J. R. Statist. Soc. A (1988)
151, *Part* 3, *pp.* 385–418

Fitting Smoothed Centile Curves to Reference Data

By T. J. COLE†

Medical Research Council, Cambridge, UK

[*Read before the* Royal Statistical Society *on Wednesday April 20th, 1988, the President* Sir John Kingman *in the Chair*]

SUMMARY

A general method is described for fitting smooth centile curves to reference data, based on the power transformation family of Box and Cox. The data are defined by values or ranges of values of the independent variable t, and best fitting powers $\hat{\lambda}_i$ assuming normality are estimated for each group i. Corresponding estimates for the generalized mean and coefficient of variation $\hat{\mu}_i$ and $\hat{\sigma}_i$ are also obtained. The $\hat{\lambda}_i$, $\hat{\mu}_i$ and $\hat{\sigma}_i$ plotted against t_i are fitted by smooth curves $L(t)$, $M(t)$ and $S(t)$ respectively, which together define a smooth curve for the 100αth centile given by

$$C_{100\alpha}(t) = M(t)[1 + L(t)\, S(t)\, z_\alpha]^{1/L(t)},$$

where z_α is the normal equivalent deviate for tail area α. The method is validated by comparison with published growth standards and illustrated on weight and height data in children. A section describing the practical details of the method is also included.

Keywords: ANTHROPOMETRY; BOX–COX TRANSFORMATION; HEIGHT; NORMAL DISTRIBUTION; QUANTILES; SKEWNESS; SMOOTHING; WEIGHT

1. INTRODUCTION

Centile reference charts are used in medicine to observe clinical measurements on individual patients in the context of population values. If the population centile corresponding to the subject's value is atypical this may indicate an underlying pathological condition. The chart can also provide a background to compare the measurement with as it changes with time. Such charts are used widely in paediatrics, for measurements related to growth and development such as anthropometry. They can also be useful to watch changes in say blood biochemistry following a clinical event, e.g. surgery or the onset of acute illness. In this context, the importance of the reference chart lies in seeing whether the subject's measurements cross centile lines with passing time, implying a change in clinical status.

It is important to stress that the reference population used to construct the chart is not viewed here as a 'normal' population—there is no implication that the 50th centile is a 'norm' to which individuals should aspire. This is particularly true of growth charts, say of height or weight, where for example the reference population may be made up of affluent western children whereas the chart will be *used* on children from relatively poor areas of the world. It would be wrong to argue that the western pattern of growth is *optimal*—the high levels of obesity for example in much of the western world make this unlikely—the population should act merely as a disinterested reference against which other individuals or populations can be compared.

The general form of a centile chart is a series of smoothed centile curves, showing how selected centiles for the measurement change when plotted against some

† *Address for correspondence*: MRC Dunn Nutrition Unit, Downhams Lane, Milton Road, Cambridge, CB4 1XJ, UK.

 0035–9238/88/151385 $2.00

independent variable. The independent variable is commonly age or time, and to simplify somewhat it is referred to here as time. The same arguments apply equally to other independent variables. Centiles are usually chosen from a symmetric subset of the 3rd, 5th, 10th, 25th, 50th, 75th, 90th, 95th and 97th. The centile curves are drawn to follow the centiles of the underlying distribution as closely as possible subject to some roughness penalty, thus providing a trade-off between smoothness and goodness of fit.

The observed distribution centiles used in the fitting process are obtained by splitting the population into separate age or time groups. If empirical centiles are used the more extreme are estimated relatively inaccurately, as the centile standard errors increase steeply towards the tails of the distribution. One way round this problem is to fit a theoretical distribution to the data and then to obtain the expected centiles from the known cumulative density function (Healy, 1974). This approach is commonly used for constructing charts of height by age in children, where the distribution of height is close to normal. Thus if for a particular age group the mean and standard deviation of height are ν and ε, the 100αth centile is given by

$$C_{100\alpha} = \nu + \varepsilon z_\alpha \tag{1}$$

where z_α is the normal equivalent deviate corresponding to tail area α.

However, for other commonly used measures of child growth, for example weight, skinfold thickness or limb circumferences, the data are usually more skew than a normal distribution. In this case it is common practice to assume a log-normal distribution, so that, if ν and ε are the mean and standard deviation on the natural logarithmic scale, the 100αth centile is given by

$$C_{100\alpha} = \mu \exp(\varepsilon z_\alpha) \tag{2}$$

where $\mu = \exp \nu$ is the geometric mean of the original measurement.

These two alternatives can be viewed as power transformations of the data, and there is a whole family of such powers. However, it is uncommon to work with powers other than zero or unity, so that for measurements where the skewness is too much for log-normality, e.g. hormones or immunoglobulins in clinical biochemistry, or where the degree of skewness changes with time, it is customary to work with the empirical centiles.

There is no reason in principle why a general power transformation should not be applied to the data, as described by Box and Cox (1964). The maximum likelihood estimate (MLE) for the power, which both minimizes the skewness and optimizes the fit to normality, is ideally suited to the problem of skew data. However, it only operates on individual groups and does not allow directly for time changes in the skewness.

Van't Hof *et al.* (1985), fitting centiles of skinfold thickness by age, extended the Box–Cox method to estimate a different power for each age group. They then drew a smooth curve through the age-specific powers. This curve, in conjunction with corresponding smooth curves for the mean and standard deviation, was used to generate the required set of centiles.

The purpose of this paper is to show that this technique, of using a smoothly varying Box–Cox transformation, is of much wider application than Van't Hof *et al.* (1985) suggest. Not only does it provide a coherent set of smoothed centiles with relatively little computation, but the shape of the power curve (not to be confused with the type II error curve) provides information about the changing skewness of the distribution which is not provided by other methods of centile fitting. In addition the mean and standard deviation curves that the method generates are of direct interest in their own right.

2. METHODS

2.1. *Box–Cox Power Family*

Box and Cox (1964) proposed two alternative families of transformations,

$$y^{(\lambda)} = \begin{cases} (y^\lambda - 1)/\lambda & (\lambda \neq 0) \\ \log y & (\lambda = 0) \end{cases} \tag{3}$$

and

$$y^{(\lambda)} = \begin{cases} [(y+\delta)^\lambda - 1]/\lambda & (\lambda \neq 0) \\ \log(y+\delta) & (\lambda = 0) \end{cases} \tag{4}$$

involving the unknown parameters λ and δ. Models (3) and (4) require that $y > 0$ and $y > -\delta$ respectively. The parameters are chosen to maximize the likelihood of the observed sample $y = \{y_1 \ldots y_n\}$, assuming it to be normally distributed. By introducing the Jacobian of the transformation, Box and Cox (1964) showed that the MLE of λ is that which minimizes the variance of the scaled variable

$$f^{(\lambda)} = y^{(\lambda)}/\dot{y}^{\lambda-1}$$

or

$$f^{(\lambda)} = y^{(\lambda)}/\mathrm{gm}(y+\delta)^{\lambda-1}$$

where $\dot{y}$ and $\mathrm{gm}(\cdot)$ indicate the geometric mean. It is clear that $f^{(\lambda)}$ is in the same dimensions as y whatever the value of λ, so that as λ varies $\mathrm{var}(f^{(\lambda)})$ remains in the same units as $\mathrm{var}(y)$. In addition, since $\mathrm{var}(f^{(\lambda)})$ is the quantity to be minimized, by definition it changes relatively slowly in the region of the minimum, so that slight differences in λ have little effect on the variance.

The log-likelihood is proportional to $-\log[\mathrm{var}(f^{(\lambda)})]$, and this is approximately quadratic in the region of the maximum. Hence it can be computed for a series of values of λ and the MLE $\hat{\lambda}$ obtained from a fitted quadratic curve.

2.2. *Alternative Scaling*

An alternative scaling transform for use with model (3) (but not model (4)) is given by

$$\begin{aligned} g^{(\lambda)} &= f^{(\lambda)}/\dot{y} \\ &= y^{(\lambda)}/\dot{y}^{\lambda}. \end{aligned} \tag{5}$$

In this case $g^{(\lambda)}$ is dimensionless and its standard deviation is analogous to the coefficient of variation of y. The standard deviation of $g^{(1)}$ differs from the coefficient of variation only in the use of the geometric rather than the arithmetic mean in the denominator. Also $\mathrm{var}(g^{(0)})$ is equivalent to $\mathrm{var}(\log y)$. This emphasizes the close link between coefficient of variation and log(standard deviation), terms which are both used here to describe the standard deviation of $g^{(\lambda)}$.

The variables $f^{(\lambda)}$ and $g^{(\lambda)}$ can be used interchangeably in model (3) but not in model (4), since in the latter case the ratio $f^{(\lambda)}/g^{(\lambda)}$ is a function of the unknown δ.

However, in practice model (3) is considerably more useful than model (4), so that the argument presented here concentrates on model (3) scaled as the variable $g^{(\lambda)}$. The advantage of $g^{(\lambda)}$ over $f^{(\lambda)}$ is that for many variables where this technique is of value the standard deviation increases fairly steadily with the mean while the coefficient of variation does not. Hence the coefficient of variation is relatively independent of the mean.

2.3. *L, M and S Curves*

The method requires that y be split into p groups, corresponding to values or ranges of values of time t, mean t_i ($i = 1 \ldots p$). MLEs of λ_i are obtained for each group, and to this end it is easier to work with y^λ than with $g^{(\lambda)}$. Let v and ε be the observed mean and standard deviation of y^λ, assumed to be normally distributed. The median of y^λ is efficiently estimated by v, so that an efficient estimate μ of median y is given by $v^{1/\lambda}$. Similarly the standard deviation σ of $g^{(\lambda)}$ is given by $\varepsilon/\lambda \dot{y}^\lambda$. For the special case $\lambda = 0$, μ is given by exp v while ε and σ coincide.

The MLE of λ is the value which minimizes σ, so σ is obtained for several values of λ and a quadratic in λ fitted to $\log[\mathrm{var}(g^{(\lambda)})]$. The minimum can then be found by interpolation. If the fitted quadratic is given by

$$\log[\mathrm{var}(g^{(\lambda)})] = \alpha + \beta\lambda + \gamma\lambda^2 \tag{6}$$

then $\hat{\lambda} = -\beta/2\gamma$ and the standard error of $\hat{\lambda}$ is $(n\gamma)^{-0.5}$, where n is the sample size.

Fitting the quadratic requires the log-variance for at least three distinct values of λ. In practice the results lie very close to a quadratic curve, so that three values specify it quite adequately. In addition the exact value of $\hat{\lambda}$ is not critical since it is to be smoothed. This means that choosing the values -1, 0 and $+1$ for λ covers a reasonable range and avoids non-integral powers, saving processor time. In this case the values of β and γ are $(V_+ - V_-)/2$ and $(V_- - 2V_0 + V_+)/2$ respectively, where V is the logarithm of the variance and the suffices $-$, 0 and $+$ refer to λ values of -1, 0 and $+1$. These formulae are given in simplified form in Section 2.9.

Now the $\hat{\lambda}_i$, $\hat{\mu}_i$ and $\hat{\sigma}_i$ are plotted in turn against t_i, and smooth curves $L(t)$, $M(t)$ and $S(t)$ are drawn for each, referred to as the power, mean and standard deviation curves. Strictly, the standard deviation curve ought to be fitted to the $\hat{\sigma}_i$ plotted on a logarithmic scale, since this corresponds more closely to the log-likelihood of the sample. In practice it makes little difference.

The smoothing can be done using whatever method is convenient, e.g. cubic splines (Silverman, 1985), kernel methods (Gasser *et al.*, 1984), polynomials, other specifically tailored mathematical functions (Jenss and Bayley, 1937; Preece and Baines, 1978) or simply fitting by eye.

It is possible to do the curve fitting in two stages, first obtaining $L(t)$ and then using $L(t_i)$ for each group rather than $\hat{\lambda}_i$, to obtain $\hat{u}_i$ and $\hat{\sigma}_i$. This way the estimates of $M(t)$ and $S(t)$ are slightly more consistent, but at the cost of extra computation.

The standard deviation obtained from the $S(t)$ curve can be restored to original y units by multiplying by $L(t)\dot{y}_i^{L(t)}$. However, this requires the values of $\dot{y}_i$ to be smoothed in addition to the $\hat{\mu}_i$, to obtain values for $t \neq t_i$. As the two means are in practice very similar, particularly if $L(t_i)$ is close to zero, the method is somewhat simplified if $M(t_i)$ rather than $\dot{y}_i$ is used.

With this simplification, the L, M and S curves can be used to generate *any* smooth centile curve over the whole age range. If the required centile is 100α, then the equation of the centile curve $C_{100\alpha}(t)$ is given by

$$C_{100\alpha}(t) = M(t)[1 + L(t)\,S(t)\,z_\alpha]^{1/L(t)} \tag{7}$$

The equivalent form if $L(t)$ is zero is given by

$$C_{100\alpha}(t) = M(t)\,\exp[S(t)z_\alpha]$$

although it is unlikely for $L(t)$ to be *exactly* zero given that it varies continuously.

2.4. *Testing for Normality*

The method assumes that after power transformation the data are normally distributed. This is tested formally using the Shapiro–Wilk W statistic (Royston, 1982), which is based on the squared correlation coefficient of the ordered sample values plotted against their expected normal order statistics. This is known to be a good omnibus test in the absence of skewness.

2.5. *Multiple Categories and Covariance Adjustment*

Box and Cox (1964) defined $\mathrm{var}(y)$ as the residual variance after fitting a quite general design matrix. There are two particular forms of design matrix which are likely to be useful in this context—one allowing for different categories of individual within each age group, the two sexes for example, and the other allowing for covariates, e.g. age within each age group. For data in different categories, the method provides distinct $M(t)$ curves while pooling the data to estimate $L(t)$ and $S(t)$. This is worthwhile for improving the estimates $\hat{\lambda}_i$ and $\hat{\sigma}_i$ but is only sensible if the L and S curves are likely to be similar in the different categories. The alternative would be to keep the categories separate and to estimate all three curves for each.

Incorporating age as a covariate allows for a linear trend in the mean and so gives a direct estimate of the instantaneous variance (Healy, 1962). The variance to be minimized here is the residual variance after fitting the covariate, and the obtained mean can be adjusted to some central covariate value.

2.6. *Assessment of Individuals*

In the area of nutritional status assessment, it is common practice to express measurements for individual subjects in one of three ways: either as a centile, a standard deviation score or a fraction of the median. The form of equation (7) allows the three ways to be unified, in the following manner. Consider a subject whose observed measurement is Y and whose age (assuming this to be the independent variable) is T. By rearranging equation (7), the corresponding standard deviation score for the subject is given by

$$Z = \frac{[Y/M(T)]^{L(T)} - 1}{L(T)S(T)}. \tag{8}$$

The term $Y/M(T)$ is Y expressed as a fraction of the median, Z is the standard deviation score and the centile corresponding to Z is given by $\Phi(Z)$, the normal probability integral for Z. Thus if, in future, centile chart tabulations include the S

and L curves in addition to the M curve, it will be possible to obtain Z to full accuracy. Conversely if information on L is missing it can be assumed to be either unity or zero, in which case equation (8) simplifies to respectively

$$Z=\frac{Y/M(T)-1}{S(T)} \tag{9}$$

or

$$Z=\frac{\log[Y/M(T)]}{S(T)}. \tag{10}$$

Bearing in mind that $S(T)$ is the coefficient of variation or log(standard deviation), and $Y/M(T)$ has mean 1, equations (9) and (10) are equivalent to the standard deviation score $(Y - \text{mean } Y)/\text{sd}(Y)$. If in addition $S(T)$ is unknown, it is clear that $Y/M(T)$ is the approximate index.

