STATISTICS IN MEDICINE, VOL. 11, 1305–1319 (1992)

SMOOTHING REFERENCE CENTILE CURVES: THE LMS METHOD AND PENALIZED LIKELIHOOD

T. J. COLE

MRC Dunn Nutrition Unit, Downhams Lane, Milton Road, Cambridge CB4 1XJ, U.K.

AND

P. J. GREEN

Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K.

SUMMARY

Reference centile curves show the distribution of a measurement as it changes according to some covariate, often age. The LMS method summarizes the changing distribution by three curves representing the median, coefficient of variation and skewness, the latter expressed as a Box–Cox power. Using penalized likelihood the three curves can be fitted as cubic splines by non-linear regression, and the extent of smoothing required can be expressed in terms of smoothing parameters or equivalent degrees of freedom. The method is illustrated with data on triceps skinfold in Gambian girls and women, and body weight in U.S.A. girls.

INTRODUCTION

Reference centile curves are used widely in medical practice as a screening tool. They identify subjects who are unusual, in the sense that their value of some particular measurement, for example, height or weight or plasma prolactin, lies in one or other tail of the reference distribution. The need for centile curves, rather than a simple reference range, arises when the measurement is strongly dependent on some covariate, often age, so that the reference range changes with the covariate. The case for making the centile curves smooth is to some extent cosmetic – the centiles are more pleasing to the eye when smoothed appropriately – but there is also the underlying justification that physiologically, small changes in the covariate are likely to lead to continuous changes in the measurement, so that the centiles *ought* to change smoothly. In such cases, fitting discontinuous curves could lead to substantial bias.

The literature on fitting smooth centiles to reference data has mushroomed in the last few years.[1–9] The techniques discussed fall into two broad categories, those that require 'commonality' between adjacent centile curves, and those that do not. Commonality here means that the spacings between centiles are constrained to be related to each other, whereas without commonality the centiles can be entirely independent of their neighbours, to the extent that in the limit, adjacent curves may touch or even cross.

To ensure commonality, some form of distributional assumption has to be made, implicitly or explicitly. Healy *et al.*[1] expressed the assumption as a constraint, that the spacings between centiles be expressible as a low-order polynomial in the underlying standard deviation score (SD score or Z score). Pan *et al.*[6] extended this idea to a series of piecewise polynomials. Thompson

0277–6715/92/101305–15$12.50
© 1992 by John Wiley & Sons, Ltd.

Received August 1991
Revised January 1992

and Theron[7] by contrast fitted a family of Johnson curves to the data, with suitable smoothness constraints.

Cole[2,8] assumed an underlying skew normal distribution of the measurements, so that a suitable power transformation[10] would render the distribution normal. In the latter method, the distribution at each covariate value is summarized by three parameters, the Box–Cox power λ, the mean μ and the coefficient of variation σ, and the initials of the parameters give the name to the LMS method. The three parameters are constrained to change smoothly as the covariate changes, and can, like the centiles, be plotted against the covariate. Thus, one advantage of the LMS method is that the three curves, L, M and S, completely summarize the measurement's distribution over the range of the covariate, and in addition they may be of interest in their own right. Other distributionally based methods share this advantage.

A key assumption of the LMS method is that after a suitable power transformation, the data are normally distributed. Anthropometry measurements, particularly weight and height, tend to follow this pattern,[2,9] and many other variables are of the same form. The main problem with the assumption may be the presence of kurtosis, which the transformation does not adjust for, but kurtosis tends to be less important than skewness as a contributor to non-normality.

Green[3] highlighted the subjective and complex nature of Cole's fitting algorithm for the LMS method – in particular the need to group the data, essentially arbitrarily, according to the value of the covariate. He pointed out that the method of maximum penalized likelihood[11] could be used to provide smooth estimates of the L, M and S curves directly, thus avoiding the need to group the data, and the only input required from the user would be the choice of smoothing constants for the three curves.

The purpose of this paper is to describe in more detail the method of maximum penalized likelihood as applied to the LMS method.

