STATISTICS IN MEDICINE, VOL. 17, 407–429 (1998)
Statist. Med., **17**, 407–429 (1998)

BRITISH 1990 GROWTH REFERENCE CENTILES FOR WEIGHT, HEIGHT, BODY MASS INDEX AND HEAD CIRCUMFERENCE FITTED BY MAXIMUM PENALIZED LIKELIHOOD

T. J. COLE[1]*, J. V. FREEMAN[2] AND M. A. PREECE[2]

[1] *MRC Dunn Nutrition Centre, Cambridge, U.K.*
[2] *Institute of Child Health, London, U.K.*

SUMMARY

To update the British growth reference, anthropometric data for weight, height, body mass index (weight/height^2) and head circumference from 17 distinct surveys representative of England, Scotland and Wales (37,700 children, age range 23 weeks gestation to 23 years) were analysed by maximum penalized likelihood using the LMS method. This estimates the measurement centiles in terms of three age-sex-specific cubic spline curves: the *L* curve (Box–Cox power to remove skewness), *M* curve (median) and *S* curve (coefficient of variation). A two-stage fitting procedure was developed to model the age trends in median weight and height, and simulation was used to estimate confidence intervals for the fitted centiles. The reference converts measurements to standard deviation scores (SDS) that are very close to Normally distributed – the means, medians and skewness for the four measurements are effectively zero overall, with standard deviations very close to one and only slight evidence of positive kurtosis beyond ± 2 SDS. The ability to express anthropometry as SDS greatly simplifies growth assessment. © 1998 John Wiley & Sons, Ltd.

INTRODUCTION

Age-related reference ranges have received considerable attention in recent years.[1–11] There are two main schools of thought about how they should be fitted: distribution-free (for example, quantile regression), and parametric (for example, Box–Cox transformation). The parametric approach offers further choice over the form of the age-related summary curves defining the distribution: functional (for example, exponential[7] or fractional polynomial[12]), or based on smoothing splines.[6]

The advantages of a functional form are that asymptotic behaviour can be modelled explicitly, and the summary curves are parsimoniously represented, whereas a smoothing spline requires a table of values on an age grid of suitable resolution. A possible disadvantage of the functional

* Correspondence to: T. J. Cole Ph.D., MRC Dunn Nutrition Centre, Downhams Lane, Milton Road, Cambridge CB4 1XJ, U.K. E-mail: Tim.Cole@mrc-dunn.cam.ac.uk

CCC 0277–6715/98/040407–23$17.50

Received November 1995
Revised May 1997

approach compared with splines is that the function may be insufficiently flexible to model the distribution at all ages, leading to bias.

As well as goodness-of-fit, confidence intervals for the fitted centiles are an important requirement. Wade and Ades[7] obtained them by generating contour likelihoods around each fitted curve at selected points. Ideally it would be better to obtain confidence intervals for each centile curve over the whole age range, rather than interpolating from particular ages.

This paper describes the extension of the spline smoothing approach of Cole and Green[6] to generate British national reference centile curves for weight, height, body mass index (weight/height^2) and head circumference between 23 weeks of gestation and 23 years. The method deals with asymptotic behaviour near adulthood, goodness-of-fit, and the estimation of confidence intervals. Clinically oriented papers describing some of the charts have already been published.[13,14]

DATA SOURCES

Paediatricians have made considerable use of the weight and height reference centiles first published 30 years ago in the 'citation classic' by Tanner *et al.*[15] Over the last decade these references have become more and more out-of-date, due to the secular trend in body size. In 1991 a search was made for pre-existing data sets of child anthropometry which might be combined to provide a source for new reference centiles. Subsequently extra data sets were added to extend the age range of the references down to 23 weeks gestation. Ideally the data sets were required to be recent, cross-sectional, representative of Britain and of high quality. In practice most but not all of these aims were met.

The chosen data came from twelve sources, which can best be viewed in three groups. The first six were described in detail by Freeman *et al.*:[13] the Human Anthropometry Research Group (HUMAG) at Loughborough University, contracted by clothing manufacturers to obtain summary statistics of body measurements in six age-sex groups; the National Study of Health and Growth (NSHG), a study of primary school age children set up in 1972 to monitor growth after the ending of free school milk; the Department of Health (DoH) Study of heights and weights of adults aged 16–64; the Tayside Growth Study, set up in 1989 to investigate the prevalence of short stature in children between 3 and 15 years in the Tayside region of Scotland; the Cambridge Infant Growth Study, a group of infants born in Cambridge and followed longitudinally; and the Whittington Birth Dimensions Study, carried out in 1990 at the Whittington Hospital London to test the value of different body dimensions for identifying small for dates babies.

A second group of five data sets was used to augment the data in early life, particularly weight at birth: the National Diet and Nutrition Survey (NDNS), nationally representative of children aged 1·5 to 4·5 years;[16] the MRC Dunn Nutrition Unit Premature Baby Study, neonates up to 30 weeks gestation with birthweights less than 1850g born in Ipswich, Kings Lynn, Norwich and Sheffield between 1984 and 1988;[17] all babies over 31 weeks gestation born in the Rosie Maternity Hospital, Cambridge, between July 1992 and December 1993; all babies up to 31 weeks gestation born in the Rosie Maternity Hospital, Cambridge between 1985 and 1994; and all babies up to 31 weeks gestation (including fresh stillbirths) born in the Northern Region for the years 1983, 1991 and 1992.

The third group of data sets, used to provide head circumference reference data, consisted primarily of the Edinburgh Growth Study,[18] a longitudinal survey of Edinburgh children followed from birth to age 20, supplemented with head circumference data from three of the

Table I. Details of the studies providing data

Study	Date	Age range (years)	Region*	Sample size‡	Measurements†
HUMAG Infants	1987	0– < 2	E W	791	W H
HUMAG Toddlers	1987	2– < 5	E W	1017	W H
HUMAG Boys	1978	5– < 17	E S	3501	W H
HUMAG Girls	1986	5– < 16	E S W	4285	W H
HUMAG Men	1984	16– < 23	E S W	1752	W H
HUMAG Women	1987	16– < 23	E S W	1058	W H
National Study of Health and Growth	1989–90	4·5– < 12	E S	10531	W H
Department of Health	1980	16– < 23	E S W	1474	W H
Tayside Growth Study	1989–90	4·5–14	S	1624	W H
Whittington Hospital	1987–88	33–44 weeks gestation	E	756	W H O
Cambridge Infant Growth Study	1984–90	4 weeks– 2 years	E	3868 (246)	W H O
National Diet and Nutrition Survey	1992–93	1·5–4·5	E S W	1768	W H
Cambridge Rosie Neonates	1992–93	32–44 weeks gestation	E	6544	W
Cambridge Rosie Premature Neonates	1985–94	23–31 weeks gestation	E	1088	W O
Dunn Premature Baby Study	1984–90	23–30 weeks gestation	E	298	W
Northern Region Premature Neonates	1983 1991–92	23–31 weeks gestation	E	758	W
Edinburgh Growth Study	1972–90	0– < 18 (male) 0– < 17 (female)	S	5660 (177)	O

* E = England, S = Scotland, W = Wales
† W = weight, H = height, O = occipito-frontal circumference (= head circumference)
‡ Brackets indicate the number of distinct subjects in the longitudinal studies

studies described above: the Rosie premature babies; Whittington term babies, and Cambridge Growth Study infants.

The HUMAG studies were designed to be nationally representative, based on England and Scotland for some surveys, and England, Scotland and Wales for others. The DoH and NDNS studies were random samples of England, Wales and Scotland, while the NSHG had two separate surveys, one representative of England and the other of Scotland. All the other studies were representative only of their locality. Note that Cambridge-born babies in the Dunn Premature Baby Study were excluded from the analysis, as most were also in the Rosie series.

With the exception of Tayside and the birthweight surveys (where the relevant information was not available), ethnic non-white children were excluded from the analysis due to known differences in growth pattern between ethnic groups.[19] The proportion of non-white children in Tayside is known from the 1991 census to be small (< 2·5 per cent).

Table I summarizes the various studies, giving the sample size, date of data collection, regional representativeness, and the measurements obtained. Further details of most of the studies, particularly the sampling procedures, measurement technique, quality control and other information collected are given elsewhere.[13,17,18] Information on the Cambridge neonates born after 31 weeks and those from the Northern Region was obtained from the local child health computer, while the Cambridge premature baby data were accumulated from the clinic notes (with acknowledgement and grateful thanks to Dr. Clive Glazebrook, Dr. Unni Wariyar and Dr. Janet Rennie, respectively).

Children were measured for weight (adjusted for any clothing weight), and height or length (depending on age – usually length before 2 years and height thereafter). Head circumference was measured in four studies. In addition sex and age (calculated from dates of birth and measurement) were recorded, except for NDNS where age was known to the nearest month. No adjustment was made for gestational age in the HUMAG, Cambridge or Edinburgh studies during infancy. Weight adjusted for height (Quetelet's index, or body mass index, BMI) was calculated as weight/height^2, in units of kg/m^2. For simplicity length is referred to here as height, and although there is known to be a systematic difference of up to 1 cm between them,[16] no explicit adjustment has been made.

METHODS

Data Cleaning and Standardization

The data were first scrutinized for outliers, using an informal cut-off of ± 5 standard deviation scores (SDS). Fewer than 0·1 per cent of the data were excluded on this basis, although most of the studies had been cleaned before and the obvious outliers dealt with.

