. Nonetheless, the advisory group noted that starting references at later ages, possibly at 6 months, for skinfolds could add edge effects given the peak of the curve around that age. To judge when to start MUAC and skinfolds curves, the group recommended that standard errors of centile estimates be assessed after constructing the curves using all the data.

## 3. METHODS FOR THE CONSTRUCTION OF GROWTH CURVES

In 1997, Wright and Royston provided a comprehensive review of the most commonly used methods for the construction of growth curves [8]. Since then, a number of new methods

 

Table I. Methods for the construction of attained growth curves.

| Method [Reference] | Centiles estimation | Curve-fitting method | Distributional assumptions |
|---|---|---|---|
| *Bin methods, no smoothing* | | | |
| Raw centiles, estimated separately [18] | Separately | None | None |
| *Bin and smooth methods, without distributional assumptions* | | | |
| Fixed knot splines [9] | Separately | Fixed knot splines* | None |
| Eye fitting (weight/age) [19] | Separately | Eye fitting* | None |
| Kernel regression [20, 21] | Separately | Kernel estimation* | None |
| CDC [14] Step 1 | Separately | 3-parameter linear* | None |
| CDC [14] Step 2 | Together | Cubic splines* | Box–Cox normal |
| *Bin and smooth methods, with distributional assumptions* | | | |
| Eye fitting (height/age) [22] | Together | Eye fitting† | Normal |
| Box–Cox normalization [23] | Together | Eye fitting† | Box–Cox normal |
| LMS, version-1 [24] | Together | Cubic splines or others† | Box–Cox normal |
| Variance stabilization (height/age) [25] | Together | Polynomials† | Normal |
| Polynomial fitting [26] | Together | Polynomials† | Normal |
| *Age handled continuously, without distributional assumptions* | | | |
| Quantile regression, estimated separately [27] | Separately | Quantile regression* | None |
| HRY method [28] | Together | Polynomials* | None |
| Adapted HRY method [29] | Together | Grafted polynomials* | None |
| Kernel density estimation [30] | Together | Kernel density estimation‡ | None |
| Non-Gaussian quantile curves [31] | Together | Nearest-neighbour kernel density of conditional cdf‡ | None |
| Non-Gaussian quantile curves [32] | Together | 4-parameter monotonic function (mean) and linear (dispersion)† | None |
| Regression quantiles, estimated together [33] | Together | Natural splines† | None |
| *Age handled continuously, with distributional assumptions* | | | |
| Multilevel models [34] | Together | ML estimation of linear and non-linear models† | Normal |
| Aitkin [35] | Together | Linear models† | Normal |
| Thompson and Theron [36] | Together | Polynomials† | Johnson system |
| LMS, version-2 [15] | Together | Cubic splines† | Box–Cox normal |
| Wade and Ades [37] | Together | Exponential functions† | Box–Cox normal |
| Wade and Ades (with correlations) [38] | Together | ML exponential (spread, skewness); polynomial (mean)† | Box–Cox normal |

 

Table I. *Continued.*

| Method [Reference] | Centiles estimation | Curve-fitting method | Distributional assumptions |
|---|---|---|---|
| FPET method [39] | Together | Fractional polynomials† | (Modulus)-exponential-normal |
| Additivity and variance stabilization (AVAS) [40] | Together | Non-parametric regression AVAS† | Normal |
| Mean and dispersion additive models (MADAM) [41] | Together | Parametric or non-parametric functions (MADAM)† | Normal |
| S-distribution [42] | Together | Polynomials† | S-distribution |
| GAMLSS [43] | Together | Linear parametric or additive non-parametric† | Various |
| LSMT [44] | Together | Cubic splines or (fractional) polynomials† | Box–Cox-*t* |
| LSMP [45] | Together | Cubic splines or (fractional) polynomials† | Box–Cox-power-exponential |

*Applied to centiles.
†applied to distributional parameters.
‡centiles calculated from density fitting.

have been proposed. Table I summarizes 30 existing methods that could potentially be used for the construction of the WHO attained growth curves [9, 14, 15, 18–45]. We have included methods that although are not state-of-art, they are of historical importance, e.g. the method by Roche *et al.* [18] reporting raw centiles, or methods that used ‘eye fitting’ as smoothing technique [19, 22, 23]. The methods have been grouped according to whether they use binning (i.e. age groups) or treat the age continuously, and according to whether they use distributional assumptions or not.

### *3.1. Bin methods (using age grouping) without distributional assumptions*

From the group of methods that estimate centiles separately using age grouping and no distributional assumptions (often called ‘bin and smooth centiles’), the work of Hamill *et al.* [9] can be cited as the first growth curves to be fitted mathematically. These were derived by dividing the data into 1-year age groups, calculating seven centiles from the 5th to the 95th in each group, and then using cubic splines to smooth each centile across age. Knot placements were made the same for all the centiles on each chart. Gasser *et al.* [20] and, more recently, Guo *et al.* [21] applied kernel regression to smooth empirical centiles.