METHODS

The LMS method

The variable of interest, denoted by y, is assumed positive. Suppose that y has median μ, and that y^λ (or if $\lambda = 0$, $\log_e(y)$) is normally distributed. It is then appropriate to consider the transformed variable

$$x = \frac{(y/\mu)^\lambda - 1}{\lambda}, \qquad \lambda \neq 0 \tag{1}$$

or

$$x = \log_e(y/\mu), \qquad \lambda = 0$$

based on the family of transformations proposed by Box and Cox.[10] This transformation maps the median μ of y to $x = 0$, and is continuous at $\lambda = 0$. For $\lambda = 1$ the standard deviation (SD) of x is exactly the coefficient of variation (CV) of y, and this remains approximately true for all moderate λ. The optimal value of λ is that which minimizes the SD of x.

Denoting the SD of x (and the CV of y) by σ, the SD score of x and hence of y is given by

$$\begin{aligned} z &= x/\sigma \\ &= \frac{(y/\mu)^\lambda - 1}{\lambda\sigma}, \qquad \lambda \neq 0 \end{aligned} \tag{2}$$

or

$$z = \frac{\log(y/\mu)}{\sigma}, \qquad \lambda = 0$$

and it is assumed that z has a standard normal distribution.

Assume now that the distribution of y varies with covariate t, and that λ, μ and σ at t are read off the smooth curves $L(t)$, $M(t)$ and $S(t)$. It follows that

$$z = \frac{[y/M(t)]^{L(t)} - 1}{L(t)S(t)}, \qquad L(t) \neq 0 \tag{3}$$

or

$$z = \frac{\log[y/M(t)]}{S(t)}, \qquad L(t) = 0.$$

Rearranging (3) shows that the centile 100α of y at t is given by

$$C_{100\alpha}(t) = M(t)(1 + L(t)S(t)Z_\alpha)^{1/L(t)} \qquad L(t) \neq 0 \tag{4}$$

or

$$C_{100\alpha}(t) = M(t)\exp[S(t)Z_\alpha] \qquad L(t) = 0$$

where Z_α is the normal equivalent deviate (NED) of size α. This shows that if the L, M and S curves are smooth, then so are the centile curves.

Maximum penalized likelihood

For the case of n independent observations $\{y_i\}$ at corresponding covariate values $\{t_i\}$, the log-likelihood function ℓ derived from (3) is given (apart from the constant) by

$$\ell = \ell(L, M, S) = \sum_{i=1}^{n} \left(L(t_i) \log \frac{y_i}{M(t_i)} - \log S(t_i) - \frac{1}{2} z_i^2 \right) \tag{5}$$

where $\{z_i\}$ are the SD scores corresponding to $\{y_i\}$. The curves $L(t)$, $M(t)$ and $S(t)$ are estimated by maximizing the penalized likelihood

$$\ell - \tfrac{1}{2}\alpha_\lambda \int \{L''(t)\}^2 \mathrm{d}t - \tfrac{1}{2}\alpha_\mu \int \{M''(t)\}^2 \mathrm{d}t - \tfrac{1}{2}\alpha_\sigma \int \{S''(t)\}^2 \mathrm{d}t \tag{6}$$

where α_λ, α_μ and α_σ are smoothing parameters. The three integrals provide roughness penalties according to the squared second derivatives of the L, M and S curves, so that maximizing (6) strikes a balance between fidelity to the data and smoothness of the L, M and S curves. See Silverman[12] and Green[11] for a fuller discussion of the principle. It can be shown that these forms of penalty lead to natural cubic splines with knots at each distinct value of t. Thus only the smoothing parameters α_λ, α_μ and α_σ need to be chosen in order to fit the model.

The log-likelihood function (6) can be maximized iteratively using Fisher scoring, with an updating reminiscent of ridge regression.[11] This involves deriving the first and second derivatives of ℓ with respect to L, M and S, denoted by u_λ and W_λ etc. See the Appendix for details.