It became clear at an early stage that there were appreciable offsets between the data sets. Some heterogeneity was inevitable due to differences in geography (children are larger in the South-East than the North-West) and survey date (due to the secular trend in height). It was decided to make the data sets as similar as possible, by adjusting out the differences between them while simultaneously adjusting for age group and geographical region, using the NSHG 1990 survey and the South-East region as baselines. The adjustments were done for weight and height separately, analysed by multiple regression after a natural log transformation, grouping the data into 3-month age groups and 11 geographical regions in addition to the between-data-set effects. The sexes were analysed together, including a sex by age interaction, to constrain the data set and region effects to be the same for the two sexes. Note that the HUMAG child and adult surveys were coded as distinct data sets by sex as they were carried out in different years. The results were then used to adjust individual data points to the common baseline, by dividing by the antilog of the relevant data set and region offsets.

Most of the data sets overlapped in age with others, ensuring consistent estimates for the between-data-set differences. However this was not the case for the birth and head circumference data, so they were left unadjusted.

Model fitting

The statistical basis for the calculation of reference centiles, the *LMS* method, is essentially that described by Cole and Green.[6,20] In brief, it assumes that for independent positive data y_i at ages t_i $(i = 1, \ldots, n)$ an age-specific Box–Cox transformation can be applied to the data to make them Normally distributed. In general the t_i are replicates from a smaller set of distinct ordered ages T_j $(j = 1, \ldots, m)$. The distribution of the y_i at T_j (where $t_i = T_j$) is summarized by the median $M(T_j)$, coefficient of variation $S(T_j)$, and Box–Cox power $L(T_j)$, so that the formula

$$z_i = \frac{[y_i/M(T_j)]^{L(T_j)} - 1}{L(T_j)S(T_j)} \qquad (1)$$

converts measurement y_i to its Normal equivalent deviate (NED) z_i. For the case when $L(T_j)$ is zero:

$$z_i = \frac{1}{S(T_j)} \log_e[y_i/M(T_j)]. \quad (2)$$

The quantities L, M and S are natural cubic splines with knots at each T_j, and are estimated by maximum penalized likelihood. The complexity of each spline is measured by its equivalent degrees of freedom (e.d.f.).[20] The e.d.f. for each curve are analogous to the degrees of freedom of a polynomial, and range from 2 upwards. The lower bound of 2 corresponds to an infinitely smoothed curve, that is, a straight line, while larger e.d.f. values correspond to progressively rougher spline curves.

A simple alternative to a smooth curve is a horizontal straight line corresponding to a constant value estimated over the whole age range. This is referred to subsequently as a curve with 1 e.d.f.

When the median (M) curve is complex in shape but also monotonic, a less biased estimate of the curve can be obtained by using a two-stage process based on the principle of Rao:[21] 'transform the timescale to a time metameter which makes the group average curve linear'. The process is described in the Appendix.

This two-stage process does not work if the M curve is non-monotonic, but a simple alternative is to log transform the age scale. This is also useful for modelling asymptotic behaviour at older ages. Other power transformations (with a suitable offset) are obvious alternatives.

The original FORTRAN program written to fit the model[6] has been modified and extended. Genstat 5 release 3[22] was used for the subsidiary analyses, and the graphs were drawn with Genstat and Microsoft Excel 4.

Significance testing

The choice of smoothing parameter for each of the three curves is aided by the use of a likelihood ratio test. The increment in penalized log-likelihood for a given change of q equivalent degrees of freedom is distributed approximately as $1/2\chi^2$ on q d.f. This distributional assumption *is* only approximate, but nevertheless is useful for informal testing of model adequacy.

Derivation of centiles

Once the LMS curves have been estimated, any required centile curve can be derived from them by inverting equation (1). For the 100αth centile, where α defines the lower tail area of the centile and Z_α is the corresponding NED, the measurement centile at age T_j is given by

$$C_{100\alpha}(T_j) = M(T_j)\,(1 + L(T_j)S(T_j)\,Z_\alpha)^{1/L(T_j)} \quad (2)$$

or if $L(T_j)$ is zero

$$C_{100\alpha}(T_j) = M(T_j)\exp(S(T_j)\,Z_\alpha). \quad (2a)$$

Goodness-of-fit

Goodness-of-fit is conventionally assessed using an omnibus test covering both skewness and kurtosis, for example, reference 23, but here the skewness has been explicitly removed. The simplest approach is to consider skewness and kurtosis separately, analysing the data overall and in 10 equally sized age groups.

Confidence intervals

The degree of uncertainty in the estimated L, M and S curves and derived centile curves is obtained by simulation. The underlying assumption of the LMS method is that after Normalization, the data are converted to NEDs. Therefore on this assumption it is possible to generate a set of random Normal deviates, at the same ages as the original data, and use the fitted LMS curves and equation (2) to transform them back to the original scale. This simulated data set can then be fitted from scratch, using the e.d.f. values chosen for the original data, leading to a new set of LMS curves.

Repeating the process leads to a family of such curves, and the values of the curves at each age can be ranked. Normal order statistic theory then allows approximate confidence intervals of any size to be constructed from the ranks. For example, if 25 sets of simulated data are fitted, then the extremes are on average 1·97 standard errors above and below the mean,[24] so that the range represents an approximate 95 per cent confidence interval. A larger simulated data set would conventionally be used to test the Normality assumption and estimate the confidence interval more precisely, but it involves a substantial time penalty.

Centile curves other than the median are calculated from each set of simulated LMS curves (equation (2)), and the process described above is applied to the ranked values of the centile curves at each age to obtain confidence intervals.

RESULTS

Data standardization

The data standardization results are given in Table II for height and weight, showing the estimated differences between data sets and their standard errors. The regional differences, which are generally smaller, will be published elsewhere. The response variates are transformed to natural logarithms $\times 100$ so the coefficients are percentage differences. Most of the differences are numerically small and statistically insignificant, but the t statistics exceed 5 for height and weight in the HUMAG boys and weight in the HUMAG girls. Why these two surveys should be so out of step with the others is a mystery.

Weight

There are 21,123 weight measurements for boys and 19,876 for girls, from 23 weeks gestation to 22·98 years. The M curves are fitted using the two-stage process, illustrated in Figure 1(a) for boys, the curves offset for clarity. Despite its complex shape, curve (a) with 15 e.d.f. fails to fit early in infancy. Curve (b) with 30 e.d.f. fits well in infancy, but is not smooth at older ages. We need the smoothness of 15 e.d.f. and the good fit of 30 e.d.f.

The solution is to expand the age scale when the velocity is high (infancy and puberty), and to compress it when low (mid-childhood and particularly adulthood), so as to make curve (a) straight. Refitting the M curve on this transformed age scale, again with 15 e.d.f., leads to curve (c) in Figure 1(a), which is straight apart from slight deviations in infancy and puberty. The notation 15* e.d.f. means that the curve is fitted with 15 e.d.f. on the transformed scale. Curve (d) is curve (c)

Table II. Estimates of the percentage offsets in height and weight between data sets, derived as 100 times the between-set regression coefficients for the logged data. See text for details

Data set	100 × $\log_e$ height			100 × $\log_e$ weight		
	Regression coefficient	Standard error	t	Regression coefficient	Standard error	t
HUMAG Infants	− 0·09	0·36	− 0·3	− 4·3	1·3	− 3·4
HUMAG Toddlers	+ 0·07	0·22	0·3	+ 1·2	0·8	1·5
HUMAG Boys	− 1·01	0·11	− 9·6	− 2·0	0·4	− 5·2
HUMAG Girls	− 0·32	0·11	− 3·1	− 2·7	0·4	− 7·1
HUMAG Men	− 0·27	0·24	− 1·1	− 1·1	0·9	− 1·2
HUMAG Women	+ 0·27	0·31	0·9	− 1·7	1·1	− 1·5
National Study of Health and Growth	0	–	–	0	–	–
Department of Health	− 0·85	0·23	− 3·7	− 1·7	0·8	− 2·0
Cambridge Infant Growth Study	− 0·41	0·35	− 1·2	− 0·7	1·3	− 0·6
Tayside Growth Study	− 0·44	0·13	− 3·4	+ 0·4	0·5	0·9
National Diet and Nutrition Survey	+ 0·42	0·21	2·0	+ 0·8	0·8	1·1

plotted on the original untransformed age scale, denoted by 15** e.d.f., and it fits well at all ages (that is, close to curve (b) in infancy and asymptotic in adulthood), and it is smooth throughout.

To simplify comparisons, the M curves in Figure 1 are fitted assuming a constant power transform and a constant CV at all ages (that is, 1 e.d.f. for the L and S curves). The estimated power and CV are $-0{\cdot}19$ and 0·1485 for curve (a) and $-0{\cdot}29$ and 0·1466 for curves (b)–(d), the negative power indicating more skewness than a log-Normal distribution. Curve (b) is set to 30 e.d.f. to provide the same CV as curve (d). Figure 1(b) highlights the differences between curves (a) and (d), expressed in weight SD units as $\log_e(M_{15**}/M_{15})/0{\cdot}1466$ (using an obvious notation). The differences are concentrated in infancy, with smaller effects in puberty and adulthood. The effect in infancy amounts to a swing about zero of more than 1 SD, emphasizing the need for the two-stage fitting process.