The recent Center for Disease Control and Prevention (CDC) growth curves [14] were constructed using a two-step method. In a first step, a wide variety of methods were tested to model empirical centiles observed in age bins. The method ultimately chosen differed according to the different growth measures considered. For length-for-age, weight-for-age and head circumference-for-age (birth to 36 months), a 3-parameter linear modelling of the empirical centiles was applied [46]. For height-for age, a non-linear model was selected and for weight-for-length and weight-for-height, a fifth degree polynomial was used. For weight-for-age and BMI-for-age in children older than 24 months, smoothing of the empirical centiles first involved robust locally weighted regression [47] followed by polynomial regression. The second step was an 'a *posteriori*' LMS [15] approximation of these smoothed empirical centiles to calculate $z$-scores.

### 3.2. Bin methods with distributional assumptions

From the group of methods that estimate centiles separately using age grouping with distributional assumptions (also called 'bin and smooth distribution parameter to obtain centiles'), the work of Tanner *et al.* [22] assumed that height was normally distributed throughout childhood to construct height curves. Centiles were estimated together and age grouping was employed. They adjusted the variances using the method by Healy [48] to compensate for increased variation due to grouping the data. Eye fitting was used to smooth the centile curves. Chinn [25] suggested a transformation on height to stabilize the standard deviation across age groups. Centiles were calculated using the normal distribution fitted on the back-transformed residuals generated from a cubic polynomial fitting on the transformed height. Niklasson *et al.* [26] also used polynomial fitting for smoothing the mean and standard deviation curves, estimated under normality or power–normality assumption, to update the Swedish references.

To handle age-dependent non-normality and estimating centiles together, Dibley *et al.* [49] describes the method used for constructing normalized growth curves based on the 1977 NCHS references. For weight-for-age and weight-for-height, the centiles were normalized using two separate standard deviations at each age. The first standard deviation was calculated from the 5th, 10th, 25th and 50th empirical centiles, and the second from the 50th, 75th, 90th and 95th empirical centiles. The Box–Cox power transformation [50] was the basis for some proposed methods for constructing growth references, given its capacity to 'normalize' the data to a reasonable degree in terms of skewness. Extending earlier work by Van't Hof *et al.* [23], Cole [24] proposed the first version of the LMS method using age grouping. In Cole's method, the Box–Cox powers $\lambda$ were estimated by maximum likelihood rather than minimum skewness, and the age-specific variation of the distribution was expressed by the CV rather than the standard deviation. The L, M and S functions for the Box–Cox power transformation, mean and CV, respectively, fitted across age, were used for completely characterizing any centile curve and allowing for easy calculation of $z$-scores.

### 3.3. Age handled continuously, without distributional assumptions

Koenker and Basset [27] estimated centiles separately considering age as a continuous variable and introducing the concept of regression quantiles, where the $\alpha$-quantile curve was the linear function. Their method minimized a weighted sum of positive residuals with weight $\alpha$, and negative residuals were made positive and weighted by $(1-\alpha)$. An extension to this work with a cubic spline replacing the linear function, was later proposed [51] using the modification by Efron [52], minimizing the square error loss function. Other regression-quantiles-based methods were proposed in the same vein [31, 32], but estimating centiles together.

For centiles estimated together, there has been recent work in terms of non-parametric methods. The HRY method [28] estimated each centile as a rough curve using the modification of Cleveland's locally weighted scatter plot smoother [47]. Each centile was smoothed using a polynomial function. However, a series of constraints was applied to the coefficients to force commonality on the centiles, preventing the centile curves from crossing. One weakness of the method is the inflexibility of the polynomials. Pan *et al.* [29] overcame this by using grafted polynomials, enabling them to fit centiles to height and weight data for over 9000 Chinese children up to 6 years of age.

Heagerty and Pepe [33] generalized an earlier work by He [53] that proposed a restricted version of regression quantiles that eliminated the problem of quantile (i.e. centile) crossing. The method provided three functions that summarize the complete model. While in He [53] the third function was fixed, Heagerty and Pepe [33] allowed this distribution to vary as a function of covariate(s). The distribution was estimated through local kernel smoothing of the empirical distribution function of the standardized residuals. Earlier work by Rossiter [30] used the kernel approach for density estimation for calculating centile curves.

### *3.4. Age handled continuously, with distributional assumptions*

For centiles estimated together and for which normality is assumed and age is treated as a continuous variable, Aitkin [35] estimated the mean with a linear regression model, and the residual variance about the mean with a log-linear model, using maximum likelihood estimation. In the same group of methods, Altman [54] fitted the spread across age by modelling the absolute residuals about the fitted mean as a function of age. Royston [55] suggested fitting centiles using low-order polynomials for the mean and spread. In more recent work, Royston and Wright [39] proposed a more flexible class of functions composed of low-order fractional polynomials for each of the parameters of a modulus-exponential-normal (MEN) distribution.