The complexity of each fitted cubic spline curve, for example λ, is measured by its 'equivalent degrees of freedom' (e.d.f.),[14] defined as

$$\text{e.d.f.}_\lambda = \operatorname{trace}(W_\lambda + \alpha_\lambda K)^{-1} W_\lambda \tag{7}$$

calculated at convergence. This expression appears to imply a one-to-one correspondence between each α and its e.d.f., but in fact the matrices W_λ, W_μ and W_σ depend on all three fitted curves (8) and hence on all three α's. However, in practice it is found that each e.d.f. is substantially influenced only by its corresponding α, and very little by the other two.

Initial choices for the smoothing are obtained as follows:

$$\alpha_\mu = \frac{n\,r(t)^3}{400\,r(y)^2}$$

$$\alpha_\sigma = 2\bar{\mu}^2\alpha_\mu$$

$$\alpha_\lambda = \bar{\sigma}^4\alpha_\sigma,$$

where $r(t)$ is the range of t, $r(y)$ is the range of fitted y, and $\bar{\mu}$ and $\bar{\sigma}$ are typical values of μ and σ. The dependence on $r(t)$ and $r(y)$ provides the mathematically correct equivariance to scale and origin in the data, while the ratios between the α's are derived heuristically from consideration of the values of W_λ, W_μ and W_σ, and the constant $n/400$ is chosen empirically to provide e.d.f. of about 5 or 6 for approximately uniformly distributed t's. The empirical relationship between α and e.d.f. is approximately linear on a log-log scale, with a slope near to -5.

A FORTRAN program has been developed to implement the method. The outer loop iteration converges when the change in penalized log-likelihood is less than 0·01, while the inner loop criterion is for the sum of absolute changes in σ^* (9) to be less than 0·001. The degree of smoothing required can be specified either by α or by e.d.f. The latter proves more convenient in practice, and has also been suggested by Hastie and Tibshirani.[14]

DATA

Two sets of data are used to illustrate the method. The first is an anthropometry survey of 892 girls and women up to age 50 in three Gambian villages, seen during the dry season of 1989; 620 (70 per cent) of the subjects were aged under 20. There were 733 distinct ages in the dataset. Five anthropometric measurements were taken, but just the triceps skinfold is discussed here.

The second example consists of body weight in 4011 U.S. girls aged between 1 and 21 years, obtained as part of the American HANES1 Health and Nutrition Survey.[15] This dataset, consisting of 1657 distinct ages fairly uniformly distributed, was used by Cole[8] to illustrate the separate age group method of fitting the LMS method.

RESULTS

Figures 1 (*a*)–(*c*) show the L, M and S curves for triceps skinfold in Gambian females from birth to age 50 years, over a range of fitted e.d.f. between 3 and 15. For clarity the curves are offset from each other (by 0·1 units for L, 0·5 mm for M, and 0·03 units for S), and the central bold curve (with 9 e.d.f.) is the baseline. The fitted e.d.f. extend up to 15 to allow for sufficient detail to emerge in the curves during the growth phase.

The smoothest curves, with 3 e.d.f., are almost quadratic in shape, and provide a very poor fit to the data. As the e.d.f. increase the curves become more complex, and with 6 or 9 e.d.f. the broad shapes of the curves are clear. Conversely with 15 e.d.f. the curves are obviously undersmoothed. All the L curves (Figure 1 (*a*)) show a period around puberty when the Box–Cox power of the distribution dips sharply, below a value of $-0{\cdot}6$, indicating an increase in skewness to the right. The effect of smoothing is to flatten out the trough, so that it becomes wider and more diffuse as the e.d.f. fall. The M curves (Figure 1 (*b*)) all show high triceps values in early childhood, followed