To allow the CV to vary with age, the S curves with each of the M curves in Figure 1(a) are refitted with 9 e.d.f. The corresponding penalized log-likelihood ratios are 432·5 for curve (a), 455·4 for (b) and 569·1 for (d), indicating highly significant variability in the CV with age. In addition the fit is appreciably better on the modified age scale.

The power transform is then allowed to change with age, fitting L curves with 6 e.d.f. They give log-likelihood ratios of 277·1, 265·4 and 249·5 for (a), (b) and (d), respectively, demonstrating that the degree of skewness is highly age-dependent, and that the three curves fit equally well. Figure 2 shows the L and S curves corresponding to M curves (b) and (d) in Figure 1. The (d) curves have a more clearly defined fall in infancy and peak at puberty.

For girls' weight, the fitted L, M and S curves have 5, 15 and 9 e.d.f. Figure 3 compares the three curves in the two sexes and confirms their similar shape. The L curves show clear trends in skewness with age, from Normality at birth ($L = 1$) to a log transform at 1 year ($L = 0$), to an inverse transform at 9 years ($L = -1$) and back to a log at 14 years. Figure 4 illustrates this for the boys. Both L curves peak during puberty (2 years earlier in girls than boys), corresponding to a reduction in skewness, then level off in adulthood at a value of about $-0{\cdot}7$.

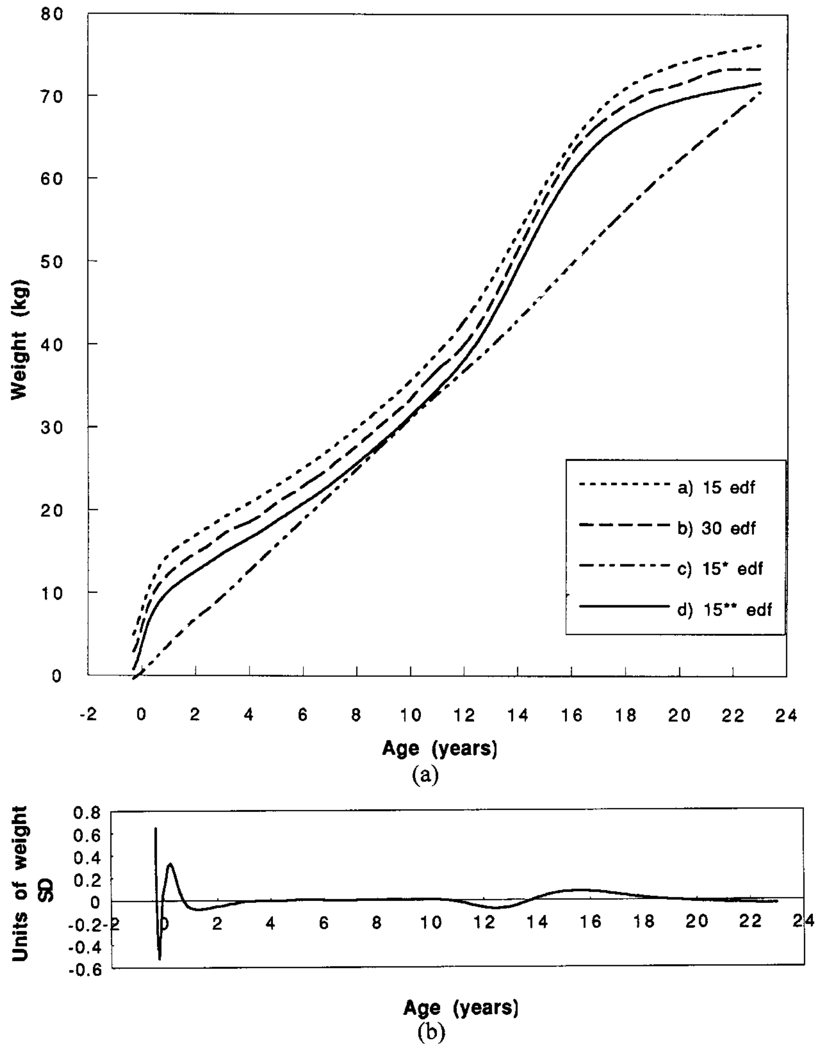

Figure 1. Illustration of the two-stage process for fitting the boys' weight M curve. (*a*) Comparison of boys' weight M curves. (*b*) Comparison of M curves with 15** e.d.f. and 15 e.d.f. MSD units

The S curves fall steeply to a minimum at 1 year, then rise fairly linearly to a peak in puberty. Compared with the L curves, the peaks on the S curves occur 1 year earlier in boys and 2 years earlier in girls. The M curves for both sexes show a steep rise in weight for the first few months of life, then a more steady rise until puberty. Boys are slightly heavier than girls until puberty, when girls are briefly heavier. Past puberty boys become 20 per cent heavier than girls, and continue to increase in weight after girls have stopped.

The shapes of the L and S curves are reminiscent of the rate at which weight changes during childhood. Figure 5 illustrates this with the derivatives with respect to age of the M curves for the

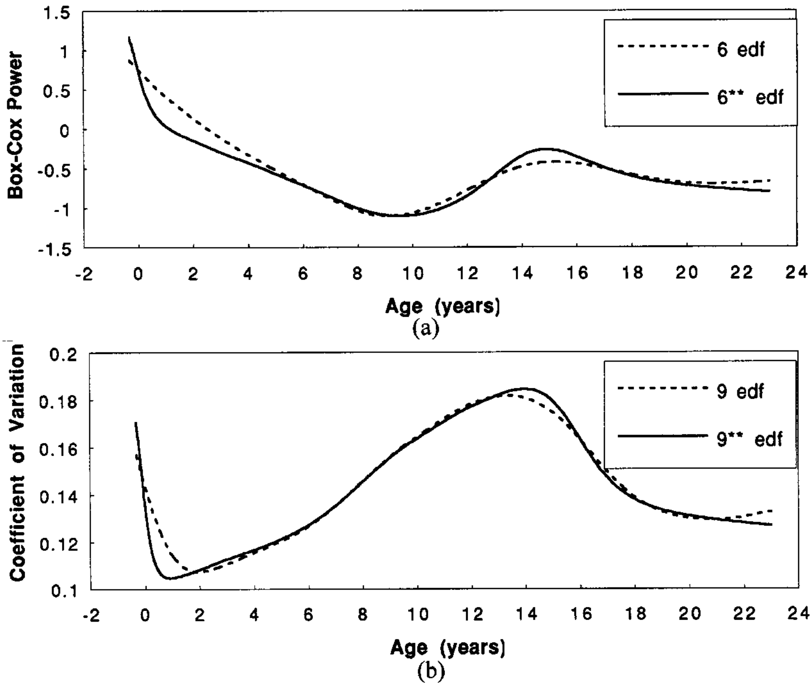

Figure 2. *L* and *S* curves for boys' weight fitted on the original and transformed age scales: (*a*) *L* curves; (*b*) *S* curves

two sexes. The curves are rougher in mid-childhood than the *M* curves, as is inevitable with first derivatives of spline curves. They show an initial rise and a sharp fall in weight 'velocity' during infancy, then a period of fairly steady growth, followed by a pubertal peak in weight velocity and growth slowing down as adulthood approaches. The initial rise, which highlights the complexity of the *M* curve's shape, is due to the trend in birthweight with gestational age in premature infants, the velocity peaking at 36 weeks.

Figure 6 shows the weight reference by sex in nine-centile format.[25] The centiles are spaced two-thirds of a standard deviation score (SDS) apart, covering the range $\pm$ 2·67 units, and the extreme curves correspond to the 0·4th and 99·6th centiles. The skewness of the centiles changes with age, and the 99·6th centiles are rough during puberty because the *L* and *S* curves peak at different ages.

Table III summarizes the distribution of weight SDS in the two sexes, as calculated from (1). The median, mean and skewness are all near 0 and the SD near 1, and most are far from significant. However, the kurtosis exceeds 0·3 and is highly significant. Table IV shows the proportions of data in the channels around the nine centiles of Figure 6, and there is an excess of some 40 per cent in the two tails beyond $\pm$ 2·67 SDS, that is, 0·5–0·6 per cent rather than the 0·38 per cent expected. Beyond $\pm$ 2 SDS the kurtosis is less marked, 2·5 per cent as against 2·3 per cent expected. Thus in practical terms the heaviness of the extreme tails is of little consequence.