Rigby and Stasinopoulos [56] presented a flexible model for variance heterogeneity that assumed normal errors. Both mean and variance were modelled using semi-parametric models, called 'mean and dispersion additive model' (MADAM). More recently, they have proposed the use of a subclass of MADAM for modelling the mean and variance (or log-variance), including parametric or smooth non-parametric functions of age for constructing age-related centiles [41]. Tango [40] proposed a non-parametric regression based on additivity and variance stabilization (AVAS) procedure, but ultimately using the normal assumption for constructing the centiles.

When departure from the normal distribution occurred, the use of more flexible families of distributions for constructing growth curves has been investigated. Cole and Green [15] extended Cole's LMS method [24] which adjusts for skewness. They improved the previous method which applied age grouping to allow the data to be analysed continuously on age with the L, M and S curves fitted by maximum penalized likelihood, applying cubic splines smoothing. Wade and Ades [37] developed a similar method to fit centiles with age trends in the Box–Cox power, mean and SD functions being specified by exponential parametric functions, rather than cubic splines. Sorribas *et al.* [42] proposed a parametric method based on the S-distribution that covers a wide range of shapes and types of skewness, to construct smoothed centile curves across ages.

The problem of significant kurtosis in the residuals from the LMS method has been highlighted [8, 12, 16]. Thompson and Theron [36] produced centiles based on the Johnson family of 4-parameter distributions that specify location, scale and two shape parameters [57]. They modelled the location and scale parameters using polynomials in age and the shape parameters as constants and fitted the model by maximum likelihood [36]. Royston and Wright [39] also proposed a completely parametric method, namely the fractional polynomials and exponential transformation (FPET) method. Fractional polynomials of age [58] are used to model the parameters of a chosen 3- or 4-parameter distribution estimated by maximum likelihood. They consider the distributions 'exponential-normal' (EN) and 'modulus-exponential-normal' (MEN), based on transformation of the data towards normality.

Rigby and Stasinopoulos [43] proposed a class of univariate statistical models—generalized additive models for location, scale and shape (GAMLSS)—that could be applied to data presenting skewed and/or kurtotic continuous or discrete distributions. The systematic part of the model was expanded to allow modelling the location (e.g. median) and other parameters of the distribution of the response variate as linear parametric and/or additive non-parametric functions of explanatory variables. Maximum (penalized) likelihood estimation was used to fit the models. It was suggested that, for each fitted GAMLSS model, the randomized quantile residuals [59] should be used to check model's adequacy and especially the response variable's distribution. These residuals are $z$-scores, which always have a standard normal distribution if the model is correct.

Within the GAMLSS framework, Rigby and Stasinopoulos [44] proposed the LMST method of centiles estimation for a response variable exhibiting both skewness and kurtosis larger than 3 (i.e. leptokurtic data), which is based on the Box–Cox $t$ (BCT) distribution. This model assumes that a transformed response variable (using the same transformation used by Cole and Green [15]) has a Student $t$ distribution with degrees of freedom parameter $\tau > 0$ (with $\tau$ treated as a continuous parameter). The Box–Cox $t$ distribution has 4 parameters, denoted by $\mu$, $\sigma$, $\nu$, $\tau$, related to location (median), scale (CV), skewness (power transformation to symmetry), and kurtosis ($t$ distribution degrees of freedom), respectively. The first three parameters $\mu$, $\sigma$, $\nu$ are those used in the LMS method. When $\tau$ tends to infinity, the Box–Cox $t$ distribution converges to the Box–Cox normal distribution used in the LMS method.

To overcome the limitation of the Box–Cox $t$ distribution caused by the fact it handles only leptokurtic data, Rigby and Stasinopoulos [45] developed a more flexible distribution, the Box–Cox-power-exponential (BCPE) distribution. This distribution is able to model any type of kurtosis (lepto, platy or mesokurtosis). The Box–Cox normal distribution is a particular case of the BCPE distribution for the case the fourth parameter $\tau$ is equal to 2 (i.e. mesokurtic case). They called this generalization of the LMS method the LMSP method.

Both the LMST and LMSP methods provide highly flexible models for location, scale, skewness and kurtosis of the response variate using the GAMLSS [43]. Each of the 4 parameters can be modelled using parametric (e.g. fractional polynomial) or non-parametric (e.g. cubic spline) functions of the explanatory variable, e.g. age. The model fitting is achieved by maximum (penalized) likelihood. A simple formula for computing centiles and $z$-scores can be provided. Additional covariates or factors can also be included in the model, if required, allowing adjustments to the parameters, especially $\mu$, for other explanatory variables (e.g. sex, parental height, etc).