2.7. *Robustness of Power Estimate*

One uncertainty with the method is to what extent outliers influence the estimation of $\hat{\lambda}$. The theoretical advantage of centiles is that they are robust to the presence of outliers, so it is important to know how sensitive $\hat{\lambda}$ is to their presence. Its value is chosen to minimize skewness and to optimize kurtosis, of which only the former is affected to any extent by outliers. Thus it is permissible to trim the data (by deleting the smallest and largest values) and to re-estimate $\hat{\lambda}$. If either of the two deleted values is influential $\hat{\lambda}$ will change appreciably. This trimming may be continued in principle until the proportion of data deleted in each tail approaches the tail area corresponding to the most extreme published centiles (i.e. the 3rd and 97th), and the change in $\hat{\lambda}$ during the process measures the robustness of the procedure. However, the estimates of the variance after trimming will be biased downwards progressively, although this can be adjusted for by fitting a doubly truncated normal distribution (Healy, 1978).

2.8. *Validation*

To test whether existing tabulated centile charts can be expressed in the *LMS* form, i.e. that they are normally distributed on an appropriate power scale, it is necessary to estimate the three curves from tabulated centiles. Consider the set of centiles $C_{100\alpha}$ for one particular age, and assume that when raised to some power λ they approach normality. They are then of the form (1), where ν the mean and ε the standard deviation are unknown constants and the z_α are known. For given λ, the linear form of equation (1) allows ν and ε to be estimated from the regression of $C_{100\alpha}$ on z_α, and so long as the z_α are symmetric, $\hat{\nu}$ is the intercept of the regression equation and $\hat{\varepsilon}$ the regression coefficient.

The method then continues as described in Section 2.3, obtaining μ and σ from ν and ε. L is obtained by interpolation as the value of λ that minimizes σ. M and S are given by μ and σ.

The values of L, M and S at each age make up the three required curves. They are likely to be fairly smooth, particularly the M curve, as a result of the smoothing applied to the chart originally. Substituted into equation (7) they can be used to recreate the original centiles, with which they can be compared.

2.9. *Simplified Description of LMS Method*

The purpose of this section is to provide practical directions for using the method.

For each distinct age or age group, λ is estimated as follows. The standard deviation of the measurement (weight in the examples here) is calculated for three known values of λ. The simplest (and computationally most economical) choices of λ and 1, 0 and -1, corresponding to the transformations weight itself, (natural) log-weight and inverse weight. The geometric mean of weight is also required, which is the antilogarithm (or exponential) of mean log-weight.

The weight standard deviation is divided by the geometric mean weight to give a form of coefficient of variation, while the inverse weight standard deviation is *multiplied* by the geometric mean; the log-weight standard deviation is left unchanged. The three standard deviations, or more accurately coefficients of variation, are now found to be very similar. The aim is to interpolate between them to find the minimum value for the coefficient of variation, and this then corresponds to the best value of λ.

Call the coefficients of variation as obtained from weight, log-weight and inverse weight s_+, s_0 and s_- respectively. The estimate of λ is given by

$$\frac{\log(s_-/s_+)}{2\log(s_- s_+/s_0^2)} \tag{11}$$

and its standard error is

$$[n \log(s_- s_+/s_0^2)]^{-0.5}$$

where n is the sample size.

This process is repeated for each group, and the resulting λ values are plotted against age. A smooth curve L is then drawn through the points, either by computer or by eye, so that λ can be read off the curve at any age.

To find the mean and coefficient of variation for each age group, weight is raised to the power λ as obtained either from the original calculation or as read off the $L(t)$ curve—the choice is unimportant. The mean and standard deviation of the λ-transformed weight is calculated, and this standard deviation is divided both by λ and by the geometric mean of weight raised to the power λ. If the result is negative, make it positive. This is the minimum coefficient of variation, and so should be slightly smaller than s_-, s_0 and s_+. The mean of the transformed weight is back transformed to mean weight by raising it to the power $1/\lambda$.

Just as with the power, the means and coefficients of variation for each age group are plotted against age, and the smooth curves M for the mean and S for the coefficient of variation are obtained. The L, M and S curves can be fitted either by computer or by eye, but the advantage of fitting them by computer is that they can be expressed numerically, which simplifies reading values off the curves subsequently.

Once L, M and S have been obtained they can be substituted into equation (7) for given values of age t and normal equivalent deviate z_α, to obtain a complete set of centile curves. For reference, the values of z_α (the SD score) corresponding to particular centiles ($C_{100\alpha}$) are given in Table 7, later. If L is negative, this must be taken into account in equation (7).

For the more involved cases described in Section 2.5, the residual standard deviations for weight, log-weight and inverse weight can be obtained from analysis of variance or regression analysis. Otherwise the method is as already described.

3. DATA

3.1. *Validation*

The recently published set of 1980 Dutch growth standards (Roede and Van Wieringen, 1985) is used to validate the method. The set of standards is based on a national survey of 41 870 children aged between 3 weeks and 19 years, of whom 8301 were under 1 year old. The weight centiles were obtained empirically and subsequently smoothed by eye, and these are used for validation. The 3rd, 10th, 25th, 50th, 75th, 90th and 97th centiles are tabulated by sex for 26 ages between 3 weeks and 64 weeks, and by half-years from 1 year to 19.5 years. Between 100 and 300 children were seen at each age. The centiles for the first year are used here, with $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}$ being computed for each tabulated week in the two sexes.

3.2. *Two Examples*

Two datasets are used to illustrate the method on raw data. The first is the Cambridge infant growth study (Whitehead *et al.*, 1988), a longitudinal study of 132 children born in the city of Cambridge in 1983–84 and seen every 4 weeks (plus or minus 2 days) between 4 weeks and 52 weeks of age, and also at 78 weeks. Half the cohort was also seen when 2 weeks old. Six measures of anthropometry were recorded at each visit, of which just weight is used here. The values of $\hat{\lambda}_i$ ($i=1 \ldots 15$) were obtained by pooling the two sexes, and $L(t)$ the power curve was fitted as a cubic in weeks of age. The smoothed λ values from $L(t)$ were then used to obtain $\hat{\mu}_i$ (by sex) and $\hat{\sigma}_i$ (pooled). The median curves $M(t)$ for each sex were fitted by the Jenss curve (Jenss and Bayley, 1937)

$$y = a_o + a_1 \mathrm{Age} - \exp(c_0 + c_1 \mathrm{Age})$$

or equivalently

$$y = a_0 + a_1 \mathrm{Age} + a_2 r^{\mathrm{Age}} \tag{12}$$

which has been shown (Berkey, 1982) to provide a good fit to weight data during infancy. The curves were fitted using the OPTIMISE directive of GENSTAT, which exploits the linear form of the latter equation (for given r) in the fitting process. The standard deviation curve $S(t)$ was also summarized as a cubic in age. In fitting the two cubic equations, each polynomial term was included only if it reached significance at $P < 0.05$. (This significance test is slightly dubious owing to the longitudinal nature of the data, but the main concern here is estimation rather than significance). The sexes were pooled for estimating $L(t)$ and $S(t)$ since separate estimations had shown that they were not materially different. Thus the complete weight centile chart for the two sexes was summarized by four four-parameter functions.

The second example comes from the second and third health examination surveys of the USA (National Center for Health Statistics, 1970, 1973). Taken together these provide data on the anthropometry of children between the ages of 6 years and under 18 years, with from 400 to 600 children of each sex seen in each year of age. Here the weights and heights of boys were analysed, by single years of age, and within each group age was used as covariate to obtain the residual variance; the group means were adjusted to age mid-year. The height powers showed a clear linear trend, and the smoothed values were used to estimate $M(t)$ and $S(t)$. Apart from this smooth curves were not fitted to the data.

4. RESULTS

4.1. *1980 National Survey of Dutch Infants*

Figs 1 and 2 give the Box–Cox power estimates $\hat{\lambda}$ and the corresponding coefficients of variation $\hat{\sigma}$ by sex and week of age for the Dutch infant weight centiles. The power changes relatively little during the year, and the values for the sexes agree to within 0.2. The trends in the coefficient of variation are also similar between the sexes, and fall monotonically with age except for the boys between 24 weeks and 36 weeks. The values of $\hat{\mu}$ obtained for the boys of each age are shown as the median curve in Fig. 3 along with the constructed curves for the other six centiles. Fig. 3 also shows the boys' centiles as published, and the two sets of curves are clearly in close agreement. Table 1 summarizes the magnitude of the discrepancies for each centile. The largest bias, 0.28 %, occurs on the 90th centile, while the standard deviations of the discrepancies vary between 0.15 % and 0.47 %. The absolute departures range between -0.9 % and $+1.0$ %, and 87 % of the $26 \times 7 \times 2 = 364$ points are within $+0.5$ %.

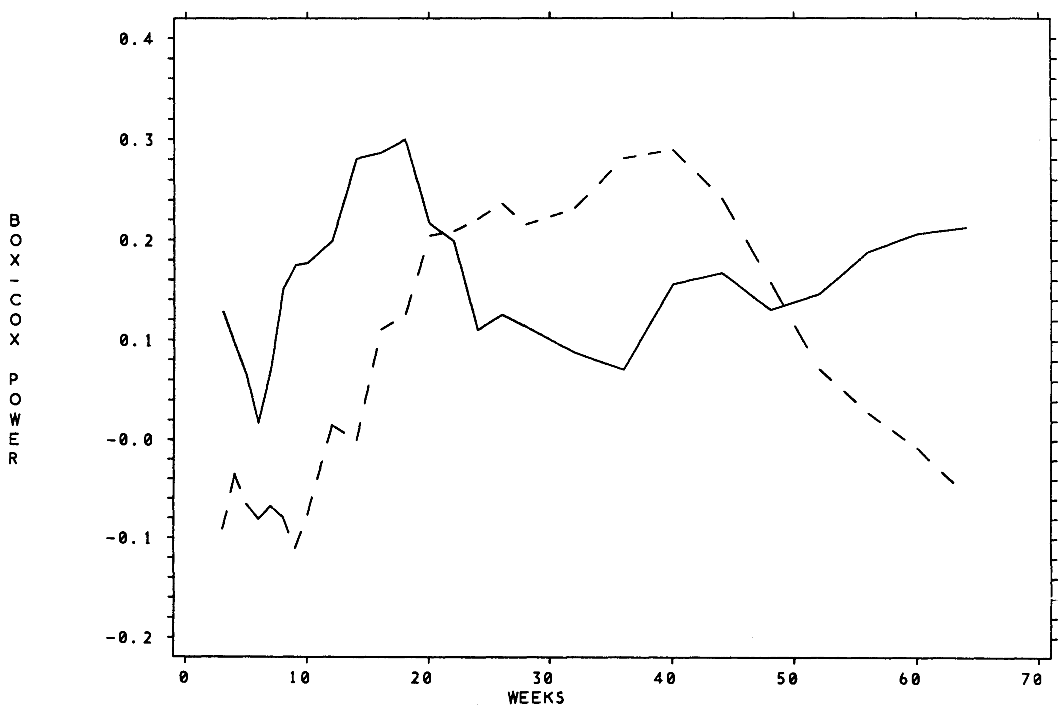

Fig. 1. 1980 Dutch national survey infants' weight: power curves $L(t)$ for boys (full curve) and girls (broken curve) showing the change in Box–Cox power during the first year

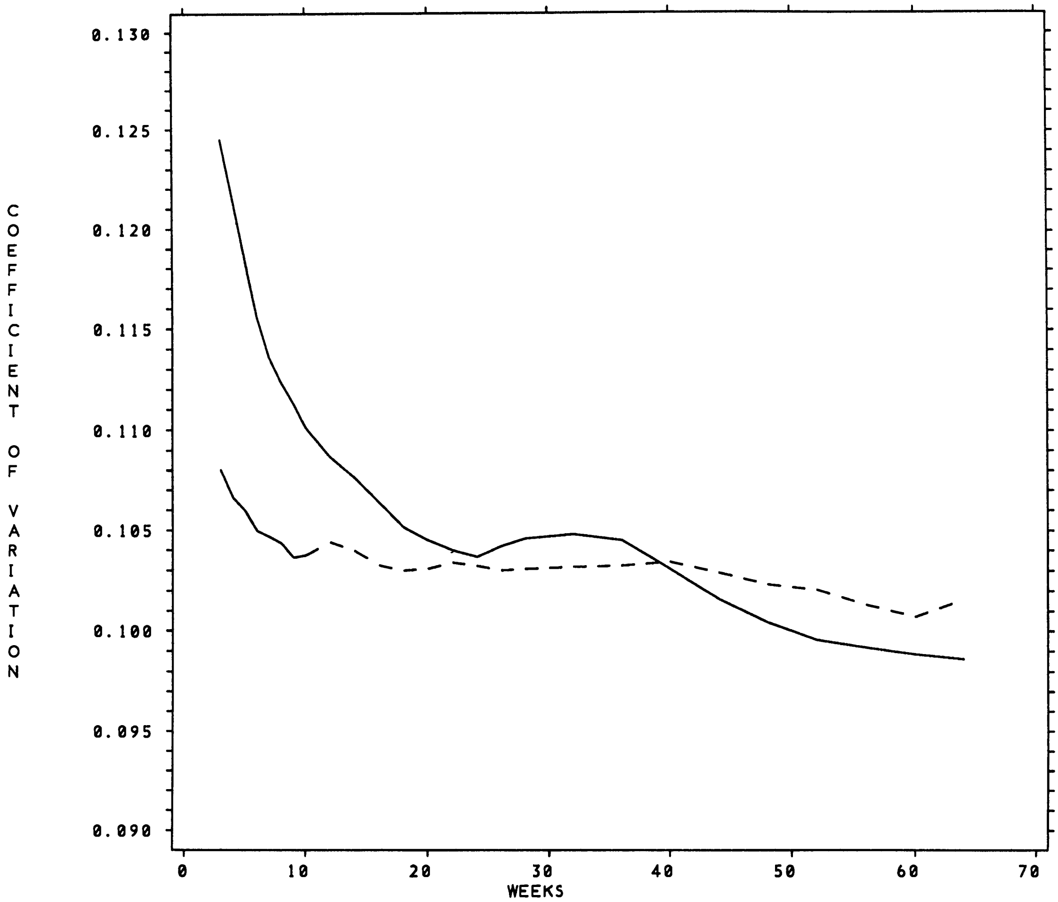

Fig. 2. 1980 Dutch national survey infants' weight: coefficient of variation curves $S(t)$ for boys (full curve) and girls (broken curve) showing the change in coefficient of variation during the first year

4.2 *Cambridge Infant Growth Study*

Table 2 gives the results by week of age for weight in the Cambridge infant growth study. Shown are the best fitting power $\hat{\lambda}$ and the coefficient of variation $\hat{\sigma}$ for the sexes combined, and the mean weights for the two sexes. In addition the corresponding smoothed estimates are shown. The standard errors of $\hat{\lambda}$ are about 0.6. Table 3 gives the coefficients of the Jenss curves $M(t)$ (12) by sex, and the cubic curves $L(t)$ and $S(t)$ for power and coefficient of variation. Figs 4 and 5 respectively show $\hat{\lambda}$ and $\hat{\sigma}$ with the fitted cubic curves, where the fitted curves follow the data quite reasonably. The power curve covers a much wider range of values than the Dutch survey (Fig. 1), but the coefficients of variation in the two studies are very similar. Fig. 6 shows the constructed weight centile chart for boys, where the change in skewness from 2 weeks to 78 weeks is clearly seen.