Looking at the distribution of weight SDS by tenths of age, most groups are similar to the overall pattern. A notable exception is in infancy, where the SD changes rapidly in both sexes,

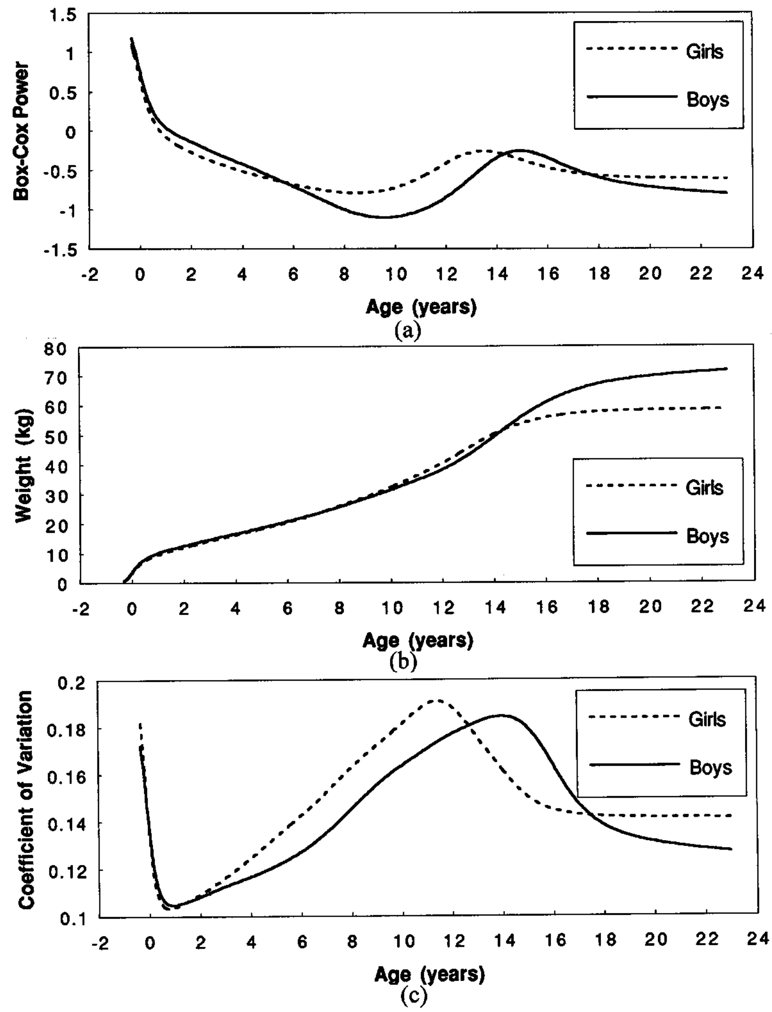

Figure 3. *L*, *M* and *S* curves for weight by sex: (*a*) *L* curves; (*b*) *M* curves; (*c*) *S* curves

with the SD for premature infants appreciably above 1, and for term infants well below 1. This is illustrated in Figure 7, where the fitted reference is compared with an analogous reference based only on birthweight data. The 1990 reference centiles are too close together from 28 to 36 weeks, and too far apart at 44 weeks. There is also significant positive skewness in the third tenth of age, corresponding to the first year after birth.

The estimation of confidence intervals is illustrated in Figure 8, which shows 25 sets of *L*, *M* and *S* curves, based on 6, 15 and 9 e.d.f., respectively, obtained by simulation assuming that weight SDS is Normally distributed. The range in each plot corresponds roughly to a 95 per cent confidence interval for the curve. The *M* curve is estimated very precisely, and the *S* curve is also

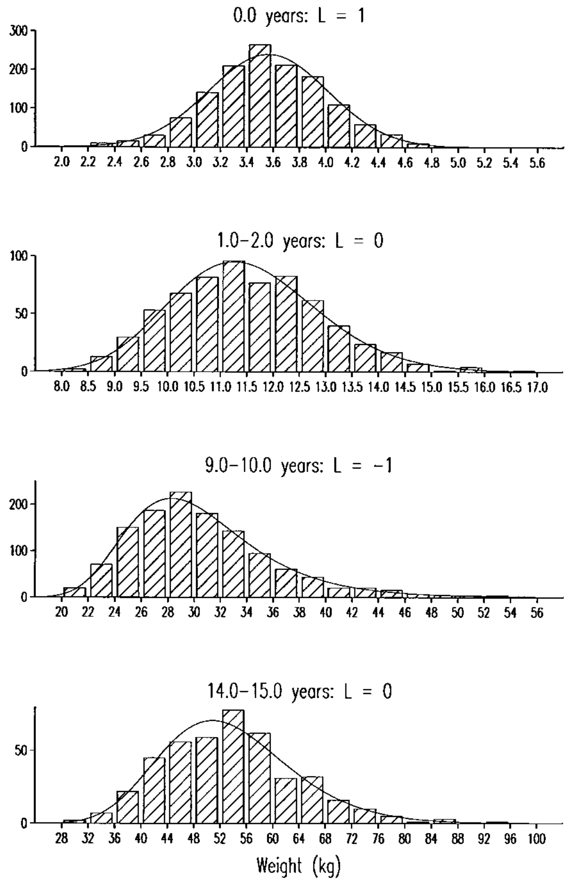

Figure 4. Histograms of boys' weight with superimposed skew Normal plots to illustrate the change in skewness with age

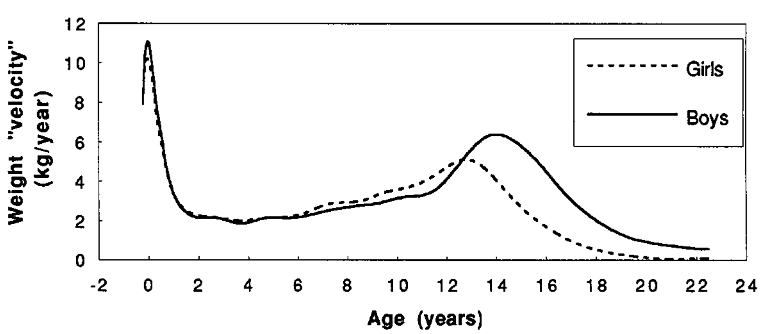

Figure 5. First derivatives of the weight M curve by sex

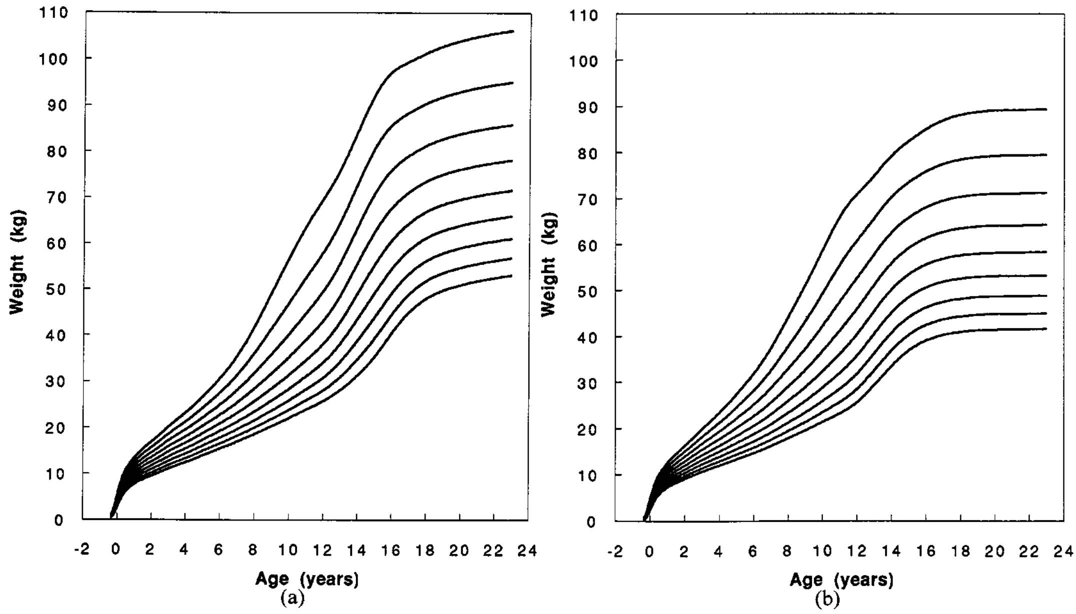

Figure 6. British 1990 weight reference centiles by sex, spaced two-thirds of an SDS apart: (*a*) boys; (*b*) girls

Table III. Summary statistics for the four measurements expressed as SDS, by sex

		n	Median	Mean	SD	Skewness	SE	Kurtosis	SE
Weight	Boys	21123	0·01	0·00	1·00	0·01	0·02	0·37	0·03
	Girls	19876	− 0·01	0·00	1·00	0·01	0·02	0·34	0·03
Height	Boys	16548	0·00	0·00	1·00	0·01	0·02	0·17	0·04
	Girls	15786	− 0·01	0·00	1·00	0·01	0·02	0·17	0·04
BMI	Boys	16491	− 0·02	0·00	1·00	− 0·06	0·02	1·36	0·04
	Girls	15731	− 0·02	0·00	1·00	− 0·03	0·02	0·67	0·04
OFC	Boys	6444	0·00	0·00	1·00	0·00	0·03	0·56	0·06
	Girls	4917	− 0·03	0·00	1·00	− 0·07	0·03	0·33	0·07

well specified, but the L curve shows a lot of variability. The curves are truncated at 25 weeks gestation, owing to difficulties fitting the model down to 23 weeks.

The sets of LMS curves are combined using equation (2) to provide 25 sets of nine simulated centiles, and the ranges of the centiles are shown in Figure 9 plotted on an SDS scale against age. The confidence intervals are on the whole centred on the nominal SDS value and small, but tend to be larger on the more extreme centiles and at the extremes of age. Variability in the timing of the pubertal peak also leads to a wider interval for the extreme centiles at this age, occurring slightly earlier for the upper than the lower centiles. The widths of the confidence intervals,

Table IV. Percentages of the measurements, expressed as SDS, falling in the 10 channels defined by the 9 centiles. The expected percentages based on a Normal distribution are also shown

		1	2	3	4	5	6	7	8	9	10
Weight	Boys	0·5	2·0	6·3	15·6	24·8	26·4	15·6	6·2	1·9	0·6
	Girls	0·6	1·7	6·3	16·4	25·1	25·1	16·0	6·4	1·9	0·6
Height	Boys	0·5	1·5	7·0	16·3	24·6	25·1	16·2	6·5	1·9	0·5
	Girls	0·4	1·8	6·5	16·2	25·0	24·6	16·3	6·7	1·8	0·5
BMI	Boys	0·5	1·6	6·2	15·9	26·4	25·6	15·5	5·5	2·4	0·6
	Girls	0·5	1·6	6·4	16·0	26·0	25·1	15·5	6·3	2·1	0·5
OFC	Boys	0·6	1·6	6·7	15·3	25·5	26·1	15·6	6·1	1·8	0·7
	Girls	0·9	1·8	5·7	15·1	27·9	23·8	16·2	6·2	2·0	0·5
Expected		0·38	1·89	6·85	16·13	24·75	24·75	16·13	6·85	1·89	0·38

averaged across age, range from 0·11 SDS units on the median to 0·27 units on the extreme centiles. Multiplied by the corresponding values of M and S at each age, they correspond in weight units to about 55 g on the median at birth up to 2·5 kg on the 99·6th centile at 20 years.