### *3.5. Methods incorporating correlations*

There has been little development towards the fitting of growth curves with correlated measurements (i.e. repeated measurements), as most existing methods are based on cross-sectional designs. Mixed-effect models can be used for modelling growth curves, since they allow the partitioning of the variance to the contribution of between- and within-subject variability. Laird and Ware [60] wrote about two-stage random-effect models for growth and the empirical Bayes estimates of parameters in a specified model. This method assumes that data are normally distributed and that inference based on non-normal data should be done only after a normalizing transformation. Milani *et al.* [61] applied this methodology to the construction of longitudinal growth norms for length. They used measurements taken at birth and at 5–8 follow-ups between 3 months and 3 years of life in Italian children. In a more general framework, Goldstein [34] proposed the use of multilevel models to resolve longitudinal problems, including those presented by growth studies. This class of mixed-effect models or random-effect models allows for complex structures of covariance and also for explanatory variables that depend on age. Multilevel modelling with non-parametric components also is found in References [62–64].

Recently, some authors expressed interest in investigating the effects of a correlation structure on the modelling of growth curves [38, 63, 64]. Wade and Ades [38] proposed an LMS-based maximum likelihood method for fitting age-related growth curves that incorporates explicit modelling of within-subject correlations. Applied to CD4 counts of uninfected children born to HIV-1-infected women, correlation structures of varying degrees of complexity were considered; exponential models were used to fit spread and skewness trends, and high-degree polynomials and exponential models to fit the median trend. The data consist of routinely collected serial measurements where the number and timing of repeats are assumed to be unrelated to the actual values of the measurements. Five different correlation structures combined with five models for the median were compared and the method of 'profile likelihood' [65] was used for the construction of confidence intervals with an iterative algorithm. Despite the presence of a strong correlation structure, the results showed that incorporating a correlation structure into the likelihood function had little effect on the choice of the model for the mean curve or on the fitted centiles. Moreover, the precision with which the centiles was estimated were not affected to any clinically important degree when the correlations were incorporated.

In the case of the longitudinal component of MGRS data set, timing of repeats is fixed and number of measurements taken per child is constant, apart from few sporadically missing visits. As a result, we do not expect bias in the estimate of the mean curve. A special effort is needed, however, to accurately estimate the confidence intervals around the centile estimates. The choice of using methods that apply to cross-sectional data on the longitudinal study component can be combined with the use of the bootstrap technique for deriving the standard errors associated with centile estimates. To accomplish this, resampling of children (i.e. vectors of measurements) will be done with replacement to create the new samples of the same size in the bootstrap procedure, which will lead us to correctly estimating the standard errors of the centile estimates.

## 4. CRITERIA AND SELECTION OF METHODS

### *4.1. Criteria for methods selection*

The selection of methods for constructing the attained growth curves should be based on the criteria that account for the MGRS data set's characteristics and the uses of the future standards. The advisory group agreed on the following points to guide the method selection:

- Outer centiles cannot be estimated with sufficient precision without relying on an underlying distribution, given the scarce information at the tails.
- The crossing of centiles must be avoided. Therefore, the simultaneous calculation of centiles is necessary.
- Only methods that allow back-transformations should be considered so that direct calculation of centiles and $z$-scores is enabled.
- Because the MGRS data set is composed of both a longitudinal component, sampled at target ages, and a cross-sectional component (where data were collected continuously across the age range), its design does not favour the use of age-grouping methods for estimating centiles. Interpolation to an age-group midpoint, followed by variability correction, could be applied to cross-sectional data, but treating age as a continuous variable likely will avoid problems associated with variability and arbitrary choices of age groupings.
- Methods that are able to address kurtosis, in addition to skewness, are preferable; genuine kurtosis can occur for some measurements and, if data are not modelled correctly, resulting fitted centiles can be distorted.

The advisory group also discussed secondary criteria to be used in method selection. For example, the method should provide easy assessment to goodness-of-fit diagnostic tools. Computational simplicity should be considered in selecting between two methods that meet all primary requirements equally well. Similarly, careful consideration should be given to any single method that is flexible enough to be applied across all measurements.

In summary, the primary criteria agreed on for method selection were the ability to:

- estimate precisely outer centiles,
- estimate centiles simultaneously in such a way that they are constrained not to cross,
- estimate $z$-scores and centiles using direct formulae,
- apply continuous age smoothing, and
- account for both skewness and kurtosis when necessary.

The secondary criteria were:

- ability to assess fit to the data,
- easy to explain and well documented,
- useful for application to different anthropometric measures so that the WHO growth curves would rely on a single approach.

### *4.2. Method selection*

The advisory group proceeded to review the 30 methods listed in Table I based on these primary and secondary criteria. Only methods that treated age as a continuous variable, were based on explicit distributional assumptions and simultaneously estimated centiles were considered.

The final short list of methods recommended for constructing the WHO growth curves was

- the FTPE by Royston and Wright [39] with the MEN distribution and fractional polynomial fitting technique,

 

- the GAMLSS method by Rigby and Stasinopoulos [43] with the Box–Cox $t$ distribution [44] or BCPE [45] distributions with a natural cubic spline-fitting technique (it also includes the fractional polynomial),
- a method to be proposed using the Johnson's family of distributions [57].