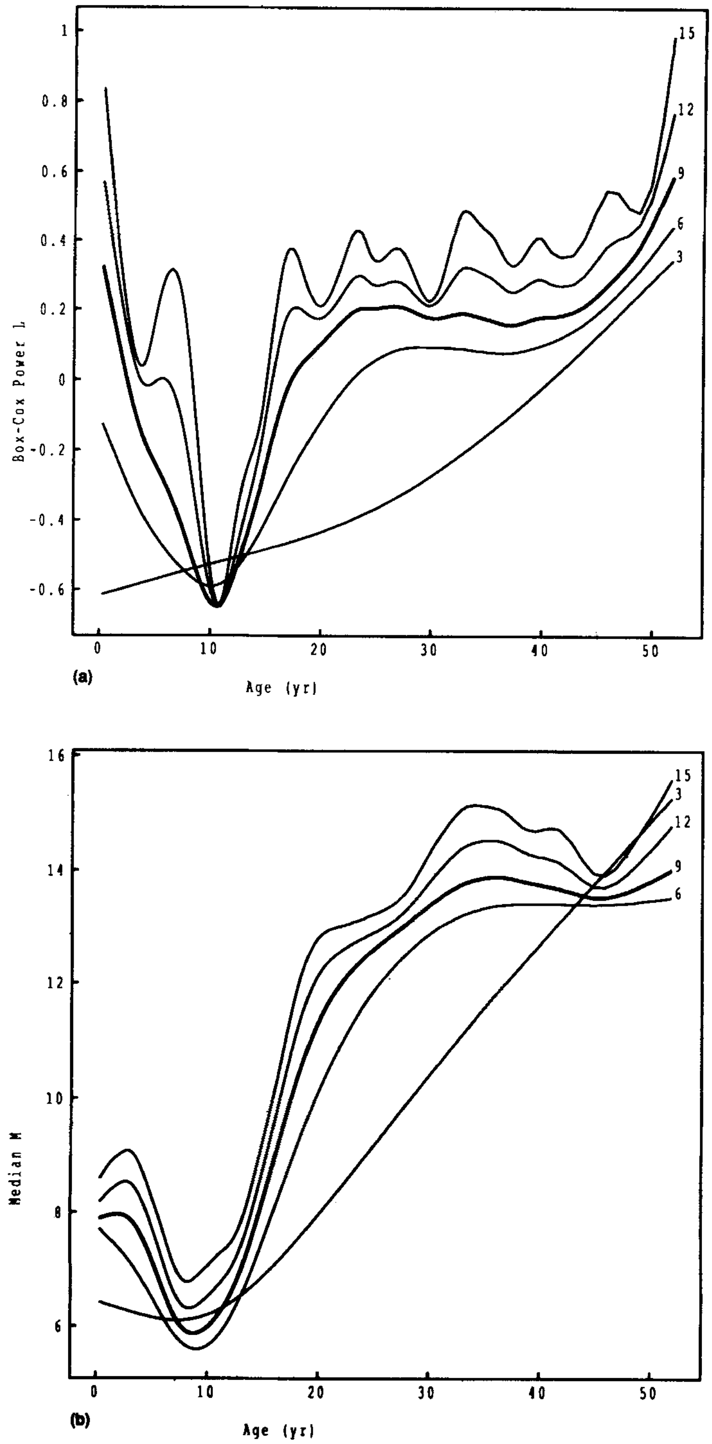
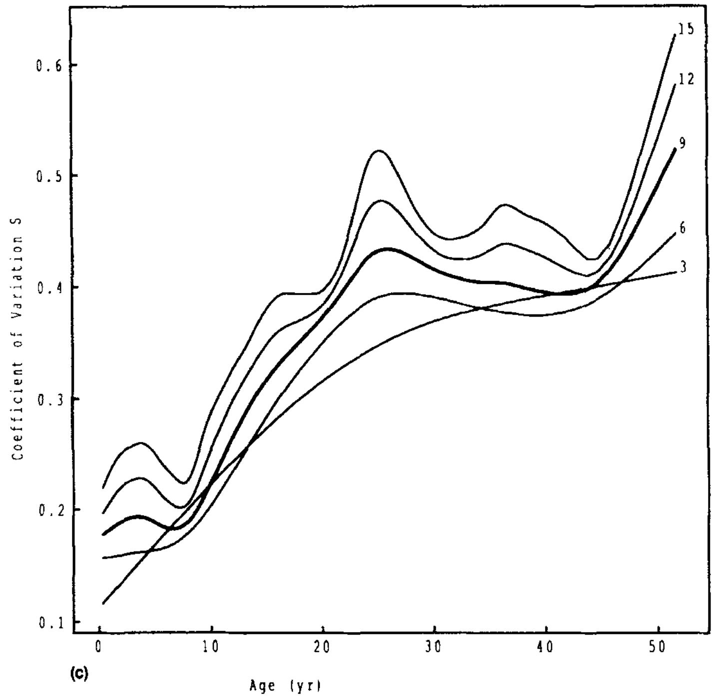