Height

There are 16,548 and 15,786 heights for boys and girls, respectively, between the ages of 33 weeks gestation and 22·98 years. The M curves have 15 e.d.f. for boys and 14 e.d.f. for girls, with S curves of 8 e.d.f., fitted on the modified age scale with $L = 1$ at all ages. Estimating a constant value for L rather than forcing it to 1 gives a power of 0·95 for boys and 0·92 for girls, with log-likelihood ratios of 0·9 and 0·2 compared with the Normal distribution. Since the CV of height is small, the effect on the skewness of the centiles is minimal. For this reason L is fixed at 1.

Figure 10 compares the M and S curves for the two sexes, showing similar age-related features but puberty occurring two years earlier in girls. The S curves fall steeply from birth then rise again at 9 months, continuing to rise until puberty and then falling as adulthood approaches. The rise in CV before puberty is non-linear and shows a distinct shoulder in mid-childhood.

Table III summarizes the distribution of height SDS, and as for weight the distribution is very close to Normal, with zero mean, median and skewness, and an SD of 1. Kurtosis is less than for weight, although still significant. Table IV shows the proportions of the data between the nine centiles, and there is an excess of less than 1 per 1000 (0·5 per cent versus 0·38 per cent) beyond the extreme centiles, while the tails beyond SDS $= \pm 2$ exactly match the expected size of 2·3 per cent.

Body Mass Index

For body mass index there are 16,491 measurements in boys and 15,731 in girls, between 33 weeks gestation and 22·98 years. As the M curve is non-monotonic, age is transformed to $\log_e(\text{age} + 0{\cdot}27)$ before fitting, the offset allowing the premature data to be included. The chosen LMS e.d.f. for boys are 4, 13 and 7, respectively, and for girls 4, 12 and 7. Figure 11 compares the fitted LMS curves by sex.

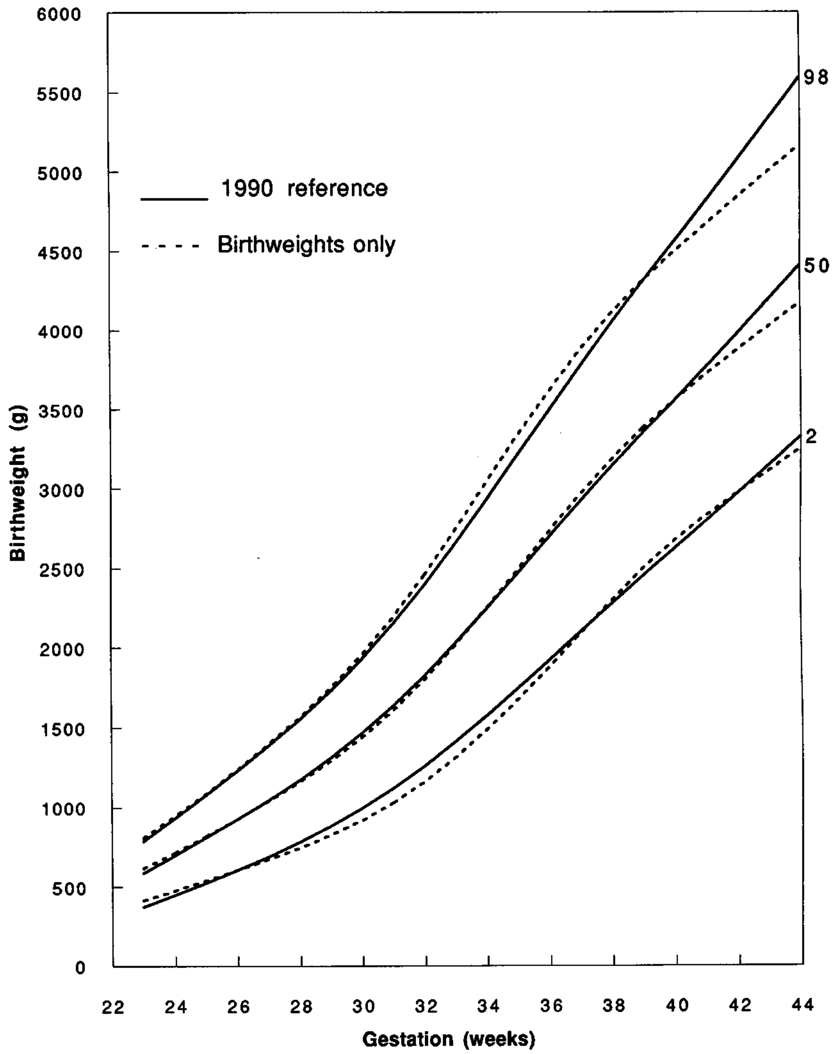

Figure 7. 2nd, 50th and 98th boys' weight centiles up to 44 weeks gestation, as obtained from the 1990 reference and a birthweights-only reference. See text for details

Table III summarizes the distribution of normalized body mass index by sex, showing a somewhat poorer fit than for weight or height. Skewness in boys is small but significant, while kurtosis is large and highly significant in both sexes, and much more so than for weight or height. The median, mean and SD are close to expectation. In terms of the effect on the distribution, the skewness is much more obvious than the kurtosis (Table IV), with 2·1 per cent below -2 SDS and 2·6 per cent (girls) or 2·9 per cent (boys) above $+2$ SDS. On average 0·51 per cent of the sample exceeds $\pm 2{\cdot}67$ SDS and 2·5 per cent exceeds ± 2 SDS. These departures from Normality are not of practical importance.

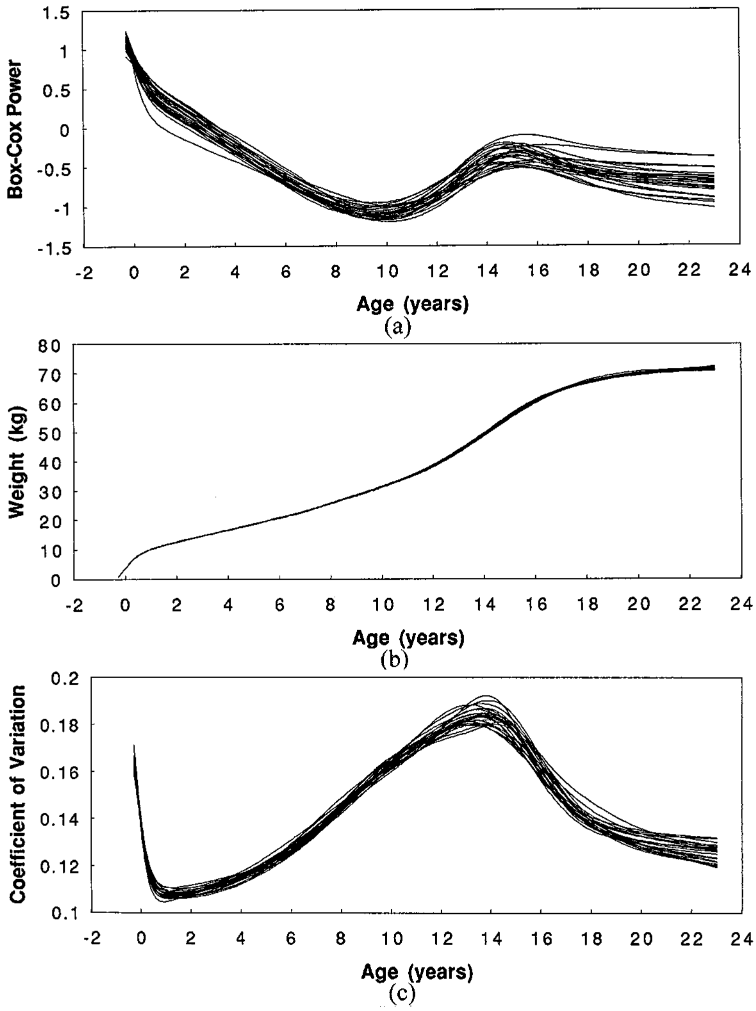

Figure 8. Sets of 25 simulated *L*, *M* and *S* curves for boys' weight: (*a*) *L* curves (6 e.d.f.); (*b*) *M* curves (15 e.d.f.); (*c*) *S* curves (9 e.d.f.)

Head circumference

There are 6444 head circumference measurements for boys and 4917 for girls, from 23 weeks gestation to 18 years (boys) or 17 (girls). The data are truncated at the upper end due to reducing numbers. As for weight and height, the age trend in head circumference is too complex for a single spline curve. However, the fitting process for the two-stage curve is unstable, for reasons that are not clear, and instead the curves are fitted on an age scale transformed to $\log_e(\text{age} + 0{\cdot}65)$.