The LMS method [15], which uses the Box–Cox normal distribution, is a particular case of the GAMLSS method, for which kurtosis is not modelled and a cubic splines technique is used for fitting. Therefore, it was naturally retained for testing, which was considered important, since it has been used successfully in many cases.

Given the many selected methods, the group proposed a preliminary study to decrease the number of methods to be tested fully. As a first step, the group recommended that MGRS data be used to evaluate various distributions relevant to the selected methods by comparing the goodness of fit of five distributions

- the Box–Cox normal (3 parameters),
- the Box–Cox $t$ (4 parameters),
- the BCPE (4 parameters),
- the MEN (4 parameters),
- the Johnson system (4 parameters).

Initially, the goodness of fit of each of the five distributions will be assessed separately for the data in each separate age group (in the longitudinal study) for each growth variable. The rationale for this first step was that if one family of distributions initially dominates across separate age groups, the same distribution is likely to dominate when age smoothing is introduced.

## 5. ASSESSING GOODNESS OF FIT

For testing and comparing the final methods, many diagnostic tools and tests have been proposed in the literature. Some were cited by Van Buuren and Fredriks [16] for checking the quality of the fit

(i) Visual inspection of the shape of the centile curves,
(ii) Centiles plotted with data points,
(iii) Empirical and fitted centiles plotted on top of each other,
(iv) Observed and expected counts of children with measurements below fitted centiles,
(v) Tests of normality for the $z$-scores,
(vi) Detrended normal $Q$–$Q$ plot of the $z$-scores,
(vii) Worm plots.

Royston and Wright [39] proposed, among other tools, the use of model augmentation for testing goodness of fit. The reduction in deviance that results from including $k-1$ age-group indicators in a model that includes a constant and the parametric terms is tested against $\chi^2$ with $k-1$ degrees of freedom. If the test is significant, there is evidence of model inadequacy. The authors pointed out that, as this test does not take into account time ordering, it may lack power. Also, the number of age groups is chosen subjectively and different categorizations may lead to different conclusions about a given model.

 

Another possible test to check a model's goodness of fit is the $Q$-statistic, described by Royston and Wright [66]. The distribution of fitted $z$-scores is tested for normality across age using a combination of tests for the four moments of the distribution, the modified D'Agostino tests, and Shapiro–Wilk test. This provides guidance to which moments of the distribution are inadequately modelled. The $Q$-statistic from a particular age range indicates whether the corresponding moment is inadequately model in that age range. The $Q$-tests together with the 'worm plots' should provide trustworthy diagnostic.

Most of these goodness-of-fit techniques apply to the final growth curve, after smoothing across age, as for example, the worm plots, the $Q$-test, and graphical tools (i)–(iii) in the list above. Some, however, can be applied at age-specific level, like D'Agostino normality tests on the $z$-scores, detrended $Q$–$Q$ plots, and observed against expected counts comparisons.

For the preliminary study of the distributions, the group recommended that the goodness of fit of the selected distributions for each age group be examined using the following diagnostic tools:

(a) Comparison of the log-likelihood or global deviance equals $-2$ times the log-likelihood;
(b) detrended normal $Q$–$Q$ plots of the fitted $z$-scores, i.e. plotting empirical quantiles minus the normal quantiles against the normal quantiles;
(c) comparison of distribution model and empirical sample per cents;
(d) normality tests on the fitted $z$-scores, e.g. D'Agostino tests [17].

Figure 3 illustrates the comparison among the five distributions for the subscapular skinfold thickness for boys at 4 months. This example was carried out using the software GAMLSS [67], kindly provided by the authors. For these specific data, kurtosis is accounted for and $z$-scores are corrected when using any of the 4-parameter distributions. We notice a decrease of around 10 in global deviance, associated with an increase of 1 parameter in the model, comparing the fit using the BCPE distribution with the fit using the Box–Cox-power-exponential distribution to the data, which is highly significant using the generalized likelihood ratio test for nested models. We also observe the visual change in the detrended normal $Q$–$Q$ plot, which show the considerable flattening of the worm formed by the points. The D'Agostino tests of normality, as previously described in Figure 1, detects the residual kurtosis when applying the Box–Cox normal distribution. Complete comparisons for all age groups in the longitudinal study for all measurements are to be reported separately.

Nevertheless, simply comparing the goodness of fit at specific age groups may lead to different results from those obtained if smoothing across age is applied. In a further step, the method(s) using the distribution(s) with better performance across age groups should be fully tested, i.e. including the age-smoothing component.