Figure 1. Triceps skinfold in Gambian females from birth to age 50 years, fitted by a series of spline curves with between 3 and 15 equivalent degrees of freedom

(*a*) Box–Cox power (*L*). The curves are offset by 0·1 units from their neighbours, with the baseline in bold

(*b*) Median (*M*). The curves are offset by 0·5 mm from their neighbours, with the baseline in bold

(*c*) Coefficient of variation (*S*). The curves are offset by 0·03 units from their neighbours, with the baseline in bold



(c)

Figure 1. (Continued)

by a fall and then a second rise continuing beyond age 30. The S curves (Figure 1 (c)) demonstrate that the coefficient of variation is raised during the early peak in triceps skinfold, and that it falls briefly before rising steeply during puberty.

Figures 2(a) and 2(b) show seven centiles, from the 3rd to the 97th, for triceps skinfold in Gambian females, obtained from the LMS curves using equation (4). In Figure 2(a) the curves each have 9 e.d.f., while in Figure 2(b) they have 12. Figure 2(a) is reasonably well smoothed over the whole age range, whereas in Figure 2(b) the centiles beyond age 20 are somewhat ragged. However, the corresponding centiles during childhood are convincingly smooth, showing that the extent of smoothing is greater for the children than the adults.

Figure 3 illustrates the empirical relationship between α and e.d.f. for the Gambian triceps skinfold data. The L, M and S curves are fitted in turn with e.d.f. of between 2 and 15, and the e.d.f. are then plotted against the corresponding values of α (7). On log-log axes the plots are close to linear and parallel, particularly for e.d.f. > 4. This same pattern is also found with other data (not shown), confirming the empirical relationship $\alpha = C\ \text{e.d.f.}^k$, where k usually lies between -4 and -6.

Figures 4(a)–(c) show the fitted L, M and S curves for U.S. girls' weight between 1 and 21 years, obtained by setting the e.d.f. to 7, 10 and 7, respectively. The S curve (Figure 4(c)) shows that the CV of weight increases until 12 years, the age of peak weight velocity, and then declines again, while the L curve (Figure 4(a)) demonstrates a slightly earlier period during puberty when the skewness of the weight distribution falls and then rises again, which is due to heterogeneity in the timing of the pubertal growth spurt.[8] Figure 5 gives the corresponding set of seven centile curves, from the 3rd to the 97th, obtained from Figures 4(a)–(c).