The chosen e.d.f. for the *M* curve are 12 for boys and 11 for girls, with 5 for the *S* curve in both sexes. The CV of head circumference is very small, less than 3 per cent past infancy, and there is no

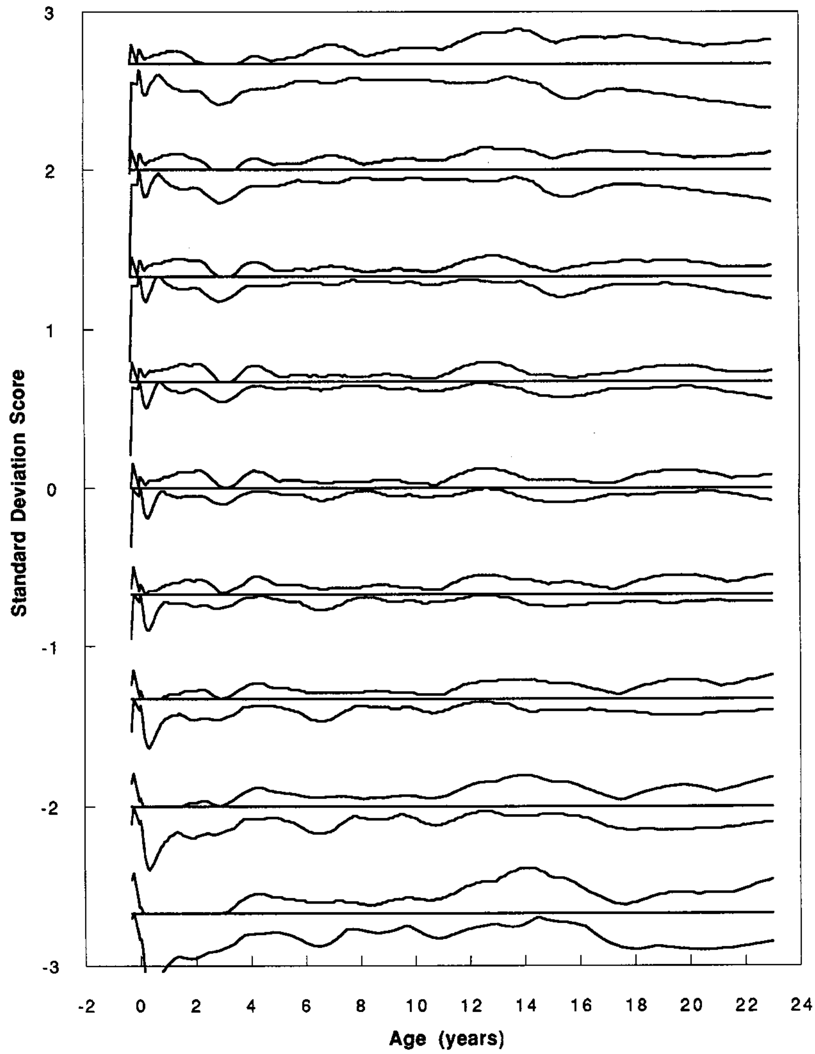

Figure 9. 95 per cent confidence intervals for boys' weight centiles

evidence of skewness, so L is set to 1 throughout. Figure 12 compares the M and S curves for head circumference by sex, showing very similar trends with age. Most growth in head circumference occurs in the first 5 years of life, although there is a suggestion of a pubertal spurt in boys but not in girls.

Table III summarizes the mean, median, SD, skewness and kurtosis of normalized head circumference. The first two moments correspond to those of the Normal distribution, but there is slight evidence of skewness in the girls and the tails are relatively heavy. Table IV shows that 0·6 per cent of the sample are beyond $\pm 2{\cdot}67$ SDS and 2·5 per cent beyond 2 SDS.

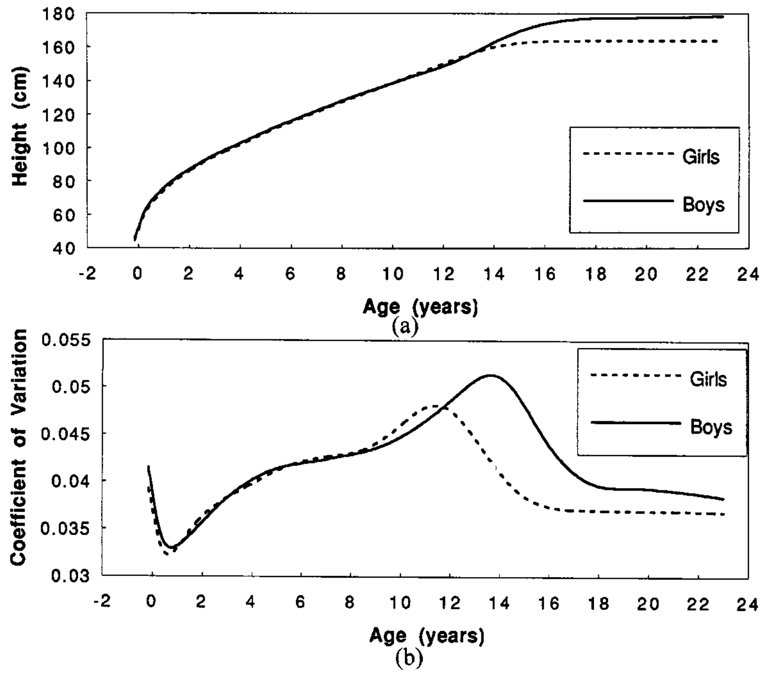

Figure 10. M and S curves for height by sex: (*a*) M curves; (*b*) S curves

DISCUSSION

The British 1990 growth reference consists of centiles for weight, height, body mass index and head circumference from birth to 23 years. Fitted by the *LMS* method and maximum penalized likelihood, they allow measurements to be converted to Normally distributed SDS.

The reference is thought to be the first covering all childhood to model both the centile shapes and centile spacings statistically. The American NCHS reference centiles[26] were fitted with cubic splines, and more recently Guo *et al.*[27] used kernel regression to model centiles of skinfold thickness, but in each case the spacings between centiles were not modelled, and the lower age range was 1 year not birth.

The reference data used to construct the centiles here were obtained by merging several distinct data sets. Compared with a single dedicated survey this was less than ideal, but it was the only realistic alternative given the cost of obtaining a new national sample. Nine of the surveys were nationally representative (Table I), but two of them, the HUMAG boys and girls, proved substantially different from the others, and this led to the data set differences being adjusted out statistically. In most cases the adjustments were small and insignificant (Table II), and their effects on the fitted centiles similarly small – a slight increase in the M curves, and a trivially small reduction in the S curves, between 5 and 16 years. However this remains a potential weakness of the growth reference.

Originally the data sets were adjusted for each sex separately, which led to the boys' and girls' weight centiles in infancy being too close together,[28] and necessitated a re-analysis constraining

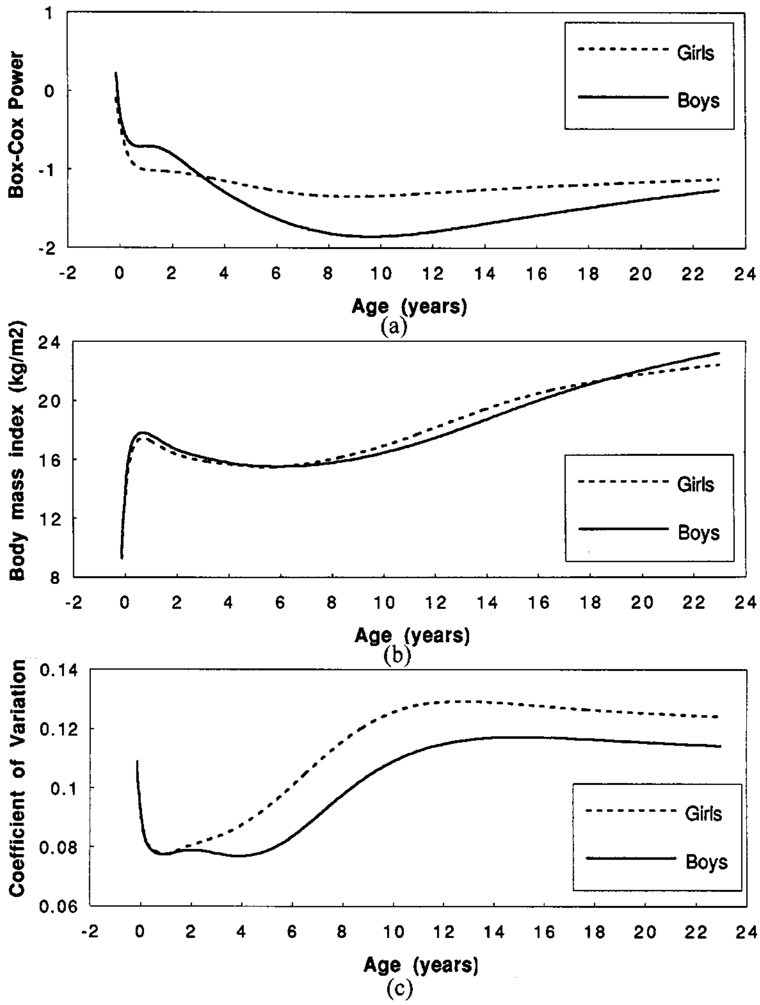

Figure 11. L, M and S curves for body mass index by sex: (*a*) L curves; (*b*) M curves; (*c*) S curves

the data set adjustments to be the same in the two sexes. There is no sex bias in the current version of the reference.