## 6. CONCLUSIONS

The WHO statistical advisory group reviewed 30 methods for attained growth curves. Using primary and secondary pre-set criteria, the advisory group selected three methods for testing, combining five distributions and two smoothing techniques. Because the number of resulting methods was considered too large to be fully applied and compared, a preliminary study was recommended to evaluate distributional goodness of fit of the five distributions. The results of this evaluation will be reported separately. Methods based on distributions with

 

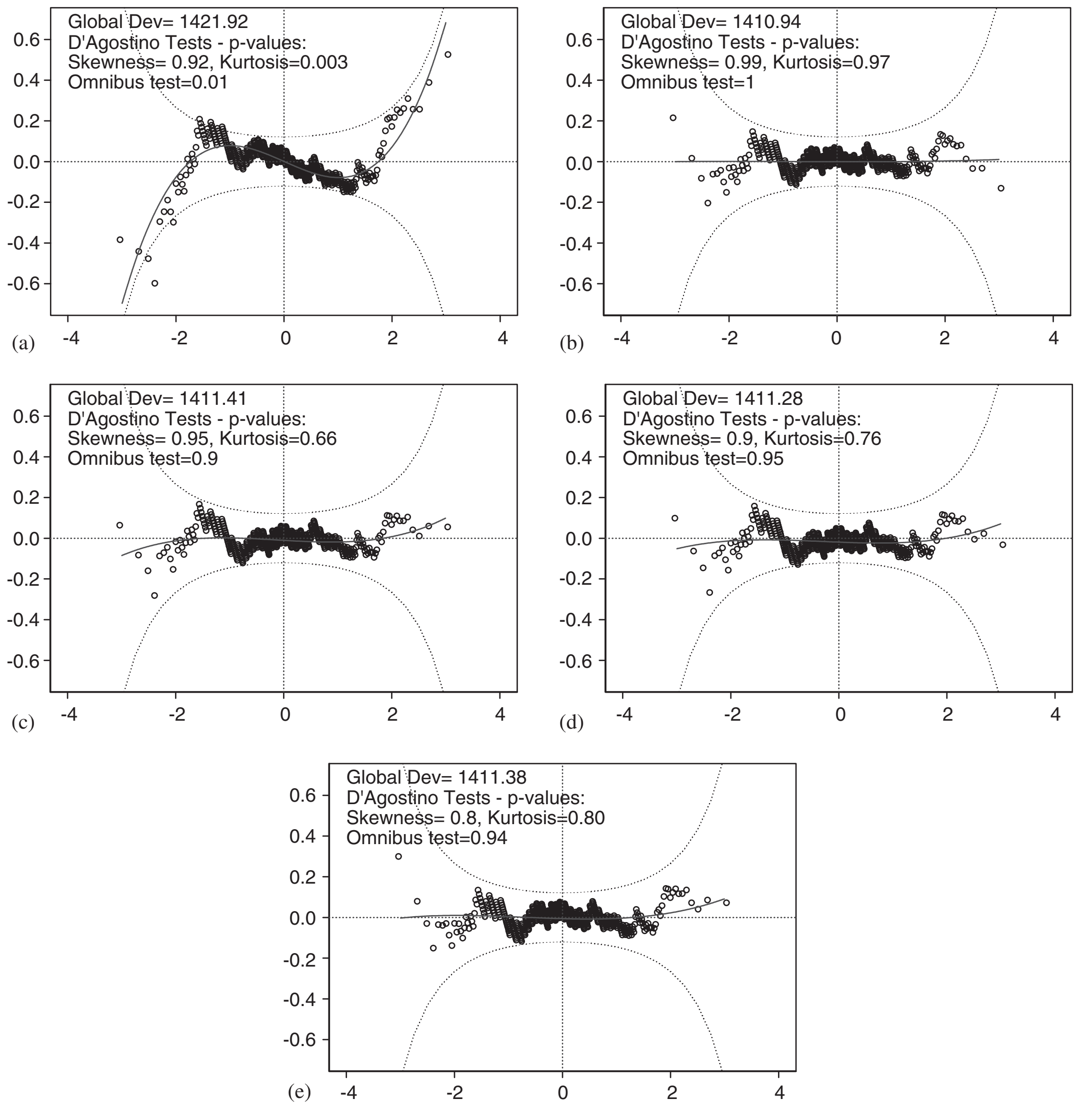


Figure 3. Goodness-of-fit comparisons for subscapular skinfold thickness for boys at 4 months fitting distributions: (a) Box–Cox normal; (b) Box–Cox T; (c) Box–Cox-power-exponential; (d) Modulus-exponential-normal; and (e) Johnson SU.

poor performance will be eliminated and the remaining methods fully tested and compared. The WHO statistical advisory group also stressed the importance of evaluating and field testing the growth standards before releasing them for general use.

 

ACKNOWLEDGEMENTS

We are grateful to Professor Tim Cole from the Institute of Child Health, London, U.K., for conducting the first review of methods in 1996 and providing comments to the subsequent updates.