Table 4 shows the effect on $\hat{\lambda}$ of trimming the data, one pair at a time. After 16 weeks of age, deleting the first pair increases $\hat{\lambda}$ by about 0.3, which shows that the

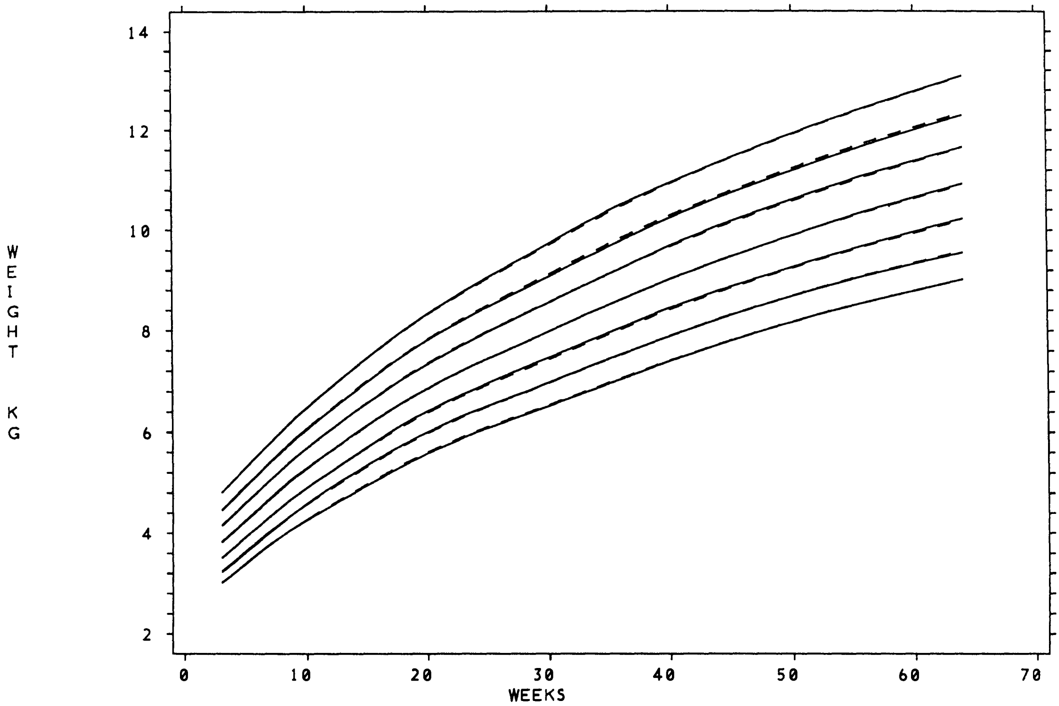

Fig. 3. 1980 Dutch national survey infants' weight: 3rd, 10th, 25th, 50th, 75th, 90th and 97th centiles for boys as published (Roede and Van Wieringen, 1985) (full curves) and as calculated from the curves $L(t)$, $M(t)$ and $S(t)$ using equation (7) (broken curves): the calculated 50th centile is also the $M(t)$ curve

heaviest subject skews the distribution to the right. Deleting the second pair has little further effect on $\hat{\lambda}$.

4.3. *US Health Examination Surveys*

Table 5 gives the results by year of age for boys' weight in health examination surveys 2 and 3. $\hat{\lambda}$ is close to -1 until age 11, from 12 to 16 it increases to about -0.5, and then at 17 it drops back towards -1. The standard errors are generally less than 0.2,

TABLE 1

1980 Dutch national survey infants' weight: percentage differences between standard weights as tabulated by Roede and Van Wieringen (1985) and as predicted by equation (7), averaged over age and sex†

	Centile						
	3	*10*	*25*	*50*	*75*	*90*	*97*
Mean (%)	−0.14	0.04	0.12	0.10	0.07	−0.28	0.08
Standard deviation (%)	0.38	0.47	0.34	0.25	0.25	0.19	0.15

† The curves $L(t)$, $S(t)$ and $M(t)$ in equation (7) are shown in Figs 1–3.

TABLE 2

Cambridge infant growth study weight: results by weeks of age for the power and coefficient of variation, sexes combined, and for the mean, sexes separate†

Weeks						*Boys*		*Girls*	
	n	$\hat{\lambda}$	$L(t)$	$\hat{\sigma}$	$S(t)$	$\hat{\mu}$	$M(t)$	$\hat{\mu}$	$M(t)$
2	62	0.79	1.08	0.130	0.130	3.65	3.68	3.60	3.64
4	130	1.02	0.75	0.131	0.126	4.22	4.21	4.10	4.07
8	131	0.26	0.18	0.117	0.121	5.20	5.15	4.89	4.85
12	132	−0.16	−0.29	0.114	0.116	5.92	5.93	5.48	5.51
16	131	−0.76	−0.66	0.109	0.113	6.56	6.58	6.06	6.08
20	129	−1.15	−0.95	0.112	0.110	7.11	7.13	6.57	6.58
24	129	−1.13	−1.15	0.109	0.108	7.59	7.60	7.03	7.02
28	125	−1.20	−1.29	0.109	0.107	8.02	8.02	7.39	7.41
32	129	−1.29	−1.37	0.107	0.106	8.39	8.38	7.76	7.75
36	129	−1.48	−1.39	0.106	0.105	8.71	8.71	8.07	8.06
40	128	−1.57	−1.37	0.102	0.105	9.06	9.01	8.38	8.35
44	127	−1.12	−1.32	0.105	0.104	9.26	9.28	8.60	8.60
48	124	−1.22	−1.24	0.106	0.104	9.51	9.53	8.83	8.84
52	126	−1.17	−1.15	0.101	0.104	9.78	9.77	9.05	9.07
78	104	−0.77	−0.76	0.096	0.096	11.13	11.13	10.28	10.28

†In each case the observed and fitted values are shown; the fit is by cubic curve for the power and coefficient of variation, and by Jenss curve for the means.

TABLE 3

Cambridge infant growth study weight: coefficients of the fitted Jenss curves for the mean, sexes separate, and the fitted cubic curves for the power and coefficient of variation, sexes combined†

		$a_0 + a_1\text{Age} + a_2 r^{\text{Age}}$		
$M(t)$	a_0	a_1	a_2	r
Boys	7.59	0.046	−4.5	0.943
Girls	7.50	0.037	−4.3	0.953
		$a_0 + a_1\text{Age} + a_2\text{Age}^2 + a_3\text{Age}^3$		
	a_0	a_1	a_2	a_3
$\log S(t)^2$	−4.03	−0.030	0.00063	−0.0000045
$L(t)$	1.44	−0.187	0.00383	−0.000023

†Age is in units of weeks.

so that in every age group $\hat{\lambda}$ is highly significantly less than zero. The period 12–16 covers the boys' adolescent growth spurt, which influences the value of the coefficient of variation $\hat{\sigma}$ as well as $\hat{\lambda}$. $\hat{\sigma}$ rises steadily from a value of 0.14 at age 6 to 0.22 at age 13, dropping back to 0.16 at age 17. None of the 12 age groups is significantly non-normal by the Shapiro–Wilk test after transformation ($P > 0.1$), while *all* are when untransformed ($P < 0.001$), and all but three are with a logarithmic transform ($P < 0.05$). Fig. 7 illustrates the SD scores on the transformed scale corresponding to the standard centile points at each age. The agreement is good on the whole, and better in the middle than in the tails.

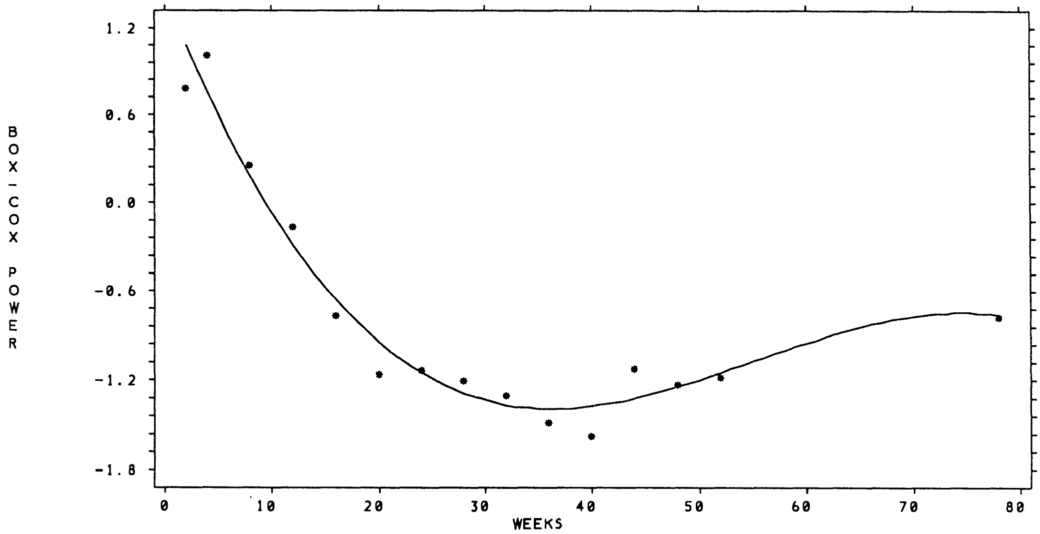

Fig. 4. Cambridge infant growth study weight: estimates of the Box–Cox power at each age with the fitted cubic curve, sexes pooled

The corresponding results for height are in Table 6 and Fig. 8. Here $\hat{\lambda}$ is more variable from age to age, with standard errors four times greater than for weight. The mean value of $\hat{\lambda}$ is 1.1, but there is a highly significant positive trend with age (regression coefficient 0.28 SE $0.056y^{-1}$, $P < 0.001$). Even so, only in the oldest age group is $\hat{\lambda}$ significantly different from unity ($P < 0.05$). The value of $\hat{\sigma}$ is about one-quarter that for weight, but as with weight it increases with age from 6 to 13 and then decreases

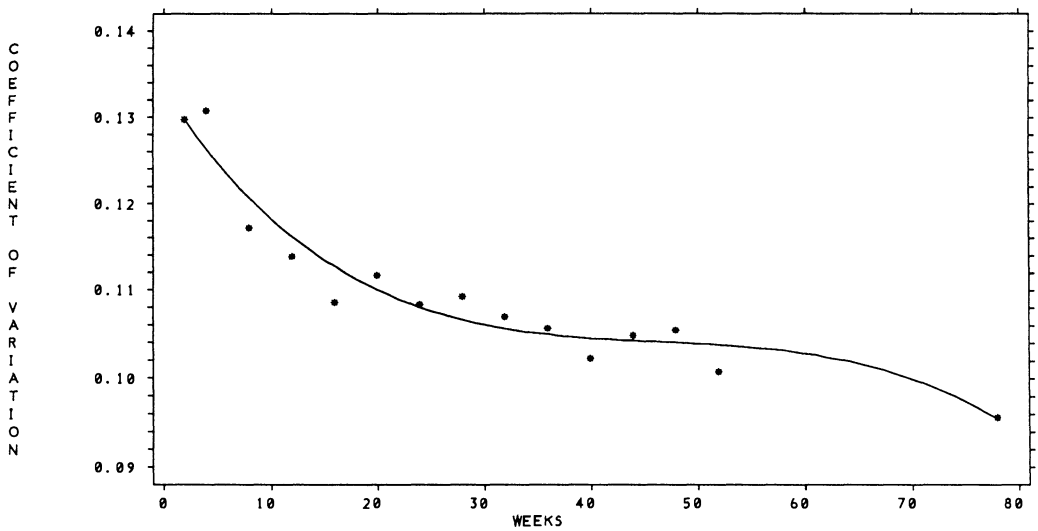

Fig. 5. Cambridge infant growth study weight: estimates of the coefficient of variation at each age with the fitted cubic curve, sexes pooled

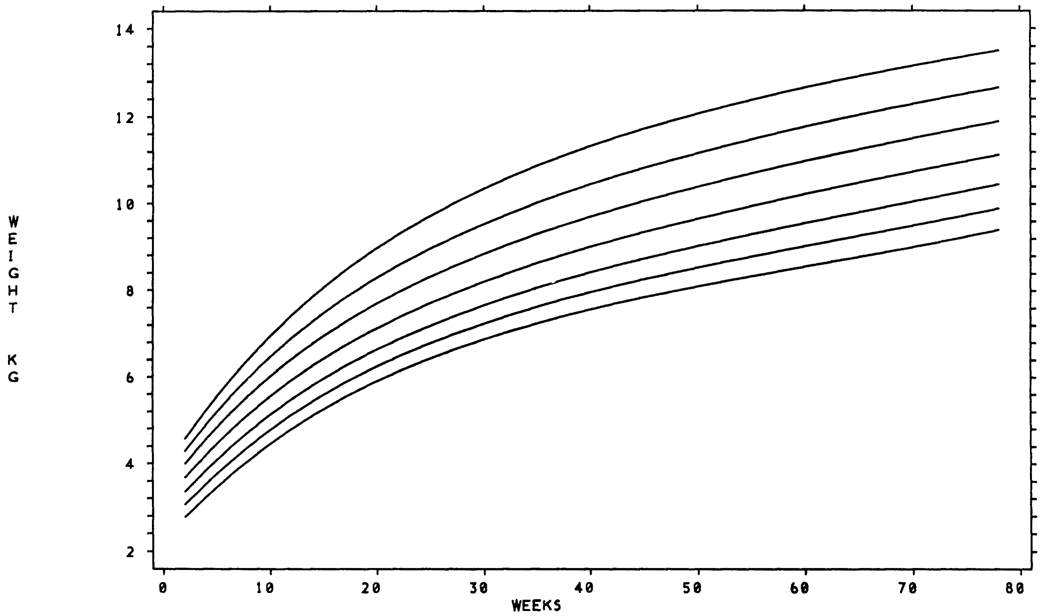

Fig. 6. Cambridge infant growth study weight: fitted centile curves for boys as derived from equation (7): $L(t)$ and $S(t)$ are shown in Figs 4 and 5, while $M(t)$ is shown here as the 50th centile

TABLE 4

Cambridge infant growth study weight: change in the estimated power relative to the original value after deleting the largest and smallest value†

Weeks	*Pair* 1	2
2	0.00	−0.19
4	−0.04	−0.15
8	−0.59	−0.81
12	−0.06	−0.32
16	−0.11	−0.03
20	+0.27	+0.32
24	+0.28	+0.28
28	+0.45	+0.46
32	+0.26	+0.52
36	+0.17	+0.38
40	+0.33	+0.49
44	+0.35	+0.29
48	+0.22	+0.29
52	+0.31	+0.46
78	+0.48	+0.59

†Results are shown after deleting one and two such pairs.