With these provisos, the reference sample is representative of British ethnically Caucasian children, and none is excluded on health grounds. Hence it is a growth *reference* not a growth *standard*. This extends even to weight at birth, by contrast to recent birth weight standards[29, 30] where infants born between 28 and 31 weeks were excluded if labour was induced. The argument is that such infants are induced because they are growing poorly, and hence are lighter than those delivering spontaneously.

This is an example of the more general problem – should the reference sample be representative of the whole population, or should it be based on a selected group? At older ages the criterion for

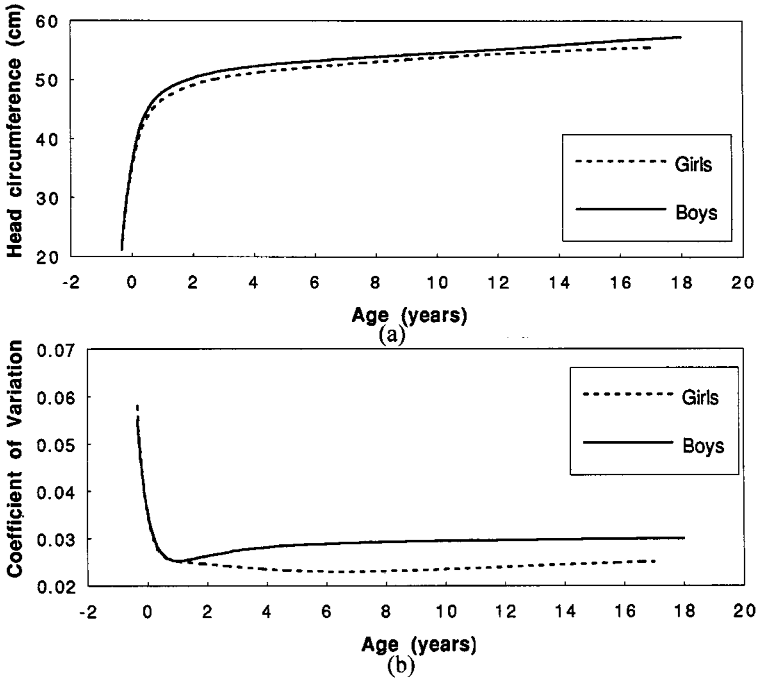

Figure 12. *M* and *S* curves for head circumference by sex: (*a*) *M* curves; (*b*) *S* curves

inclusion is that the child should be representative, and to use a different criterion for premature babies is inconsistent. As it happens, the poor fit of the birthweight reference at early gestations (see Figure 7) raises the lower centiles, and this increases the chance of a poorly grown infant being detected. So in practice the reference is similar to others. Even so, the decision to include all babies will not satisfy everyone.

The reference sample excludes children from ethnic minorities on the grounds that they may grow differently from white children. This is also controversial, as in principle it prevents the growth of ethnic minority children from being assessed. However, the alternative, of producing ethnic-group-specific references, is even less satisfactory, as the required large and representative samples are simply not available, and the definition of ethnicity is in any case a minefield. A better answer would be to use a series of small-scale surveys to summarize the growth status of specific ethnic minority children in terms of their mean SDS on the British reference, and this could be used to recalibrate the reference for use with such groups.

The age range of the reference is unusually wide, extending for weight from 23 weeks of gestation to 23 years. Both extremes pose problems for statistical modelling, with the centiles in prematurity changing slope rapidly and those in adulthood approaching an asymptote. The refitting of the age scale to ensure an unbiased *M* curve works well for monotonically increasing weight and height, even though the first derivatives of the curves are non-monotonic. It also has the useful side-effect that changes in variability and skewness can be modelled on a scale of constant velocity, so that at ages when growth is fastest, the CV and skewness can change rapidly too. However the procedure fails to deal with non-monotonic data such as the BMI. Log transforming the age scale is only a partial solution, and perhaps a fractional polynomial transformation[12] would work better.

In theory the two approaches of age transformation and age refitting could be used together, transforming the age scale using say a fractional polynomial, and then fitting a two-stage M curve. A third alternative, to log or square root transform the measurement scale, may be useful for modelling exponential decay in the M curve.[7]

The LMS method expresses the reference data as SDS with a mean, median and skewness very close to zero, and the SD is consistently close to 1. Thus the first three moments of the distribution are effectively Normal. There is some evidence of kurtosis in all four measurements, but its practical significance is slight. The proportion of children lying beyond the 0·4th and 99·6th centiles is 20–40 per cent higher than expected, but this represents an excess of only 1 or 2 children per 1000. For the 2nd and 98th centiles the excess is again no more than 1–2 per 1000. Healy[31] pointed out that when fitting centiles the main interest is in the tails of the distribution, where the data are inevitably sparse, so that some extrapolation is needed. The evidence here shows that extrapolating to $\pm 2{\cdot}67$ SDS is not unjustified. The kurtosis could be removed by fitting a 4-parameter Box–Cox transformation with an age-dependent offset, but it would make little practical difference.

This supports the use of the 9-centile format proposed by Cole,[25] where the 0·4th and 99·6th centiles are added to the conventional 7 centiles to reduce the false positive rate associated with screening based on the 3rd or 5th centiles. The results show that the extra centile is likely to pick up 5 or 6 rather than the nominal 4 children per 1000.

It is not that well known that the skewness of the plotted centiles depends not only on the Box–Cox power but also on the CV of the measurement. This can be seen from equation (2), where the 100α centile expressed as a fraction of the median is $(1 + LSZ_{\alpha})^{1/L}$. Thus for height, where the CV is only one-quarter that for weight, it is very difficult to detect skewness in the distribution.

Approximate confidence intervals for the LMS curves and centiles are obtained by simulation. They quantify the sampling variation in the LMS curves and their corresponding centiles on the assumption that the underlying model is correct. Owing to the time required to fit each simulated data set, the number of simulated samples is kept small. The results (Figure 9) show that even the extreme centiles are precisely estimated at most ages except the youngest and oldest. This emphasizes the need to have large sample sizes at the extremes of the age range, to minimize edge effects. However, the 95 per cent confidence interval touches the nominal centile curve (Figure 9) on the lower centiles in the first year after birth, where the set of simulated curves fails to bracket the original curve. At this age, a referee points out, the simulated L curves (Figure 8) all lie above the original L curve (Figure 2), and this might be expected to lead to a transiently skew distribution, as was indeed observed (see Results).

The shapes of the LMS curves for the four measurements are of interest in their own right. The M curves are already familiar, but the L curves, and particularly the S curves, provide new insight into the structure underlying each reference chart. In each case there is good agreement between the sexes, which provides reassurance that the shapes are actually meaningful. For weight, height and head circumference the CV is high in early infancy, and falls to a minimum around 9 months. The sharpness of the turning point, particularly for height, has not previously been noted. It probably reflects the end of the period of catch-up after birth, when smaller babies grow relatively fast and the lower tail of the distribution shrinks compared with the upper tail. This also explains the sharp increase in weight skewness early on. The troughs on the S curves occur at the same age as peak BMI, which is curious but hard to explain.

The L curve for weight is reminiscent of the weight 'velocity' curve, falling after birth and then peaking transiently during puberty. This suggests that the changes in skewness correspond to deceleration/acceleration/deceleration in the weight distribution. The S curve shows a similar pattern of change. Bowditch[32] and Boas[33] described an increase in the skewness of height at puberty, which they explained as a correlation between height distance and height velocity. Unfortunately it was not possible to show such a change in the present height data, possibly because of the relatively small numbers at that age, but skewness in height is hard to measure anyway. If it matches the pattern seen with weight it is a transient *decrease*, not an increase, in skewness. This suggests that Bowditch and Boas were referring to the increase in skewness leading up to puberty, rather than the transient shift the other way *during* puberty. The pubertal shift in skewness is due to a correlation between weight and maturity; heavy and early-developing children have an earlier age of peak weight velocity, when variability is greatest, and this stretches the upper tail of the weight distribution. Late developers are relatively light and peak later, so that they increase the variability by stretching the lower tail, but at a later age. Hence the transient swing in skewness.

Overall, the results justify the use of the British growth reference for the assessment of growth data, allowing measurements to be converted to SDS that are Normally distributed. Not only is this useful for single measurements, but velocity[34] and conditional[35–37] references are also greatly simplified when based on SDS.

APPENDIX

The first stage of the process is to fit the model using the untransformed timescale. The resulting M curve is then used to transform the timescale monotonically to make the M curve linear when plotted on the new timescale – ideally this requires the M curve itself to be monotonic. For a typical time interval the slope of the M curve is given by

$$V_j = \frac{\delta M_j}{\delta T_j} = \frac{M_{j+1} - M_j}{T_{j+1} - T_j}\,.$$

The weighted mean slope $V_\bullet$ over the age range is given by $\Sigma\,\mathrm{abs}(\delta M_j)/\Sigma\,\delta T_j$, where the absolute value allows the M curve to be non-monotonic. The timescale is then transformed so that the interval δT_j becomes

$$\delta T_{j'} = [\mathrm{abs}(\delta M_j) + \varepsilon]/V_\bullet$$

where ε is added to ensure that $\delta T_{j'}$ is positive when δM_j is zero. The slope $\delta M_j/\delta T_{j'}$ for each time interval on the transformed scale is then equal to $\pm (V_\bullet - \varepsilon)$. If the M curve is non-monotonic this generates a discontinuity at each turning point, where the slope changes sign. Smoothing the discontinuity generates spurious wiggles in the curve at the points of inflexion on either side.