REFERENCES

1. de Onis M, Wijnhoven TMA, Onyango A. Worldwide practices in child growth monitoring. *The Journal of Pediatrics* 2004; **144**:461–465.
2. WHO. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. *Technical Report Series No. 854*, World Health Organization, Geneva, 1995.
3. de Onis M, Garza C, Habicht JP. Time for a new growth reference. *Pediatrics* 1997; **100**(5):E8.
4. Garza C, de Onis M, for the WHO Multicentre Growth Reference Study Group. Rationale for developing a new international growth reference. *Food and Nutrition Bulletin* 2004; **25**(Suppl. 1):S5–S14.
5. de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: planning, study design and methodology. *Food and Nutrition Bulletin* 2004; **25**(Suppl. 1):S15–S26.
6. Onyango AW, Pinol AJ, de Onis M, for the WHO Multicentre Growth Reference Study Group. Managing data for a multi-country longitudinal study: experience from the WHO Multicentre Growth Reference Study. *Food and Nutrition Bulletin* 2004; **25**(Suppl. 1):S46–S52.
7. de Onis M, Onyango AW, Van den Broeck J, Cameron WC, Martorell R, for the WHO Multicentre Growth Reference Study Group. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. *Food and Nutrition Bulletin* 2004; **25**(Suppl. 1):S27–S36.
8. Wright EM, Royston P. A comparison of statistical methods for age-related reference intervals. *Journal of the Royal Statistical Society*, *Series A* 1997; **160**:47–69.
9. Hamill PVV, Drizd TA, Johnson CL, Reed RB, Roche AF. NCHS growth curves for children birth—18 years. *Vital and Health Statistics Series 11, No. 165*, National Center for Health Statistics, Washington, DC, 1977.
10. Prader A, Largo RH, Molinari L, Issler C. Physical growth of Swiss children from birth to 20 years of age. *Helvetica Paediatrica Acta* 1989; **52**(Suppl.):1–125.
11. Rolland-Cachera MF, Cole TJ, Sempé M, Tichet J, Rossignol C, Charraud A. Body mass index variations: centiles from birth to 87 years. *European Journal of Clinical Nutrition* 1991; **45**:13–21.
12. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine* 1998; **17**:407–429.
13. Fredriks AM, Van Buuren S, Burgmeijer RJF, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in the Netherlands 1955–1997. *Pediatric Research* 2000; **47**:316–323.
14. Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, Flegal KM, Guo SS, Wei R, Mei Z, Curtin LR, Roche AF, Johnson CL. CDC growth charts: United States. *Advance Data from Vital and Health Statistics*, *No. 314*, NCHS, Hyattsville, Maryland, 4 December 2000.
15. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine* 1992; **11**:1305–1319.
16. Van Buuren S, Fredriks M. Worm plot. A simple diagnostic device for modeling growth reference curves. *Statistics in Medicine* 2001; **20**:1259–1277.
17. D'Agostino RB, Belanger A, D'Agostino Jr RB. A suggestion for using powerful and informative tests of normality. *The American Statistician* 1990; **44**:316–321.
18. Roche AF, Guo S, Baumgartner RN, Chumlea WC, Ryan AS, Kuczmarski RJ. Reference data for weight, stature, and weight/stature2 in Mexican Americans from the Hispanic Health and Nutrition Survey (HHANES 1982–1984) *The American Journal of Clinical Nutrition* 1990; **51**:917S–924S.
19. Roede MJ, van Wieringen JC. Growth diagrams 1980. Netherlands' third nation-wide survey. *Tijdsschrift voor Sociale Gezondheidszorg* 1985; **63**:1–33.
20. Gasser T, Köhler W, Müller H-G, Kneip A, Largo R, Molinari L, Prader A. Velocity and acceleration of height growth using kernel estimation. *Annals of Human Biology* 1984; **11**:397–411.
21. Guo S, Roche AF, Baumgartner RN, Cameron Chumlea WM, Ryan A. Kernel regression for smoothing percentile curves: reference data for calf and subscapular skinfold thicknesses in Mexican Americans. *The American Journal of Clinical Nutrition* 1990; **51**:908S–916S.
22. Tanner JM, Whitehouse RH, Takaishi M. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965; Part I. *Archives of Disease in Childhood* 1966; **41**:454–471.
23. Van 't Hof MA, Wit JM, Roede MJ. A method to construct age references for skewed skinfold data, using Box–Cox transformations to normality. *Human Biology* 1985; **57**:131–139.
24. Cole TJ. Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society*, *Series A* 1988; **151**:385–418.

 