TABLE 5

Health examination surveys 2 and 3, boys' weight: results by year of age for the number, power, mean, coefficient of variation and Shapiro–Wilk W statistic

Age	*n*	λ	*SE*	μ	σ	*W*
6–	580	−1.14	0.18	21.6	0.140	0.9858
7–	631	−1.22	0.17	24.1	0.142	0.9898
8–	622	−0.85	0.17	27.1	0.159	0.9930
9–	599	−1.17	0.14	29.9	0.183	0.9824
10–	575	−1.06	0.17	32.8	0.172	0.9913
11–	620	−1.00	0.15	37.1	0.189	0.9902
12–	641	−0.70	0.16	41.2	0.204	0.9852
13–	626	−0.51	0.14	48.2	0.218	0.9835
14–	617	−0.51	0.15	54.9	0.205	0.9861
15–	614	−0.60	0.17	60.1	0.177	0.9903
16–	555	−0.56	0.17	63.5	0.169	0.9885
17–	433	−0.79	0.23	66.7	0.160	0.9876

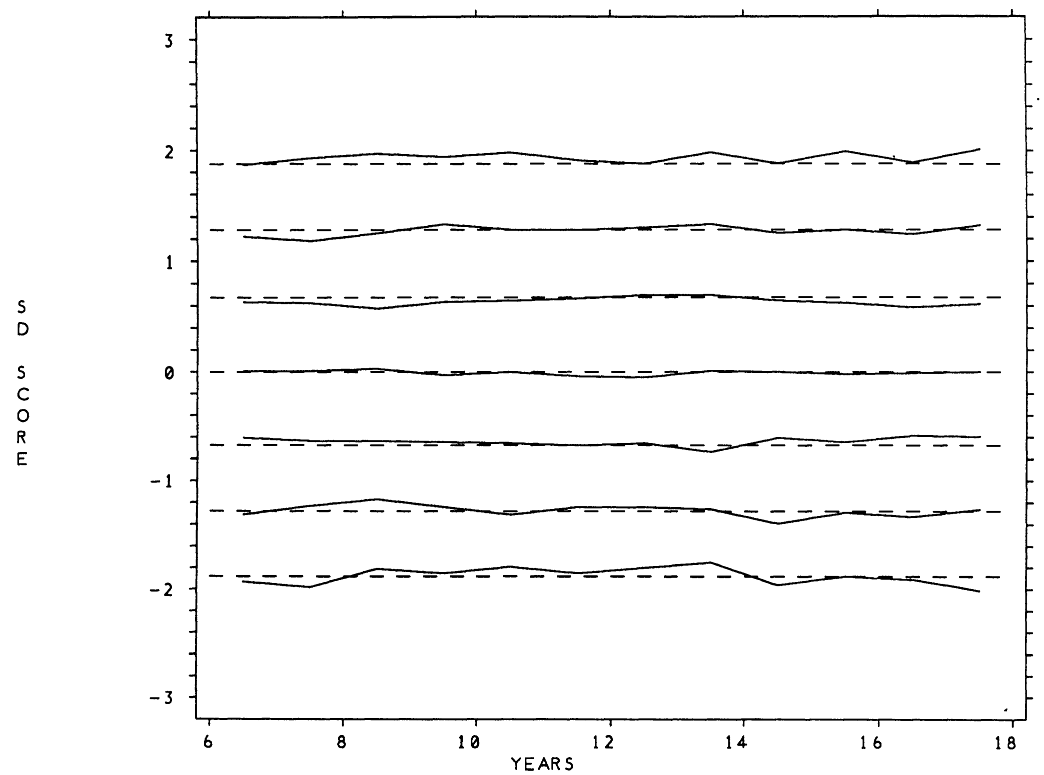

Fig. 7. Health examination surveys 2 and 3, boys' weight: SD scores corresponding to specified centiles transformed according to equation (8) (full curves) plotted against age, with the normal SD scores for the same centiles shown as broken curves

TABLE 6

Health examination surveys 2 and 3, boys' height: results by year of age for the power, linearly smoothed power, mean, coefficient of variation and Shapiro–Wilk W statistic

Age	λ	*SE*	*L*	μ	σ	*W*
6–	−1.1	0.8	−0.47	1.19	0.0414	0.9839
7–	−0.1	0.7	−0.18	1.24	0.0415	0.9832
8–	1.2	0.7	0.10	1.30	0.0428	0.9908
9–	0.2	0.7	0.39	1.36	0.0462	0.9890
10–	0.8	0.7	0.67	1.40	0.0452	0.9814
11–	1.4	0.7	0.95	1.46	0.0455	0.9887
12–	0.0	0.6	1.24	1.52	0.0522	0.9882
13–	1.1	0.6	1.52	1.60	0.0549	0.9836
14–	2.7	0.7	1.81	1.67	0.0502	0.9835
15–	1.9	0.7	2.09	1.72	0.0432	0.9845
16–	2.0	0.8	2.37	1.74	0.0402	0.9904
17–	3.1	1.0	2.66	1.76	0.0402	0.9812

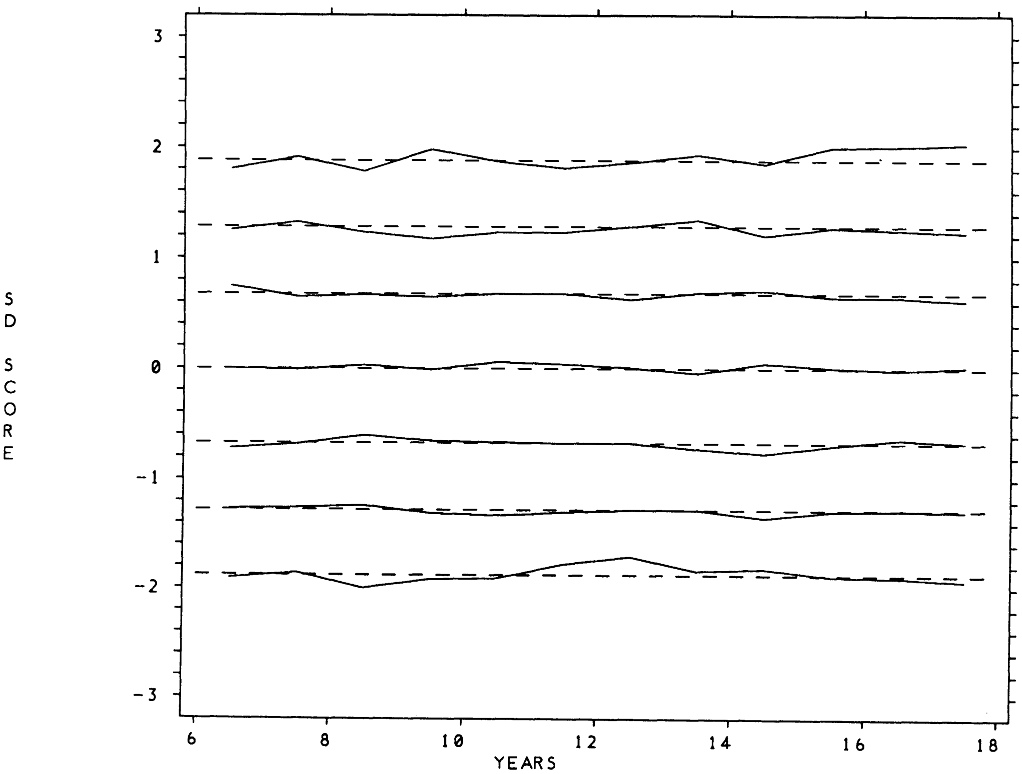

Fig. 8. Health examination surveys 2 and 3, boys' height: SD scores corresponding to specified centiles transformed according to equation (8) (full curves) plotted against age, with the normal SD scores for the same centiles shown as broken curves

again. Also as with weight none of the age groups is significantly non-normal after transformation ($P > 0.1$), while the oldest group is significant when untransformed ($P < 0.05$). Fig. 8 gives the transformed SD scores at each age, using the linear fitted powers.

Table 7 summarizes the results over age for the SD scores in Figs 7 and 8. The means and standard deviations of the discrepancies in SD score relative to the expected values are shown for each centile, with the expected SD score shown for reference. The largest bias occurs for 97th centile weight where the mean observed SD score is 1.937, which corresponds to an expected centile of 97.3. The other SD score means are only trivially different from their expected values. The variation in SD score is

TABLE 7
Health examination surveys 2 and 3, boys' weight and height: differences in SD score (observed less expected) for the seven centiles shown in Figs 7 and 8, averaged over age

				Centile			
	3	*10*	*25*	*50*	*75*	*90*	*97*
Expected SD score	−1.881	−1.282	−0.674	0.000	0.674	1.282	1.881
Weight							
Mean	0.003	0.008	0.036	−0.008	−0.038	−0.008	0.056
Standard deviation	0.082	0.057	0.041	0.023	0.037	0.046	0.049
Height							
Mean	0.008	−0.010	−0.006	0.013	−0.011	−0.030	0.023
Standard deviation	0.077	0.030	0.043	0.029	0.036	0.047	0.087

TABLE 8
Health examination surveys 2 and 3, boys' weight and height: change in the estimated power relative to the original value for weight and height after deleting the largest and smallest value†

Age	*Weight*				*Height*			
	Pair 1	*2*	*4*	*8*	*Pair 1*	*2*	*4*	*8*
6–	+0.05	+0.08	+0.16	+0.32	+0.15	+0.18	+0.29	+0.48
7–	−0.06	−0.03	+0.04	+0.15	+0.03	+0.09	−0.18	−0.54
8–	−0.04	−0.09	−0.15	−0.14	0.00	−0.07	+0.02	−0.37
9–	−0.03	−0.04	−0.14	−0.12	−0.15	−0.05	+0.16	+0.47
10–	−0.05	−0.04	−0.04	−0.01	−0.01	+0.10	+0.25	+0.33
11–	−0.09	−0.11	−0.16	−0.17	−0.17	−0.12	−0.26	−0.50
12–	0.00	0.00	−0.01	−0.04	+0.01	+0.02	+0.06	−0.09
13–	−0.02	+0.01	+0.07	+0.06	0.00	−0.03	−0.06	−0.27
14–	+0.05	+0.10	+0.18	+0.26	+0.02	−0.06	−0.08	−0.01
15–	−0.02	0.00	+0.01	+0.03	−0.18	−0.22	−0.05	+0.31
16–	−0.04	−0.05	0.00	+0.08	+0.06	+0.20	+0.04	−0.42
17–	−0.04	+0.01	+0.06	+0.06	+0.02	−0.06	−0.17	−0.23

†Results are shown after deleting one, two, four and eight such pairs, by age group.

least on the median and increases towards the tails of the distribution—height and weight are similar in this respect.

Table 8 shows how insensitive the estimates of λ are to the extreme values in each tail of the weight and height distributions, by age group. In most cases the effect of excluding the largest and smallest value is trivial, with $\hat{\lambda}$ changing by less than 0.1. Excluding further pairs has progressively more effect, but only rarely does this exceed the standard error of the original estimate. Also of interest, but not shown, is the way that the standard errors of $\hat{\lambda}$ increase steeply as the data are trimmed. Thus the choice of $\hat{\lambda}$ is determined largely in the extreme tails of the distribution. This is more labile for height than for weight, as might be expected given the larger standard errors for height.

5. DISCUSSION

Producing centile charts has always been something of a black art—the centile lines need to be drawn such that they are both smooth and close to the empirical centiles. It is not surprising that this trade-off problem is often solved by drawing the lines by eye. Roede and Van Wieringen (1985) have described in detail their sequential procedure for fitting curves to observed weight centiles. Essentially it involves first obtaining a smoothed median curve and then expressing the other observed centiles relative to the smoothed median. Smooth curves are then drawn for the other centiles, and these are adjusted in the light of the original centiles.

For height they use a simpler method; normality is assumed initially, and smoothed curves representing the mean and standard deviation are obtained. These are then used to obtain the centile curves. Comparison of the curves with the observed centiles shows close agreement except during puberty, where the smoothed curves are modified to take account of the skewness.

Tanner *et al.* (1966) assumed a normal distribution for height throughout childhood, although they acknowledged the existence of some skewness during puberty. For weight they drew the centile curves directly from the observed centiles. In both cases they 'shrank' the centiles towards the mean to compensate for increased variation due to grouping the data (Healy, 1962).

Hamill *et al.* (1977) experimented with the Pearson family of curves to represent the distribution at each age; in this family the shape of the distribution is defined by the mean, standard deviation, skewness and kurtosis. They used polynomial stepwise regression to smooth the age changes in these four quantities, a method analogous to that described here. However, they found the polynomial curves too inflexible, and so fitted cubic splines to the individual centile curves instead. This approach led to difficulties with spacing the centile curves appropriately.

A completely different approach to the problem has been provided by Healy *et al.* (1988), who propose an entirely nonparametric method of centile curve fitting. Firstly the centiles are smoothed nonparametrically, and then a set of polynomials whose coefficients are constrained to be linearly related across centiles are used to fit the separate centiles. This powerful method has much in common with the *LMS* method in that it produces centiles which are

(a) smooth,
(b) close to the data and
(c) constrained to accord with neighbouring centiles.

What it does not have is an underlying distributional form to convert centiles to SD scores.

The present method is a logical extension of the height technique used by Roede and Van Wieringen (1985), which in turn extends the methods of Tanner *et al.* (1966). Smooth curves are obtained for the mean (M) and coefficient of variation (S), but after applying a smoothed Box–Cox power transform (L). This approach provides systematic adjustments to the centile curves to cope with changes in skewness, for example during puberty. It is also a maximum likelihood procedure under normality assumptions, in that the log-variance is minimized at each age subject to mild smoothness constraints on the three curves. The total likelihood for the sample is represented by minus the area under the logarithm of the S curve.

The validation study on the Dutch national survey (Figs 1–3) provides strong support for the method. The close agreement between the published centiles and those expected from a transformed normal distribution shows that Roede and Van Wieringen (1985) could have used this method and still derived essentially the same answer. As a bonus, the method shows the changing skewness and coefficient of variation of the population during the first year of life (Figs 1 and 2).