The second stage of the process is to fit the model again using the transformed timescale. This ensures that in regions of high acceleration (for example, infancy), where the velocity on the original scale changes rapidly, the velocity on the transformed scale is effectively constant and the acceleration close to zero, so the spline curve fits well. In addition, the coefficient of variation (S curve) and skewness (L curve) are also assumed to change on a time scale of constant velocity. Thus when growth is rapid and the M curve steep (for example, infancy or puberty), the distribution is allowed to change more rapidly than at ages when the M curve is relatively shallow (for example, mid-childhood or adulthood).

Finally the original timescale is restored. As well as dealing with high acceleration, this also provides a way of modelling asymptotic behaviour. When the M curve is nearly flat, for example, in adulthood, it is compressed almost to nothing on the transformed timescale. The L and S curves are fitted on the transformed scale and the original scale restored, which expands the curves and shrinks their slopes towards zero. Thus when the M curve is almost flat, the L and S curves are constrained to be fairly flat too.

This two-stage process poses problems for the interpretation of the fitted e.d.f. and the penalized likelihood. Because the curve is fitted twice, the number of e.d.f. ought theoretically to be doubled, but in practice the degree of smoothing on the second pass is small. Also, the straightening of the M curve after the first pass means that the corresponding roughness penalty is reduced almost to zero, which needs to be taken into account when comparing one- and two-stage M curves. It does not apply though to the corresponding L and S curves, which are effectively independent of the M curve's shape.

ACKNOWLEDGEMENTS

We acknowledge our debt to Peter Green for his developing the original *LMS* program, and for his advice on estimating confidence intervals. We are also very grateful to the following for making their data available: Sue Chinn; Clive Glazebrook; Gaye Henson; Peter Jones; Alan Lucas; Alison Paul; Shirley Ratcliffe; Jan Rennie; Unni Wariyar; Liz White, and Roger Whitehead. Two anonymous referees made valuable comments on previous versions of the paper. J.V.F. was supported by the Child Growth Foundation. Finally, we are happy to record that this work would probably not have been done were it not for the continuing efforts of Tam Fry.

Clinical and community charts of weight, height, body mass index and head circumference, plus a diskette containing the chart *LMS* coefficients in Microsoft Excel format, can be obtained through the Child Growth Foundation, 2 Mayfield Avenue, London W4 1PW, U.K.

REFERENCES

1. Healy, M. J. R., Rasbash, J. and Yang, M. 'Distribution-free estimation of age-related centiles', *Annals of Human Biology*, **15**, 17–22 (1988).
2. Cole, T. J. 'Fitting smoothed centile curves to reference data', *Journal of the Royal Statistical Society, Series A*, **151**, 385–418 (1988).
3. Pan, H. Q., Goldstein, H. and Yang, Q. 'Non-parametric estimation of age-related centiles over wide age ranges', *Annals of Human Biology*, **17**, 475–481 (1990).
4. Thompson, M. L. and Theron, G. B. 'Maximum likelihood estimation of reference centiles', *Statistics in Medicine*, **9**, 539–548 (1990).
5. Pan, H. Q., Goldstein, H. and Guo, D. 'A two-level cross-sectional model using grafted polynomials', *Annals of Human Biology*, **19**, 337–346 (1992).
6. Cole, T. J. and Green, P. J. 'Smoothing reference centile curves: the LMS method and penalized likelihood', *Statistics in Medicine*, **11**, 1305–1319 (1992).
7. Wade, A. M. and Ades, A. E. 'Age-related reference ranges – significance tests for models and confidence intervals for centiles', *Statistics in Medicine*, **13**, 2359–2367 (1994).
8. Wade, A. M., Ades, A. E., Salt, A. T., Jayatunga, R. and Sonksen, P. M. 'Age-related standards for ordinal data: modelling the changes in visual acuity from 2 to 9 years of age', *Statistics in Medicine*, **14**, 257–266 (1995).

9. Nychka, D. and Ruppert, D. 'Nonparametric transformations for both sides of a regression model', *Journal of the Royal Statistical Society*, *Series B*, **57**, 519–532 (1995).
10. Wright, E. M. and Royston, P. 'A comparison of statistical methods for age-related reference intervals', *Journal of the Royal Statistical Society*, *Series A*, **160**, 47–69 (1997).
11. Thompson, M. L. and Fatti, L. P. 'Construction of multivariate centile charts for longitudinal measurements', *Statistics in Medicine*, **16**, 333–345 (1997).
12. Royston, P. and Altman, D. G. 'Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion)', *Applied Statistics*, **43**, 429–467 (1994).
13. Freeman, J. V., Cole, T. J., Chinn, S., Jones, P. R. M., White, E. M. and Preece, M. A. 'Cross-sectional stature and weight reference curves for the UK, 1990', *Archives of Disease in Childhood*, **73**, 17–24 (1995).
14. Cole, T. J., Freeman, J. V. and Preece, M. A. 'Body Mass Index refernce curves for the UK, 1990', *Archives of Disease in Childhood*, **73**, 25–29 (1995).
15. Tanner, J. M., Whitehouse, R. H. and Takaishi, M. 'Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965 Parts I and II', *Archives of Disease in Childhood*, **41**, 454–471, 613–635 (1966).
16. Gregory, J. R., Collins, D. L., Davies, P. S. W., Hughes, J. M. and Clarke, P. C. *National Diet and Nutrition Survey*: *children aged 1.5 to 4.5 years*: *Volume 1*, HMSO, London, 1995.
17. Lucas, A., Gore, S. M., Cole, T. J., Bamford, M. F., Dossetor, J. F. B., Barr, I., Dicarlo, L., Cork, S. and Lucas, P. J. 'Multicentre trial on feeding low birth weight infants – effects of diet on early growth', *Archives of Disease in Childhood*, **59**, 722–730 (1984).
18. Ratcliffe, S. G., Masera, N., Pan, H. and McKie, M. 'Head circumference and IQ of children with sex chromosome abnormalities', *Developmental Medicine and Child Neurology*, **36**, 533–544 (1994).
19. Rona, R. J. and Chinn, S. 'National Study of Health and Growth: social and biological factors associated with height of children from ethnic groups living in England', *Annals of Human Biology*, **13**, 453–471 (1986).
20. Green, P. J. and Silverman, B. W. *Nonparametric Regression and Generalized Linear Models*, Chapman & Hall, London, 1994.
21. Rao, C. R. 'Some statistical methods for the comparison of growth curves', *Biometrics*, **14**, 1–7 (1958).
22. Payne, R. W. *Genstat 5 Release 3 Reference Manual*, Oxford University Press, Oxford, 1993.
23. Royston, J. P. 'Estimating departure from normality', *Statistics in Medicine*, **10**, 1283–1293 (1991).
24. Fisher, R. A. *Statistical Tables*, 1970.
25. Cole, T. J. 'Do growth chart centiles need a face lift?', *British Medical Journal*, **308**, 641–642 (1994).
26. Hamill, P. V. V., Drizd, T. A., Johnson, C. L., Reed, R. B. and Roche, A. F. *NCHS Growth Curves for Children Birth – 18 years*, National Center for Health Statistics, Washington DC, 1977.
27. Guo, S., Roche, A. F., Baumgartner, R. N., Chumlea, W. C. and Ryan, A. S. 'Kernel regression for smoothing percentile curves: reference data for calf and subscapular skinfold thicknesses in Mexican Americans', *American Journal of Clinical Nutrition*, **51**, 908S–916S (1990).
28. Wright, C. M., Corbett, S. S. and Drewett, R. F. 'Sex differences in weight in infancy and the British 1990 national growth standards', *British Medical Journal*, **313**, 513–514 (1996).
29. Lucas, A., Cole, T. J. and Gandy, G. M. 'Birth weight centiles in preterm infants reappraised', *Early Human Development*, **13**, 313–322 (1986).
30. Yudkin, P. L., Aboualfa, M., Eyre, J. A., Redman, C. W. G. and Wilkinson, A. R. 'New birth-weight and head circumference centiles for gestational ages 24 to 42 weeks', *Early Human Development*, **15**, 45–52 (1987).
31. Healy, M. J. R. 'In the Discussion to Cole (1988)', *Journal of the Royal Statistical Society*, *Series A*, **151**, 411–412 (1988).
32. Bowditch, H. P. 'The growth of children studied by Galton's percentile grades', in *22nd Annual Report of the State Board of Health of Massachusetts*, Wright and Potter, Boston, 1891, pp. 479–525.
33. Boas, F. 'Observations on the growth of children', *Science*, **72**, 44–48 (1930).
34. Cole, T. J. 'Growth monitoring with the British 1990 growth reference', *Archives of Disease in Childhood*, **76**, 1–3 (1997).
35. Cole, T. J. 'Growth charts for both cross-sectional and longitudinal data', *Statistics in Medicine*, **13**, 2477–2492 (1994).
36. Cole, T. J. '3-in-1 weight monitoring chart', *Lancet*, **349**, 102–103 (1997).
37. Pan, H. Q. and Goldstein, H. 'Multilevel models for longitudinal growth norms', *Statistics in Medicine*, **16**, (in press) (1997).