25. Chinn S. A new method for calculation of height centiles for preadolescent children. *Annals of Human Biology* 1992; **19**:221–232.
26. Niklasson A, Ericson A, Fryer JG, Karlberg J, Lawrence C, Karlberg P. An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977–1981). *Acta Paediatrica Scandinavica* 1991; **80**:756–762.
27. Koenker RW, Bassett GW. Regression quantiles. *Econometrica* 1978; **46**:33–50.
28. Healy MJR, Rasbash J, Yang M. Distribution-free estimation of age-related centiles. *Annals of Human Biology* 1988; **15**:17–22.
29. Pan H, Goldstein H, Yang Q. Non-parametric estimation of age-related centiles over wide age ranges. *Annals of Human Biology* 1990; **17**:475–481.
30. Rossiter JE. Calculating centile curves using kernel density estimation methods with application to infant kidney lengths. *Statistics in Medicine* 1991; **10**:1693–1701.
31. Ducharme GR, Gannoun A, Guertin MC, Jéquier JC. Reference values obtained by kernel-based estimation of quantile regressions. *Biometrics* 1995; **51**:1105–1116.
32. Wellek S, Merz E. Age-related reference ranges for growth parameters. *Methods of Information in Medicine* 1995; **34**:523–528.
33. Heagerty PJ, Pepe MS. Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Applied Statistics* 1999; **48**:533–551.
34. Goldstein H. Efficient statistical modelling of longitudinal data. *Annals of Human Biology* 1986; **13**:129–141.
35. Aitkin M. Modelling variance heterogeneity in normal regression using GLIM. *Applied Statistics* 1987; **36**: 332–339.
36. Thompson ML, Theron GB. Maximum likelihood estimation of reference centiles. *Statistics in Medicine* 1990; **9**:539–548.
37. Wade AM, Ades AE. Age-related reference ranges—significance tests for models and confidence intervals for centiles. *Statistics in Medicine* 1994; **13**:2359–2367.
38. Wade AM, Ades AE. Incorporating correlations between measurements into the estimation of age-related reference ranges. *Statistics in Medicine* 1998; **17**:1989–2002.
39. Royston P, Wright EM. A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. *Journal of the Royal Statistical Society*, *Series A* 1998; **161**:79–101.
40. Tango T. Estimation of age-specific reference ranges via smoother AVAS. *Statistics in Medicine* 1998; **17**: 1231–1243.
41. Rigby RA, Stasinopoulos DM. Construction of reference centiles using mean and dispersion additive models. *The Statistician* 2000; **49**:41–50.
42. Sorribas A, March J, Voit EO. Estimating age-related trends in cross-sectional studies using S-distributions. *Statistics in Medicine* 2000; **19**:697–713.
43. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *Applied Statistics* 2005, in press.
44. Rigby RA, Stasinopoulos DM. Box–Cox *t* distribution for modeling skew and leptokurtotic data. *Technical Report 01/04*, STORM Research Centre, London Metropolitan University, London, 2004.
45. Rigby RA, Stasinopoulos DM. Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. *Statistics in Medicine* 2004; **23**:3053–3076.
46. Guo S, Siervogel RM, Roche AF, Cameron Chumlea WM. Mathematical modeling of human growth: a comparative study. *American Journal of Human Biology* 1992; **4**:93–104.
47. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 1979; **74**:829–836.
48. Healy MJR. The effect of age-grouping on the distribution of a measurement affected by growth. *American Journal of Physical Anthropology* 1962; **20**:49.
49. Dibley MJ, Goldsby J, Staehling N, Trowbridge FL. Development of normalized curves for the international growth reference: historical and technical considerations. *American Journal of Clinical Nutrition* 1987; **46**: 736–748.
50. Box GEP, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society*, *Series B* 1964; **26**:211–252.
51. Koenker R, Ng P, Portnoy S. Quantile smoothing splines. *Biometrika* 1994; **81**:673–680.
52. Efron B. Regression percentiles using asymmetric squared error loss. *Statistica Sinica* 1991; **1**:93–125.
53. He X. Quantile curves without crossing. *The American Statistician* 1997; **51**:186–192.
54. Altman DG. Construction of age-related reference centiles using absolute residuals. *Statistics in Medicine* 1993; **12**:917–924.
55. Royston P. Constructing time-specific reference ranges. *Statistics in Medicine* 1991; **10**:675–690.
56. Rigby RA, Stasinopoulos DM. A semi-parametric additive model for variance heterogeneity. *Statistics and Computing* 1996; **6**:57–65.


 

57. Johnson NL. Systems of frequency curves generated by methods of translation. *Biometrika* 1949; **36**:149–176.
58. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). *Applied Statistics* 1994; **43**:429–467.
59. Dunn PK, Smyth GK. Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science* 1996; **11**:89–121.
60. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.
61. Milani S, Bossi A, Marubini E. Individual growth curves and longitudinal growth charts between 0 and 3 years. *Acta Paediatrica Scandinavica* 1989; **350**(Suppl.):95–104.
62. Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society*, *Series B* 1991; **53**:233–243.
63. Zeger SL, Diggle PJ. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* 1994; **50**:689–699.
64. Donnelly CA, Laird NM, Ware JH. Prediction and creation of smooth curves for temporally correlated longitudinal data. *Journal of the American Statistical Association* 1995; **90**:984–989.
65. Mathews DE. Likelihood-based confidence intervals for functions of many parameters. *Biometrika* 1988; **75**: 139–144.
66. Royston P, Wright EM. Goodness-of-fit statistics for age-specific reference intervals. *Statistics in Medicine* 2000; **19**:2943–2962.
67. Stasinopoulos DM, Rigby RA, Akantziliotou C. GAMLSS package in R, version 0.3-5 2004. (available from the authors).

 