Working with data rather than tabulated centiles gives a better idea of the inherent variability of anthropometry. Weight is often thought to be poorly behaved distributionally, and the present health examination survey results confirm that, either untransformed or after logarithmic transformation, it is far from normally distributed. However, using a more extreme power transform, with powers in the range from -1.2 to -0.5, the distribution becomes insignificantly different from normal in every age group. Also the values of $\hat{\lambda}$ obtained are well behaved from year to year, showing markedly less skewness during puberty than before or after. The fact that *none* of the age groups produces a significant W statistic is somewhat surprising, but the choice of $\hat{\lambda}$ has ensured the effective absence of skewness, and this may account for the goodness of fit.

For height the story is somewhat different, in that the distribution is fairly close to normal even before transformation, and the appropriate power transform is poorly specified. Both these factors are due to the relatively small size of the height coefficient of variation. Even so the American health examination survey results demonstrate a clear negative trend in skewness across the age range, with values of 2 or more for $\hat{\lambda}$ in the four oldest age groups. Thus the method has detected a shift in the height distribution which has not previously been seen.

Examination of the SD score plots (Figs 7 and 8) shows that for both weight and height there is good agreement between the empirical centiles and those estimated from λ, μ and σ. As might be expected the fit is better for the median and quartiles than for the more extreme centiles, but even in the tails the departures are not large. In addition they tend not to show consistent trends. The standard deviation of the SD scores for each centile (Table 7), multiplied by the corresponding coefficient of variation, gives the percentage variation in weight or height between observed and fitted centiles. The standard deviations in Table 7 range from about 0.03 to 0.08 on the extreme centiles, while typical values for σ in Tables 5 and 6 are 0.18 and 0.045 for weight and height respectively. Thus there is about 1.5% uncertainty for weight and 0.4% for height on the 3rd and 97th centiles, and about half this on the other centiles.

There may, however, be situations where certain of the observed and fitted centiles show clear systematic departures. If so, the method can be extended to provide adjustments to the fitted centiles, as in the final stage of Roede and Van Wieringen's method (1985)—the method of Healy *et al.* (1988) would be a good way to do this. The only disadvantage of second-order smoothing is that the centiles cease to be normally distributed. Seen in this light, the method provides a framework of centile curves which are adequate as they stand for many purposes, but which can be adjusted if necessary.

It is worth emphasizing here that the smooth L, M and S curves at the heart of the method can themselves be drawn by eye, if other techniques are insufficiently sensitive. Thus the method does not necessarily replace the skill that is required to draw centile lines by eye: it simply provides a suitable launching point.

A requirement of the method is that the data be divided into separate groups. This is necessary to obtain the distinct estimates of λ, μ and σ required to specify $L(t)$, $M(t)$ and $S(t)$, but how to choose the number and size of the groups is not so clear. Successive values of λ, μ and σ from one group to the next provide estimates of the linear trends $d\lambda/dt$, $d\mu/dt$ and $d\sigma/dt$, but they give no information on quadratic or higher order trends. Thus the groups should be sufficiently narrow for trends across them to be essentially linear. If age is used as a covariate this provides extra within-group information on the velocity $d\mu/dt$, which could in principle be combined with the neighbouring group means to fit $M(t)$.

Goldstein (1978) has argued that the sampling fraction by age t should be proportional to $d\mu/dt$, allowing equal numbers per expected increment in the measurement. This sampling scheme would fit in well with the present results, since at the time when $d\mu/dt$ is largest, i.e. during puberty, λ and σ are also changing relatively rapidly.

The extended Box–Cox transformation (4) involving δ has not been mentioned, and it might be thought useful for improving the fit in marginal cases. In practice it adds little to the simpler family (3), because although the offset δ is significantly different from zero in certain age groups it makes no difference to the goodness of fit. The effect of δ is to adjust the kurtosis of the distribution, whereas the lack of fit is usually due more to (say) the 90th and 97th centiles being too close together. Even if adding δ to the method *was* thought to be desirable it would mean smoothing a fourth curve, which would be a substantial cost for little benefit.

The results of the Cambridge infant growth study suggest that relatively small longitudinal studies can provide well-defined estimates for the L, M and S curves. The method largely side-steps the difficulties of estimating the centiles, so that the question of whether or not the sample size is adequate hinges on the smoothness of the computed values of λ, μ and σ plotted against age. If there is too much noise, the corresponding fitted curves will be poorly specified, and the resulting centile curves as well.

This highlights the strength of a longitudinal study over a cross-sectional survey—age changes are relatively smooth since the subjects themselves do not change. Thus the estimates of λ in the Cambridge study, based on 130 subjects, show as much consistency from age to age as the health examination survey results where $n = 600$. Admittedly the standard errors are larger and the age intervals between measurements smaller in the Cambridge study, but then growth patterns also change faster in the first year.

A problem with small samples is that $\hat{\lambda}$ may be sensitive to extreme values. A distinction is drawn here between genuine outliers which clearly should be excluded and values in the tails of the distribution which, although unusual, are genuine. The problem is less severe in large studies, e.g. the health examination surveys, where 18 points (3% of the sample) lie beyond the 3rd or 97th centile. In the Cambridge study there are only three or four such points, and their presence or absence makes more difference. There was one very fat boy in the Cambridge sample, and excluding his data increased $\hat{\lambda}$ by 0.3 (Table 4). In contrast, the value of $\hat{\lambda}$ in the health examination surveys was quite insensitive to extreme values (Table 8). If in a large sample there *was* evidence of an outlier, it would be reasonable to exclude it on the grounds that centiles obtained by the conventional way ignore the extreme 3% of the distribution anyway.

The main uncertainty of centile charts based on small longitudinal studies is to do with their representativeness—how relevant are they to larger populations? This can be judged to some extent by comparing the L, M and S curves with those of other surveys. In particular, the Cambridge study can be compared with the Dutch growth standard. Judging from Figs 1–6, their mean and coefficient of variation curves are very similar while the power curves are rather different.

This discrepancy does not mean that the Cambridge power curve is in any sense *wrong*: it simply records that the two samples are structured differently, with the very heavy babies more dispersed in Cambridge. Even excluding Cambridge's outlying fat baby does not alter this—between 16 weeks and 52 weeks the power remains near -1 while in the Dutch survey it is consistently above zero. This emphasizes that, unlike the M and S curves, the L curve is specific to its own sample, acting almost as a sample fingerprint.

Thus the precise value of the L curve, and with it the relative positions of the extreme centiles, might be considered unrepresentative in small samples. The M and S curves though ought to be adequately robust. With this proviso, the Cambridge study centile curves provide a convincing solution to the problem of obtaining growth charts from limited data.

The presentation of centile charts generally is simplified because the tabulated L, M and S curves (or preferably their mathematical functions) enable *any* centile to be calculated at *any* age. Equally the centile can be expressed with full accuracy as a standard deviation score. This is of particular use in the tails of the distribution, where individuals falling below say the 3rd centile are specified only imprecisely in centile terms, whereas their standard deviation score is fully informative. To highlight the benefit this brings, the American National Center for Health Statistics' growth standard (Hamill *et al.*, 1977) had to be published in two forms: centiles for use in America and the western world, and SD scores for the third world. In addition, to calculate individual SD scores they had to use two estimates of the standard deviation, one based on the upper centiles and one on the lower. The present method makes both these measures unnecessary.

The idea of using a smoothly changing Box–Cox transformation was originally described by Van't Hof *et al.* (1985), who demonstrated it on tricep skinfold data from the Nijmegen growth study. They concentrated on minimizing the skewness, and rather dismissed the importance of the smoothed mean and standard deviation curves that are an integral part of the method. This paper differs from theirs both in emphasizing the maximum likelihood nature of the fit and in working with the coefficient of variation rather than the standard deviation. This latter approach is firmly supported by all three examples, in that the age trends in the mean and the coefficient of variation are seen to be clearly different, which would not be the case for the mean and standard deviation. In addition the coefficient of variation is relatively constant, varying for example between 0.14 and 0.22 in Table 5 when the mean changes by a factor of 3. Thus the mean and coefficient of variation curves are to a large extent independent of each other. Van't Hof *et al.* (1985) felt that the technique was valid only on relatively noisy anthropometry data and played down its use in other contexts. The results of this paper suggest that their initial impression was wrong, and that the method is useful for dealing with anthropometry in all its forms.

ACKNOWLEDGEMENTS

I am grateful to David Brown, Patrick Royston and the referees for their helpful comments. I also thank the National Center for Health Statistics for providing the data from health examination surveys 2 and 3.

REFERENCES

Berkey, C. S. (1982) Comparison of two longitudinal growth models for preschool children. *Biometrics*, **38**, 221–234.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *J. R. Statist. Soc.* B, **26**, 211–252.

Gasser, T., Muller, H. G., Kohler, W., Molinari, L. and Prader, A. (1984) Nonparametric regression analysis of growth curves. *Ann. Statist.*, **12**, 210–229.

Goldstein, H. (1978) Sampling for growth studies. In *Human Growth* (eds F. Falkner and J. M. Tanner), vol. I. Oxford: Plenum.

Hamill, P. V. V., Drizd, T. A., Johnson, C. L., Reed, R. B. and Roche, A. F. (1977) NCHS growth curves for children birth—18 years. *Vital and Health Statistics*, ser. 11, no. 165. Washington DC: Health Resources Administration, United States Government Printing Office.

Healy, M. J. R. (1962) The effect of age-grouping on the distribution of a measurement affected by growth. *Amer. J. Phys. Anth.*, **20**, 49–50.

———(1974) Notes on the statistics of growth standards. *Ann. Hum. Biol.*, **1**, 41–46.

———(1978) A mean difference estimator of standard deviation in symmetrically censored normal samples. *Biometrika*, **65**, 643–646.

Healy, M. J. R., Rasbash, J. and Min Yang (1988) Distribution-free estimation of age-related centiles. *Ann. Hum. Biol.*, **15**, 17–22.

Jenss, R. M. and Bayley, N. (1937) A mathematical method for studying growth in children. *Hum. Biol.*, **9**, 556–563.

National Center for Health Statistics (1970) Height and weight of children 6–11 years by age, sex, race, and geographic region, United States. *Vital and Health Statistics*, ser. 11, no. 104. Washington DC: Health Services and Mental Health Administration, United States Government Printing Office.

———(1973) Height and weight of youths 12–17 years, United States. *Vital and Health Statistics*, ser. 11, no. 124. Washington DC: Health Services and Mental Health Administration, United States Government Printing Office.

Preece, M. A. and Baines, M. J. (1978) A new family of mathematical models describing the human growth curve. *Ann. Hum. Biol.*, **5**, 1–24.

Roede, M. J. and Van Wieringen, J. C. (1985) Growth diagrams 1980: Netherlands third nation-wide survey. *Tijdsch. Soc. Gezhdszg*, **63**, Suppl., 1–34.

Royston, J. P. (1982) An extension of Shapiro and Wilk's *W* test for normality to large samples. *Appl. Statist.*, **31**, 115–124.

Silverman, B. W. (1985) Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. R. Statist. Soc.* B, **47**, 1–52.

Tanner, J. M., Whitehouse, R. H. and Takaishi, M. (1966) Standards from birth to maturity for height, weight, height velocity, and weight velocity. British children, 1965—I. *Arch. Dis. Child.*, **41**, 454–471.

Van't Hof, M. A., Wit, J. M. and Roede, M. J. (1985) A method to construct age references for skewed skinfold data, using Box–Cox transformations to normality. *Hum. Biol.*, **57**, 131–139.

Whitehead, R. G., Paul, A. A. and Ahmed, E. A. (1988) DHSS present day feeding practice and its influence on infant growth. In *The Physiology of Growth, Proc. SSHB Symp.*, no. 28. Cambridge: Cambridge University Press. In the press.

DISCUSSION OF THE PAPER BY COLE

Miss Susan Chinn (United Medical and Dental Schools of Guy's and St Thomas' London): Dr Cole's paper is most welcome in bringing the problems of centile estimation before a wider audience than in the past. I strongly endorse his aim to produce smooth centile curves that are based on an underlying distribution as it is unsatisfactory to carry out separately the related procedures of centile assessment and standard deviation score calculation for an individual child.

Would Dr Cole explain how he reconciles the sentences 'Table 8 shows how insensitive the estimates of λ are to the extreme values in each tail' and 'Thus the choice of λ is determined largely in the extreme tails of the distribution'?

Data from the national study of health and growth for 6000 English boys and girls, aged 4.5–12 years, confirm Dr Cole's findings of poorly determined λ for height, and a trend in λ with age in boys, but not in girls. The usual assumption of a normal distribution is quite justified for girls of this age, and I doubt that there is any advantage in employing a transformation for boys.

For weight I also find λ significantly less than zero for year age groups from 5 to 11 years. Given the large numbers that are usually, and should be, involved in centile estimation it follows neither that the extra effort in fitting a third parameter is necessary, nor that λ is the parameter of choice. Preliminary results suggest that for weight the effort is worthwhile. Given the need for a three-parameter distribution the general power transformation is a sensible place to start, but the *LMS* method may not suffice for all biological variables, and a four-parameter distribution may sometimes be necessary.

I disagree with Dr Cole over his apparently equal treatment of age and time. Longitudinal data on a relatively small number of children are not a substitute for the same number of measurements obtained on independent samples for each age group. Longitudinal data will provide a smooth standard deviation curve because essentially the same information is being repeated at each age. Of more importance reference curves obtained in the way described, known as 'distance' standards (Tanner, 1986), should not be used to follow the growth of a child; they should only be used for the initial assessment to take the decision to follow-up the child or not. When two measurements are available some measure of rate of growth is required and is used if a child is being assessed for treatment with growth hormone. Most commonly used are 'velocity' standards, e.g. height gain by age. These have been produced by Tanner *et al.* (1966), but smoothed by eye. Herein, surely, lies the important question, whether adequate distributions can be found, or whether distribution-free estimation should be used (Healy *et al.*, 1988). Even better are conditional standards for the second measurement given the first, as proposed by Cameron (1980). For height the two standards are not identical because height gain is positively related to attained height. Longitudinal standards (Tanner *et al.*, 1966) seek to provide an assessment of a child's growth curve against reference growth curves. Unfortunately sufficient data to provide estimation of such curves are rarely available, but the Cambridge infant growth study could provide such information for the first year of life.

None of these considerations is the most crucial for the assessment of a child, which is against what reference should the child be assessed? Dr Cole mentioned the problem of intercountry comparison, but it exists within country too. It is easy to show that attained height is significantly and independently associated with many factors. It does not follow that we should adjust for each of them in a child's assessment. Stunted growth may be more common under a combination of adverse social circumstances but has the same consequences as for a child not socially disadvantaged. However, if we assess a child of West Indian origin against Caucasian reference data we may seriously delay detection of growth hormone deficiency. A West Indian child can be above the nominal 3rd centile, but below the 1st centile if compared with other children of like ethnic origin. Inclusion or exclusion of adjusting factors is not clear cut. Reference curves conditional on mid-parental height have been produced (Tanner *et al.*, 1970) but the association between a child's height and that of his parents will not be entirely genetic. We need an appropriate health outcome to determine which factors should be included, and the optimal adjustment for each, analogous to moving on from reference ranges for serum constituents to the use of the same data for prognostic prediction or differential diagnosis.

It gives me very great pleasure to propose the vote of thanks.

Mr Jon Rasbash (Institute of Education, London): I would like to talk briefly about an alternative method for centile estimation, referred to by Dr Cole (and described in detail in Healy *et al.* (1988)). The method involves two stages.

1. *Obtaining the 'raw' centiles*

The data are sorted into ascending age order. A regression is fitted to the first k points in the data set. The required centiles are obtained from the ranked residuals, using interpolation where necessary. The centile points are plotted against the median age value of the k points. This procedure has used points $1-k$ of the data; the procedure is repeated using points $2-(k+1)$, $3-(k+2)$ ... until the entire age span has been covered.

2. *Smoothing*

The centiles arising from stage 1 will be very irregular and will require smoothing. Each centile should follow a smooth curve. However, for a fixed age the gaps between the centiles should also vary smoothly. A smooth curve for the ith centile can be estimated by the polynomial

$$y_i = a_{0i} + a_{1i}t + a_{2i}t^2 + \ldots + a_{pi}t^p. \tag{1}$$

We then make the set of a_{ji} coefficients linear functions of $z_{i.}$, where $z_{i.}$ is the normal equivalent deviate of the ith centile, i.e.

$$a_{ji} = b_{j0} + b_{j1}z_i + b_{j2}z_i^2 + \ldots + b_{jq_t}z_i^{q_t}. \tag{2}$$

q_j, the order of polynomial fitted to the set of a_{ji} coefficients, may vary from one value of j to another. Combining equations (1) and (2) gives a linear model which simultaneously fits all the centiles estimated by stage 1. This model can be fitted by least squares.

Dr Cole criticizes this method because 'it does not have an underlying distributional form to convert centiles to SD scores'. It is true that the method is nonparametric; however, it is still possible to evaluate SD scores for data points. Stage 2 gives us a polynomial function which predicts the value of a measurement, y, for a given z and t. Therefore, a single measurement, for which t will be known, can be converted to its equivalent SD score by solving a polynomial equation.

If I understand Dr Cole's paper correctly, the example analyses were carried out on relatively large data sets between 50 and 150 points per age interval. With smaller data sets it may be necessary to use broad time intervals to ensure that each interval contains sufficient points to estimate $\hat{\lambda}_i$, $\hat{\sigma}_i$ and $\hat{\mu}_i$ to a satisfactory precision. This may lead to estimating the $L(T)$, $S(T)$ and $M(T)$ curves from rather few points and as a result the centiles produced from these curves may be oversmoothed.

A solution to this problem might be to employ a method similar to stage 1 of the Healy *et al.* procedure. This would involve moving a 'box' of appropriate size through the data set and estimating the parameters $\hat{\lambda}_i$, $\hat{\sigma}_i$ and $\hat{\mu}_i$ for each box position. This approach would produce far more points from which the smooth $L(T)$, $S(T)$ and $M(T)$ curves could be calculated, thereby reducing the possibility of oversmoothing.

Finally, it would be interesting to see how Dr Cole's method fares with data which depart more radically from normality than weight such as skinfold thickness.

I would like to congratulate Dr Cole on a most stimulating paper. It gives me great pleasure to second the vote of thanks.

The vote of thanks was passed by acclamation.

Dr S. Rosenbaum (Radlett): It is likely that linear measurements of the body are related and tend to be normally distributed, while girths and associated measurements, such as skinfold thickness and weight, are also related and their distributions tend to be skew. While, however, as the paper points out, a log-normal transformation is commonly employed on such data, it should be modified by the subtraction of a constant c before logarithms are taken. For example, Edwards *et al.* (1955) recommended that 1.8 mm should be subtracted from the skinfold thickness in investigations of subcutaneous fat—a general correction appropriate to all sites at which measurements are taken—while I once estimated a value of 4.2 mm for the subscapular site in a sample of young adult males. The method is described in Rosenbaum (1988) in the context of weights and, although the estimates for the available data are very variable, it is reasonable to extend Rona and Altman's (1977) values in their study of children from 16 kg at age 11 to 18 kg at age 15 and 20 kg above age 16. It is interesting that Royston (1982), in the paper referred to by Dr Cole for testing for normality using the Shapiro–Wilk statistic W, also gives an example of its use for estimating c in the log-normal transformation, taking as his example the Rona and Altman data just mentioned. Given these values what is the result of a Shapiro–Wilk test on the data of Table 5 of Dr Cole's paper? It is presumably not as good as the *LMS* method but better than the implied value $c = 0$.

Royston (1986) provides a modification of his algorithm AS 181 for the W test when there are ties which are a feature of grouped data—this may also present a problem for the *LMS* method. Another practical problem arises when data are truncated, for example, if there are minimum height and weight standards for a profession as for policemen and policewomen; an assumption such as log-normality then enables us to estimate the mean and standard deviation, and hence the centiles. Can the *LMS* method be adapted to this? (Even without truncation, the mean and standard deviation only may be available.)

Robert G. Newcombe (University of Wales College of Medicine, Cardiff): The model that has been described is very elegant. There are potential applications other than construction of growth standards. The *LMS* model is very flexible and richly parameterized and could prove useful generally to characterize a bivariate distribution in which the roles of independent variables are clear cut—a general characterization of the dependence of y on x in a non-linear regression context. For example, Quetelet's index, weight/height^2, is often used as a measure of obesity; the *LMS* model could be used to investigate deficiencies in the assumed constrained model with $M(x) \propto x^2$, to suggest improvements and to characterize a particular population. Possible applications are not limited to anthropometry.

My main misgiving is that the application of growth curves, however well constructed, in assessing the adequacy of growth of an individual may still fall short of the ideal. Often, particularly in the prenatal period, longitudinal data on a subject are assessed, and an individual crossing centile curves away from the median may warrant further investigation and intervention. A study of the longitudinal properties would be rewarding. Any of the scores set out in Section 2.6 could be used and autocorrelative properties evaluated. Appropriate surveillance of the individual could be set on a more rational basis by adapting methods used in quality control.

Whether the sample size in the Cambridge study is adequate to produce normative curves remains questionable. (To the medical statistician, the principle that sample size virtually always means the number of individuals is cardinal.) If the objective is to assess whether future subjects are drifting out of control by crossing centile curves outwards, it will not matter critically if the training data set contains one persistently unusually heavy subject. However, if the intention is to measure each future individual on a single occasion, the standards will in effect be distorted throughout by the anomalous subject who has provided several deviant values; in this situation the apparent adequacy of the small longitudinal data set is illusory.

Professor F. D. K. Liddell (McGill University, Montreal): I have considerable interest in anthropometric data, not so much with centiles as with 'average' changes over *both* age *and* period. While I have found the contributions fascinating, I am left with the impression of 'black art' (Section 5), because (Section 2.3) 'The smoothing can be done using whatever method is convenient, e.g. thus or thus, or thus or thus, or simply fitting by eye', or by procedures recommended by Mr Rasbash.

Now to the data used for illustrative purposes, or as the author claims 'validation'—but (in my view) no more than justification: Table 2 (Cambridge data) shows that the number of infants varied (even for weeks 4–52) from the full quota of 132 to only 124. Clearly for many, measurements were complete (weeks 4–52); I suggest it would have been better to use only their material, and, particularly, to do so in a truly longitudinal sense, for I agree with Miss Chinn on the virtues of assessing growth from one age to another.

My earliest relevant interests were in lung function measurements in British coal-miners, first cross-sectionally and then, five years later, longitudinally. (The relevance is not marginal, because reference data on lung function *are* required by respiratory physiologists and chest physicians (admittedly not based on one occupational group).) The clear downward gradient with age of, say, forced expiratory volume (FEV) over 1 s obtained from the first cross-section would have been a poor predictor for a later cross-section. At each age after about 20 years, the average FEV *did* decrease in the five years, but considerably less than the cross-sectional gradient had implied, with improvements in some miners. (Inevitably, 'regression to the [current] mean' would be superimposed on the trend.) The main (not the only) reason that the average fall was less than 'predicted' was that older men were shorter than younger men.

This brings us back to anthropometric measurements. Dr Sidney Rosenbaum has demonstrated marked increases in heights, and even more so of weights, of males at ages 20–24, over 100 years, maintained through the last six decades. A cross-sectional survey today of those aged 20–80 would raise many problems in disentangling changes as due to age or period.

Dr Cole's paper will surely make an important impact in its true field—or (in the light of other comments here) fields.

Mr J. P. Royston (Medical College of St Bartholomew's Hospital, London): My comments are limited to the crucial issue of the Gaussian distribution. If the transformed data are for practical purposes Gaussian, a great simplification in the calculation of centiles results; if they are not, Dr Cole's method of first resort breaks down and some further smoothing work has to be done on the centiles. But first, a small addition to the material of Section 2.4: the Shapiro–Wilk W is a good omnibus test, but it is weakest against symmetric alternatives with rather long tails. Therefore examining the kurtosis coefficients (b_2) of the transformed data might also be sensible.

It is indeed convenient that the boys' power-transformed weights in the health examination surveys (Table 5) show no sign of departing from a Gaussian distribution, despite the enormous sample size (433–641 per age group). Some statisticians have criticized the use of significance tests with such large samples on the basis that the power to detect departures is so great that one would be very unlikely in practice ever to find a sample that did not fail the test. This has not been my experience, and it is pleasing that Dr Cole has come to the same conclusion in his paper. It would have been nice if Table 5 had given the P values for the W tests so that a composite P value could have been obtained, for instance using Fisher's method which says that $-2\,\Sigma \ln P_i$ is distributed as χ^2 when the P_i are uniformly distributed on the interval (0, 1).

An alternative way of estimating λ, which follows the example given in Royston (1982), is simply to choose the value of λ which maximizes W (or perhaps more conveniently the correlation between the ordered data and the corresponding expected Gaussian order statistics) for the sample. If you wish to guard against outliers you can omit the most extreme pairs of values as Dr Cole has done in Tables 4 and 8. This has the advantage that the transformed data are in a sense as Gaussian as you can make them, but the disadvantage that the estimate of λ is not maximum likelihood. In practice the results are likely to be similar, but the W method is quite general and can be applied to any exotic transformation of the raw data for which maximum likelihood estimates of the parameters may be difficult to obtain.

Dr P. J. Green (University of Durham): I particularly liked Dr Cole's model, given by equation (7) (or, equivalently, equation (8)): this amounts to a model for regression of Y on T that is nonparametric (no parametric form for L, M or S is prescribed) but not distribution free (normality is retained).

However, the procedure for fitting the L, M and S curves described in Sections 2.3 and 2.9 seems unnecessarily complicated: it involves grouping the data by values of T (which will sometimes be necessarily arbitrary) and then a number of different stages (some with alternatives), in which fitting and smoothing are separate steps, performed separately for each of the three curves. In the absence of any characterization, properties of the estimated curves are unclear: they are defined purely operationally.

A more automatic, single-stage fitting procedure may therefore be attractive: one that might be appropriate for *sample* data, on up to a few thousand individuals, can be based on penalized likelihood. One way of doing this involves maximizing

$$\sum_{i=1}^{n} \left(L(t_i) \log \left[\frac{y_i}{M(t_i)} \right] - \log S(t_i) - \frac{1}{2} \left\{ \frac{[y_i/M(t_i)]^{L(t_i)} - 1}{L(t_i)S(t_i)} \right\}^2 \right) - \frac{1}{2}\alpha_L \int [L''(t)]^2 \, dt - \frac{1}{2}\alpha_M \int [M''(t)]^2 \, dt - \frac{1}{2}\alpha_S \int [S''(t)]^2 \, dt$$

over choice of curves L, M and S. Each of the resulting estimates are natural cubic splines with knots at the distinct values of $\{t_i\}$. There are just three constants to choose, which control the smoothness of the fitted curves.

This apparently horrible optimization problem can be handled very economically. Fisher scoring yields an iterative algorithm with an updating reminiscent of ridge regression (see Green (1987)). Derivatives of the log-likelihood with respect to values of $L(t)$ do not have finite expectations, so I used only the first few terms of a Taylor expansion of $\log[y_i/M(t_i)]$. The updating is handled by an inner iteration cycling over application of Reinsch's algorithm (1967) for cubic spline smoothing to each of the three curves in turn.

The inner iteration takes $O(n)$ time per cycle, and in practice, applied to the weights from the Cambridge infant growth study, convergence was obtained in about 15 cycles.

This use of spline smoothing is quite different from that mentioned in Section 2.3: here the spline structure is not imposed.

This approach leads to piecewise polynomial curves for L, M and S that may be easily plotted, as can the resulting centile curves. It would be straightforward to adapt the method to handle covariates, and matters such as pooling the sexes when estimating L and S but not M.

A final question: what *is* the null distribution of the Shapiro–Wilk W statistic, when a power transformation has been estimated?

Professor Sir David Cox (Imperial College, London): We have heard a most interesting account of a problem that is at first sight conceptually fairly simple but which is clearly far from straightforward and which can be tackled in quite a number of different ways. One approach which has received some attention in the econometric literature centres on a notion of quantile regression formulated as follows. The median of a population and a sample minimize respectively

$$\int |y-\theta| \, \mathrm{d}F(y), \qquad \Sigma \, |y_j-\theta|,$$

in an obvious notation. The pth quantile can be shown to minimize

$$\int [p(y-\theta)^+ + (1-p)(\theta-y)^+] \, \mathrm{d}F(y),$$

$$\Sigma \, [p(y_j-\theta)^+ + (1-p)(\theta-y_j)^+].$$

Here $z^+ = \max(z, 0)$. Suppose now that θ is replaced by some suitable function of explanatory variables with unknown parameters. Then the second minimization defines an estimated quantile regression (Koenker and Bassett, 1978). Further, replacement in the sample minimization of z^+ by $h(z^+)$ defines a family of estimates whose properties can be studied. This is explored in detail by Newey and Powell (1987).

This approach refers to a single value of p. To examine several values of p together some modification might be needed.

While my initial preference would normally be for Dr Cole's procedure, it would be interesting to know more about the relative merits of these and other approaches.

Dr J. M. Bland (St George's Hospital Medical School, London): I was drawn into this area when I was approached by a paediatrician who had collected some data. He had obtained birth weights for a series of births before the 30th week of gestation, giving 171 very little Neapolitans. He wanted to use these data to construct a centile chart of birth weights in Naples at each week of gestation before 30 weeks. He thought that this was a good and worthwhile thing to do, and he convinced me of that.

I came to this problem with an uncluttered mind and produced what I considered to be an adequate solution that did the job. As I was doing this, and afterwards, I discovered that many such charts had been published, from charts which had used different models at different gestational ages connected by smoothing, to charts which were presented without any hint of how they were derived.

Dr Cole's method appears to be a considerable improvement on all those at which I looked as well as on the method that I used.

However, my clinical colleague thought it was important that the method we used should be one that he could understand. He also thought it would be even better if he could carry it out himself, and if paediatricians scattered around the world could also carry it out themselves. Although I very much like Section 2.9 of Dr Cole's paper, I thought that it was unlikely that many of my clinical colleagues would be able to apply Dr Cole's method by following it.

I wonder how users of centile charts react to this fairly complex method of calculation, and whether they find charts based on it convincing.

Perhaps I can encourage Dr Cole to prepare a version of his paper incorporating fully worked examples which any paediatrician would be able to follow and carry out himself. Then, the next time that this happens, I can show them a paper which shows them how to do it for themselves.

The following contributions were received in writing after the meeting.

Professor M. J. R. Healy (London School of Hygiene and Tropical Medicine): It should be pointed out that the Healy–Rasbash–Yang (HRY) method (Healy *et al.*, 1988) does make it possible to convert centiles to SD scores. The method effectively fits a polynomial to the age-specific centiles on a normal plot, so that SD scores can be obtained by solving a polynomial equation.

A major difficulty with all methods for forming centile charts is that practical interest is concentrated in the tails of the age-specific distributions. Direct information on these is scarce and hence the methods effectively extrapolate from the inner centiles by making some kind of assumption about distributional form. The *LMS* method makes the quite strong assumption that the symmetrized curve is normal (the HRY method is at least theoretically more flexible). This assumption can be tested (Section 2.4), but nothing is said about what to do if the test fails. The problem is manageable out to around the 3rd and 97th centiles, which are the outermost on the usual charts, but it is much more difficult when the method is used, as would be quite common in practice, to derive SD scores outside the range (say) $\pm 2\frac{1}{2}$. It is important to realize that the use of extreme SD scores is based, by definition, on minimal empirical evidence.

Dr David C. Hoaglin (Harvard University) **and Dr John H. Himes** (University of Minnesota School of Public Health): Even though Cole's method produces centile curves in the scale of the data, its use of potentially different transformations on consecutive segments of a single set of reference data over fairly short periods seems needlessly cumbersome and difficult to interpret. Instead, we believe that it is usually possible to follow the customary strategy (as in the examples of Box and Cox (1964)) of allowing the data to guide a choice among plausible round values of λ (e.g. λ values of 1, $\frac{1}{2}$, 0, $-\frac{1}{2}$ or -1). Within Cole's framework we can make a good start by plotting interval estimates of the λ_i rather than point estimates and then asking what constant values of λ are compatible with the data. Thus, for the weights in the Cambridge infant growth study, 15 intervals, each at about 99.66% confidence, would yield 95% simultaneous confidence. If we take 0.6 as the standard error of each of these $\hat{\lambda}_i$, then the intervals (centred at $\hat{\lambda}_i$) have an approximate half-width of $2.93 \times$ standard error$(\hat{\lambda}_i) \approx 1.76$, and all 15 intervals would include $-0.74 \leqslant \lambda \leqslant 0.19$, i.e. the data seem compatible with transformation by either the logarithm or the reciprocal square root (among the simple powers).

In addition, the potential sensitivity of each $\hat{\lambda}_i$ to a few extreme values is cause for concern. We should either trim them before proceeding or adopt a resistant method of obtaining $L(t)$, $M(t)$ and $S(t)$.

Depending on the needs of the application, other approaches to producing smoothed centile curves deserve consideration. If the centiles given in the reference data suffice, a version of resistant delineation (Tukey, 1977; Himes and Hoaglin, 1989) should work well. In an application to triceps skinfold thickness of males aged 1–20 years, we found that this approach retained features, visible in the raw centiles, that the centile curves published by the National Center for Health Statistics had smoothed away. As an alternative when other centiles are desired and each segment has sufficient observations, we can use Tukey's g-and-h distributions (Hoaglin, 1985) to describe the shape of the distribution at t_i and to obtain the further centiles directly. The g-and-h family includes the Gaussian and log-normal distributions as particular cases and yields estimates of skewness and elongation parameters from symmetric centiles. We can then smooth these estimates, as well as location and scale, and construct smooth centile curves.

Dr M. C. Jones (University of Bath): I would like to propose an alternative fully nonparametric approach to fitting smooth centile curves to reference data. This contribution is more a discussion of the discussion than of the paper directly, combining much of what Dr Green and Professor Cox have said. Professor Cox has reminded us of an objective function which, when minimized, yields centiles; now include in this objective function an unspecified function of t, add on a roughness penalty of the type employed by Dr Green and minimize this to obtain a 'spline smoothing regression quantile'. Specifically, for a single centile, define $\hat{C}_{100\alpha}$ to be that f which minimizes

$$R_\alpha(f) = \sum_{j=1}^{n} \rho_\alpha(y_j - f(t_j)) + \lambda \int [f''(t)]^2 \,\mathrm{d}t$$

over an appropriate smoothness class of fs, where

$$\rho_\alpha(x) = \begin{cases} \alpha x & \text{if } x \geqslant 0 \\ -(1-\alpha)x & \text{if } x < 0. \end{cases}$$

Here, $\{(t_j, y_j), j = 1, \ldots, n\}$ is the reference data set and λ is an unspecified smoothing parameter.

Now $R_\alpha(f)$ is precisely of a form recently becoming popular in the smoothing spline literature in the guise of 'robust smoothing splines'; see section 6.2.5 of Eubank (1988) for references. We need to investigate these methods for our particular choice of the function ρ_α; a start is made in chapter 5 of Bloomfield and Steiger (1983) under the names 'generalized LAD splines' and 'quantile splines' (I am grateful to John Rice for this reference). To obtain the entire family of centile curves required, indexed by α_i, say, we choose $\hat{C}_{100\alpha_i}$ to minimize each $R_{\alpha_i}(f)$ separately. Note that it seems appropriate to use the same value of λ for each α_i. This single λ, reflecting how smoothly the entire response distribution varies with t, is a major conceptual, and consequently practical, advantage of the method proposed over that of Dr Green. Getting away from the Box–Cox parameterization is also a personal preference, although Dr Cole gives the impression that the need to do so for most reference data is not urgent. Unlike Dr Green's method, the current method would allow centile curves to touch and/or cross; if this reflects a paucity of data in the region concerned, such an anomaly could turn out to be a most useful warning. Like Dr Green's method, there is no need to group the data at all. (An alternative 'robust' kernel approach involving ρ_α is also possible; see section 4.11 of Eubank (1988)).

Reference data clearly provide an important area to which smoothing methods can usefully be applied and Dr Cole is to be congratulated for bringing this topic to the attention of a wider group of statisticians.

Ms S. B. J. Macfarlane (Liverpool School of Tropical Medicine): Dr Cole's paper is of particular interest to me as I have been involved in creating reference standards for a variety of anthropometric measures for Nigerian Yoruba children (Janes *et al.*, 1981). The measurements were taken longitudinally on elite children at exact ages. Sample sizes at any one age for each sex vary approximately between 50 and 150.

Our method for arriving at length and height centiles was to estimate means and standard deviations at each age using the method of Patterson (1950) which takes advantage of the longitudinality of the data and then to find centile points using the properties of the normal cumulative density function. Smooth curves were fitted using the cubic spline technique. The assumption of normality seemed justified at each age on the basis of the measures of skewness and kurtosis, the linearity of the probability plot and the result of the Shapiro–Wilk test.

Values of λ have been calculated for the lengths of female children aged between 1 month and 3 years and for their heights between 3.5 and 10 years using the Box–Cox method of transformation (Fig. 9). All the values of λ are below unity and two are below -3.9. Approximate confidence limits for λ are very wide with standard errors lying between 1.8 and 3.6. These results do not convince me of the advantages of transforming length and height data, particularly as standard deviation scores would be so much more difficult to calculate subsequently for individual children.

Van't Hof *et al.* (1985) proposed calculating λ for noisy data such as skinfold thickness. The pattern of change in triceps fat with age for the Nigerian children is very different from that in British children and we have not produced centile charts. The values at each age are far from normal with measures of skewness and kurtosis for females as high as 2.3 and 10.4 respectively. The λ values fitted for the triceps

Fig. 9. Nigerian growth study: estimates of the Box–Cox power for length (up to 3 years) and height (from $3\frac{1}{2}$ years) at each age for elite females

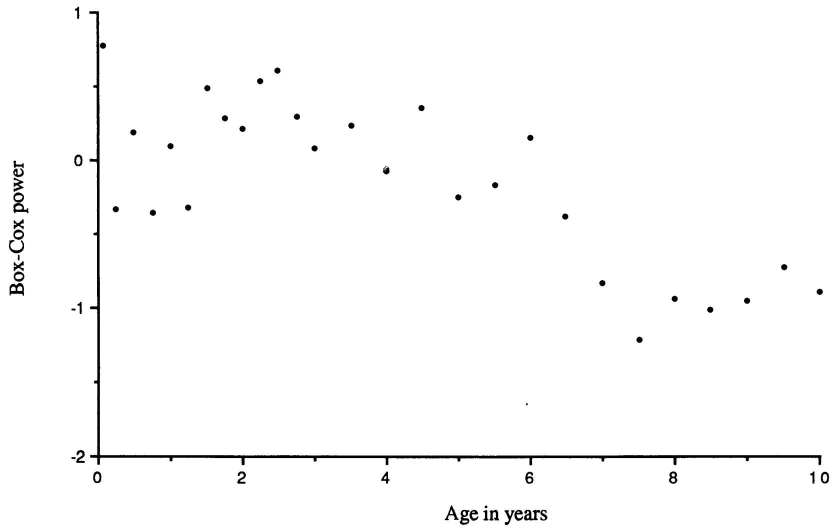

Fig. 10. Nigerian growth study: estimates of the Box–Cox power for triceps at each age for elite females

fat of female children aged between 1 month and 10 years are shown in Fig. 10. There is a clearer pattern to these values than is apparent in Fig. 9 and the standard errors for λ, ranging from 0.27 to 0.73, are considerably lower than for length and height. The resulting measures of skewness and kurtosis are near to zero and 3 respectively as would be expected for a normal distribution.

I am grateful to Dr Cole for providing a coherent approach to the fitting of centile curves and I look forward to examining the Nigerian data further to consider his conclusion that relatively small longitudinal studies can provide well-defined estimates for the L, M and S curves.

Professor A. F. Roche and Professor Shumei Guo (Wright State University, Dayton): The clinical application of reference data is stressed, but their use in epidemiological studies, clinical trials and public health surveillance is omitted. It is stated that reference charts can be used to evaluate the change in a measurement for an individual, but the distribution of changes in centile levels for a population is generally unknown.

Cole used the Box–Cox method to normalize distributions and applied a smoothly varying transformation which requires plotting L, M and S curves that relate to the power transformations, means and standard deviations respectively. He claims that raw or smoothed values for the power λ can be used. We consider smoothed values more appropriate. He offhandedly states that 'the smoothing can be done by whatever method is convenient'. Smoothing methods should be judged by the goodness of fit which cannot be determined when smoothing is done by 'simply fitting by eye'.

Cole considers the sensitivity of λ and adjustments for the reduction in variance due to trimming, but the justification for removing outliers and then adjusting the variance of the estimates is unclear. His data show that the effects of removing outliers on $\hat{\lambda}$ increase with age; these effects may be large at older ages. Some of his statements require justification. For example, he states 'if the result [value of $\hat{\lambda}$] is negative, make it positive'. It is doubtful whether the distribution will be normalized if this is done.

Cole analysed raw data that provide rather regular empirical centiles that need little smoothing. In these circumstances, his approach produces only small changes. It would have been better to fit a mathematical function to the serial data for each infant and to derive centiles from the fitted curves.

Cole claims that the spline approach of Hamill 'leads to difficulties with spacing the centile curves appropriately' but does not present supporting evidence. He states that Hamill published SD scores obtained by using estimates of the SD for the upper and for the lower parts of the distributions. This was done by the World Health Organization (1983), not by Hamill.

Dr Cole addressed this important topic logically, but the Box–Cox technique is unnecessary when empirical centiles are fairly smooth.

The **author** replied later, in writing, as follows.

The points raised, which cover a wide range of issues, fall into three broad areas: general thoughts on the construction and use of growth standards, specific comments on the performance of the *LMS* method and suggestions for new and/or improved approaches to the problem.

General thoughts on growth standards

Several discussants (Miss Chinn, Dr Newcombe, Professor Liddell, Ms Macfarlane and Professor Roche and Professor Guo) refer to the need for longitudinal data to be analysed appropriately, and for the resulting standard to be used appropriately. It is certainly true that, when at least two measurements are available on an individual, assessing recent change is of more value diagnostically than assessing single measurements. As Miss Chinn points out, this can be achieved either with velocity standards or conditional standards. The *LMS* method can be applied to both. However, any standard of this form is based on a specified time interval between measurements (1 year in the cases that she cites), which can be very restrictive when the time interval available is much shorter. It is particularly so during the first year of life, when concern about possible growth failure may develop over a period of weeks rather than months. There is surely an opportunity here for longitudinal standards based on the Cambridge infant growth study (IGS), as Miss Chinn and Dr Newcombe imply.

I am grateful to Professor Roche and Professor Guo for listing some of the other areas where centile charts for anthropometry are found to be useful. I also stand corrected on two points concerning the National Center for Health Statistics (NCHS) standard. Although Hamill *et al.* (1977) talked of the difficulties involved in the spline fitting, they were talking more of minimizing the maximum residuals than spacing the centiles appropriately. Also I did not make clear that the calculation of SD scores with an upper and lower estimate of the SD was done by the World Health Organization, not the NCHS (Dibley *et al.*, 1987).

As Miss Chinn points out, the appropriateness of the reference standard is crucial to the success or failure of a screening programme to identify growth failure. It is clear that, once a child has been identified as being at risk, subsequent growth velocity measurements are likely to confirm or deny the diagnosis—the difficulty occurs at the screening stage. The effects of various social and biological factors on child height have been studied in the national child development study (Goldstein, 1971) and the national study of health and growth (NSHG) (Rona and Chinn, 1986); the results are useful to quantify the risks of misclassification if the standard is inappropriate.

Professor Healy makes a fundamental point about the philosophy of growth standards, which is that the area of interest in the standard (the tails) is where the data are not. There is no obvious way round this dilemma, so that assumptions of regularity have to be made to justify extrapolating into the tails of the distribution. Clearly, methods which define and thus test the nature of the regularity assumption are the only ones that are able to extrapolate beyond the data with any degree of credibility. I return to this point later, when comparing the various methods raised in the discussion.

Performance of the LMS method

I am grateful to Miss Chinn, Dr Green and Ms Macfarlane for applying the *LMS* method to their own data (or, in Dr Green's case, mine!). I agree that the results for height are generally unimpressive, although there is now mounting evidence from large data sets that the *L* curve changes during puberty (Table 6; Cole, 1988). It is useful to know that for weight in the NSHG, λ is significantly less than zero until the start of puberty, confirming the pattern seen in Table 5.

Mr Rasbash asks about the *LMS* method applied to skinfold data. I cannot do better than refer him to Ms Macfarlane's Fig. 10, or alternatively Van't Hof *et al.* (1985), which both suggest that the method works well.

On the fitting of the method, Dr Bland's comments about Section 2.9 of the paper are well taken — I realize now that the section is pitched rather high for the non-statistician. A cook-book paper would be very useful for clinical practitioners, and it needs to be written.

Several discussants take me to task, correctly, for not exploiting the longitudinal nature of the IGS when estimating the age trend. The analyses suggested include Patterson's (1950) approach (Ms Macfarlane), fitting separate functions to each infant's data (Professor Roche and Professor Guo) and excluding incomplete date (Professor Liddell). This last is somewhat inefficient, and another possibility, which overlaps with Patterson's method to some extent, is to fit separate constants for each subject in the regression analysis.

Miss Chinn and Dr Newcombe are also unconvinced by my claim that the IGS is sufficiently large to provide reference curves. I note though that Ms Macfarlane's Nigerian standards are based on only slightly larger numbers, and that portions of several well-established standards (e.g. 0–5 years in Tanner *et al.* (1966) and 1–5 years in Tanner and Whitehouse (1975)) rely on longitudinal data from no more than 200 children.

As regards the mechanics of the *LMS* method, most comments refer either to the estimation of λ and the L curve or to the assumption of normality. Mr Royston suggests as an alternative to the maximum likelihood estimate, the λ value that maximizes the W statistic. This is an elegant method which can, as he says, be applied to any distribution. The main disadvantage would presumably be in obtaining expected order statistics for the more exotic distributions.

Dr Hoaglin and Dr Himes are not convinced that λ needs to change with age, despite the evidence in Tables 5 and 6 to the contrary. It is true, as they say, that for weight in the IGS λ could be held constant at 0 or $-\frac{1}{2}$, but there is abundant evidence from other much larger studies that weight is normally distributed at birth and becomes skew only later. It is surely sensible that this pattern, which the IGS also shows, should be reflected in the standard.

Professor Roche and Professor Guo are confused by two aspects of the smoothing procedure. They quote me as saying that unsmoothed λ values can be used, but this only applies to the calculation of μ and σ. The λ, μ and σ values are subsequently smoothed to obtain the L, M and S curves, before the individual centiles are derived. Also, they dislike the idea that the smoothing can be done in more than one way. They forget that any smoothing procedure, even one monitoring the goodness of fit, involves an essentially subjective trade-off between goodness of fit and smoothness. Incidentally the other statement they take exception to, making the value positive if it is negative, refers not to λ but to σ. A negative sign for σ shows that the distribution has been inverted, with high original values mapping to low transformed values and vice versa, as occurs with an inverse transformation ($\lambda = -1$).

Dr Hoaglin and Dr Himes are concerned that λ is not sufficiently robustly estimated. Although Table 8 refutes this suggestion for weight in the American surveys, it is nevertheless a valid concern. Trimming the data slightly may be a sensible precaution, although, as Professor Healy says, the data in the tails are of fundamental importance, so they should not be dismissed too lightly. On this point, Miss Chinn notes that I am inconsistent in discussing the relevance of the tails. What I meant to say (last paragraph of the results section) was: 'Table 8 shows how close even the extreme tails of the distribution are to their expected values, so that excluding them affects λ very little'. Professor Roche and Professor Guo claim that the trimming affects λ more at greater ages—this only occurs in the IGS and is due, as I discuss in the paper, to one extreme infant.

On the issue of assuming normality, I am grateful to Mr Royston for putting this into perspective. The lack of significant departures from normality in sample sizes of greater than 400 is very reassuring, even if, as Dr Green points out, the W statistic is not strictly designed for data transformed to normality. Following Mr Royston's suggestion, I have combined the significance levels for each age group to give a statistic distributed as χ^2 on 12 degrees of freedom. The result for weight, 22.4 is just significant ($P < 0.03$) while that for height at 14.8 is not ($P > 0.2$). Thus even the composite test finds only marginal evidence of non-normality.

Dr Rosenbaum's transformation log(weight $- c$) interestingly is slightly closer to normality than the best power transform, giving a χ^2 of 4.5 on 6 degrees of freedom for the composite W statistic applied to the age group 12–17 years. This is obtained by assuming that c changes linearly with age: $c = 10.5 + \text{Age}/2$, an important requirement to avoid discontinuities at year boundaries. The kurtosis of log(weight $- c$) is in the range 2–3, which is 0.1 – 0.2 higher than for weight$^{\lambda}$. This explains its slightly better fit and also illustrates Mr Royston's point that the W statistic is insensitive to departures from normality in symmetric distributions.

Dr Rosenbaum's other comment concerning ties in the W statistic is also relevant, although it does not arise in the American surveys as the data are the residuals after adjusting for age.

Professor Healy asks what is to be done if the normality test fails. The short answer is that I do not know, as it has not happened yet. However, if the non-normality were very marked (which I would not expect with anthropometry) it would be a serious problem.

Other methods

Apart from the normality assumption, the main weakness of the *LMS* method is its need to work with distinct age groups, as Dr Green points out. It is very gratifying to see four new or modified methods raised in the discussion, all of which address one or other of these problems. The methods of

Mr Rasbash and Professor Healy, Dr Green, and Professor Cox and Dr Jones operate on the complete data set, avoiding the need to split the data by age, while Mr Rasbash and Professor Healy, Professor Cox and Dr Jones, and Dr Hoaglin and Dr Himes suggest robust methods which do not assume normality.

I am glad that Mr Rasbash has taken the opportunity to explain the Healy–Rasbash–Yang (HRY) method in detail, as it involves two novel ideas worthy of wider discussion. I particularly like the polynomial sets of coefficients across centiles, which provide almost infinite flexibility to the centile spacings. However, I am less convinced that polynomials in age are the best way to model the shapes of individual centiles over a broad age range—a spline approach like Dr Green's would probably be more sensitive to the nuances of the underlying form. The suggestion of Mr Rasbash to calculate λ in overlapping groups is good and should be encouraged.

The order q of the polynomial across centiles in the HRY method is chosen to fit the plot of observed versus fitted quantiles (the Q–Q plot). For a normal distribution this plot is close to a straight line, while skewness makes it curve quadratically; kurtosis introduces a second, cubic, curve to the plot. Thus a skew distribution without appreciable kurtosis can be fitted by a quadratic polynomial. The Box–Cox transform used in the *LMS* method also gives a Q–Q plot of essentially quadratic shape ($g^{(\lambda)}$ in equation (5) plotted against y), so that, if q in the HRY method is less than 3, the two methods will give very similar results. Incidentally, Mr Rasbash and Professor Healy are right to emphasize that the HRY centiles can be converted to SD scores, and I am sorry to have given the opposite impression in the paper.

Dr Green's suggestion to fit the *LMS* method by penalized likelihood is one that I find very appealing. It has all the benefits of the *LMS* method without the cost, and the only subjective element lies in the choice of the smoothness constants α_L, α_M and α_S. I look forward to discovering just how easy his method is to implement, particularly if its computation time is only $O(n)$ per cycle.

The contribution of Professor Cox, deriving from the econometric literature, is characteristically both brief and thought provoking. I am glad that Dr Jones has risen to the implied challenge of adding flesh to the bones of Professor Cox's suggestion, with his elegant proposal to fit spline-smoothed regression quantiles. This is similar to Dr Green's version of the *LMS* method in that both minimize (maximize) a roughness-penalized function, but the two are fundamentally different in the sense that Dr Jones calculates each centile independently of its neighbours, whereas the *LMS* method obtains them as a set. Also Dr Jones's method has only one smoothing parameter where Dr Green's has three, and this might be a useful simplification. Quantile regression is also similar to step 1 of the HRY method, except that it can smooth the centiles more flexibly. However, I am not sure that the possibility of the centiles meeting or crossing is necessarily a benefit. It will be interesting to compare the results of the three approaches when they are available.

Dr Hoaglin and Dr Himes bring to the discussion some interesting ideas based on exploratory data analysis, providing robustness in the presence of outliers. The suggestion of using Tukey's g-and-h family of distributions at each t_i and then smoothing them, along with location and scale parameters, has an uncanny resemblance to the LMS method, complete with its disadvantage of having to group the ages. Dr Hoaglin (1985) has pointed out that the g-and-h distribution is analogous to the Pearson family of curves. This family was tried unsuccessfully by Hamill *et al.* (1977) when fitting the NCHS standard; to be fair though, their failure was due to the inflexibility of the smoothing polynomials rather than any deficiencies of the distribution itself. Nevertheless I remain unconvinced that the fourth kurtosis/elongation curve has much to offer over the other three curves. Certainly the g curve ought to perform just as satisfactorily as the Lcurve.

A useful idea from the EDA literature is that of Tukey's letter values, which correspond to a subset of the observed centiles. Starting with the two quartiles (fourths or F) and eighths (E), they continue with the centiles corresponding to successive halvings of the tail area; so D are the sixteenths and C the thirty-seconds. How far this continues (B, A, Z, Y, X etc.) depends on the size of the data set. This would be a rational set of centiles to calculate as step 1 of the HRY method, hence providing step 2 with extra information about the tails of the distribution.

In conclusion, three methods covered in the discussion (the *LMS* method as modified by Green, the HRY method and the Cox–Jones spline-smoothed quantile regression) convert the raw data to smoothed centiles in a single pass. Of the three the *LMS* method makes the biggest distributional assumptions, although for modest departures from normality (i.e. some skewness) it behaves very similarly to the HRY method. Both the HRY and the quantile regression approaches operate on a pre-specified set of centiles, so that extrapolating to centiles beyond the data is only possible if they are linked to an underlying distribution. The HRY method can do this extrapolation by solving the polynomial of order

q linking centile and SD score. However, for polynomials of higher order than quadratic the results are likely to be speculative at best.

As Professor Healy says, the main interest in the distribution is the region beyond the extreme observed centiles. The two requirements to predict this region at all accurately are a well-behaved distribution (such as is found with anthropometry) and an adequate summary of the distribution. The *LMS* and HRY methods both satisfy these requirements; however, the *LMS* method as presented in Section 2.9 allows the distribution to be estimated without the need for complex programming, and this is an advantage in practical applications.

REFERENCES IN THE DISCUSSION

Bloomfield, P. and Steiger, W. L. (1983) *Least Absolute Deviations: Theory, Applications, and Algorithms.* Boston: Birkhäuser.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *J. R. Statist. Soc.* B, **26**, 211–252.

Cameron, N. (1980) Conditional standards for growth in height of British children from 5.0 to 15.99 years of age. *Ann. Hum. Biol.*, **7**, 331–337.

Cole, T. J. (1988) Using the *LMS* method to measure skewness in the NCHS and Dutch national height standards. *Ann. Hum. Biol.*, to be published.

Dibley, M. J., Goldsby, J. B., Staehling, N. W. and Trowbridge, F. L. (1987) Development of normalized curves for the international growth reference: historical and technical considerations. *Amer. J. Clin. Nutr.*, **46**, 736–748.

Edwards, D. A. W., Hammond, W. H., Healy, M. J. R., Tanner, J. M. and Whitehouse, R. H. (1955) Design and accuracy of calipers for measuring subcutaneous tissue thickness. *Brit. J. Nutr.*, **9**, 133–143.

Eubank, R. L. (1988) *Spline Smoothing and Nonparametric Regression.* New York: Dekker.

Goldstein, H. (1971) Factors influencing the height of seven year old children—results from the National Child Development Study. *Hum. Biol.*, **43**, 92–111.

Green, P. J. (1987) Penalized likelihood for general semi-parametric regression models. *Int. Statist. Rev.*, **55**, 245–259.

Hamill, P. V. V., Drizd, T. A., Johnson, C. L., Reed, R. B. and Roche, A. F. (1977) NCHS growth curves for children birth–18 years. *Vital and Health Statistics*, ser. 11, no. 165. Washington DC: Health Resources Administration, United States Government Printing Office.

Healy, M. J. R., Rasbash, J. and Min Yang (1988) Distribution-free estimation of age-related centiles. *Ann. Hum. Biol.*, **15**, 17–22.

Himes, J. H. and Hoaglin, D. C. (1989) Resistant cross-age smoothing of age-specific percentiles for growth reference data. *Hum. Biol.*, to be published.

Hoaglin, D. C. (1985) Summarizing shape numerically: the g-and-h distributions. In *Exploring Data Tables, Trends, and Shapes* (eds D. C. Hoaglin, F. Mosteller and J. W. Tukey), pp. 461–513. New York: Wiley.

Janes, M. D., Macfarlane, S. B. J. and Moody, J. B. (1981) Height and weight standards for Nigerian children. *Ann. Trop. Paed.*, **1**, 27–37.

Koenker, R. and Bassett, G. S. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.

Newey, W. K. and Powell, J. L. (1987) Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819–847.

Patterson, H. D. (1950) Sampling on successive occasions with partial replacements of units *J. R. Statist. Soc.* B, **12**, 241–255.

Reinsch, C. H. (1967) Smoothing by spline functions. *Numer. Math.*, **10**, 177–183.

Rona, R. J. and Altman, D. G. (1977) National study of health and growth: standards of attained height, weight and triceps skinfold in English children 5 to 11 years old. *Ann. Hum. Biol.*, **4**, 501–523.

Rona, R. J. and Chinn, S. (1986) National Study of Health and Growth: social and biological factors associated with height of children from ethnic groups living in England. *Ann. Hum. Biol.*, **13**, 453–471.

Rosenbaum, S. (1988) 100 years of heights and weights. *J. R. Statist. Soc.* A, **151**, 276–309.

Royston, J. P. (1982) An extension of Shapiro and Wilk's *W* test for normality to large samples. *Appl. Statist.*, **31**, 115–124.

———(1986) A remark on AS 181: the *W* test for normality. *Appl. Statist.*, **35**, 232–234.

Tanner, J. M. (1986) Use and abuse of growth standards. In *Human Growth* (eds F. Falkner and J. M. Tanner), 2nd edn, vol. 3, pp. 95–109. New York: Plenum.

Tanner, J. M., Goldstein, H. and Whitehouse, R. H. (1970) Standards for children's height at ages 2–9 years allowing for height of parents. *Arch. Dis. Child.*, **45**, 755–762.

Tanner, J. M. and Whitehouse, R. H. (1975) Revised standards for triceps and subscapular skinfolds in British children. *Arch. Dis. Child.*, **50**, 142–145.

Tanner, J. M., Whitehouse, R. H. and Takaishi, M. (1966) Standards from birth to maturity for height, weight, height velocity, and weight velocity. British children, 1965. *Arch. Dis. Child.*, **41**, 454–471; 613–635.

Tukey, J. W. (1977) *Exploratory Data Analysis.* Reading: Addison-Wesley.

Van't Hof, M. A., Wit, J. M. and Roede, M. J. (1985) A method to construct age references for skewed skinfold data, using Box–Cox transformations to normality. *Hum. Biol.*, **57**, 131–139.

World Health Organization (1983) *Measuring Change in Nutritional Status*, pp. 1–101. Geneva: World Health Organization.