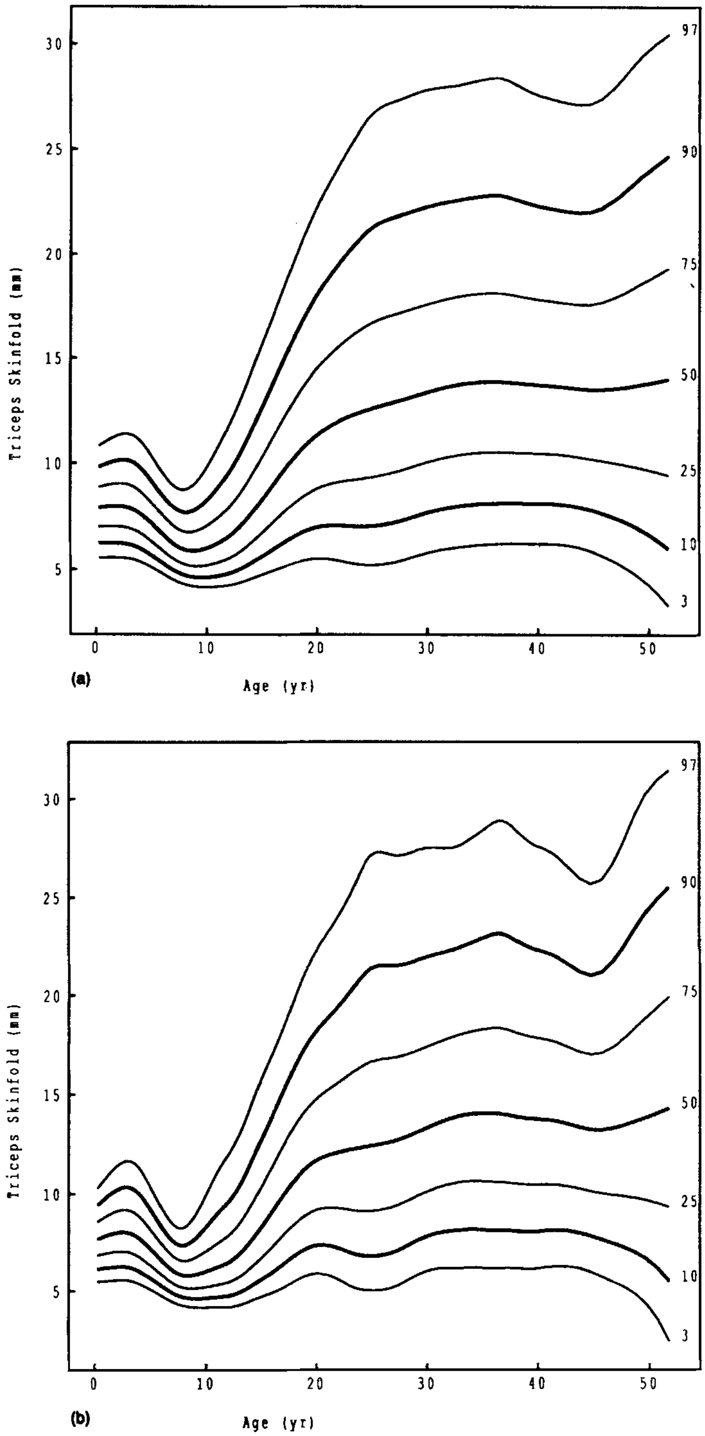


Figure 2. Seven centiles of triceps skinfold in Gambian females from birth to age 50 years, based on the LMS curves in Figure 1
(*a*) With 9 equivalent degrees of freedom (*b*) With 12 equivalent degrees of freedom

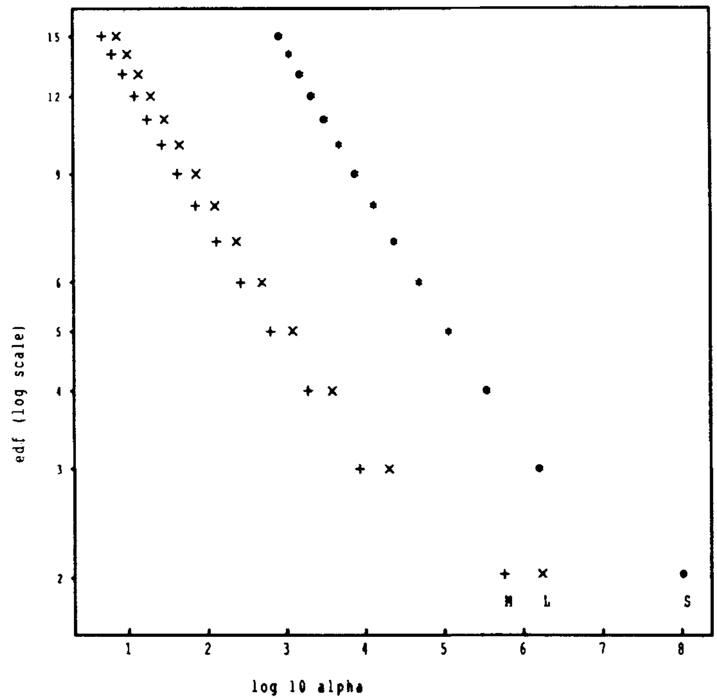
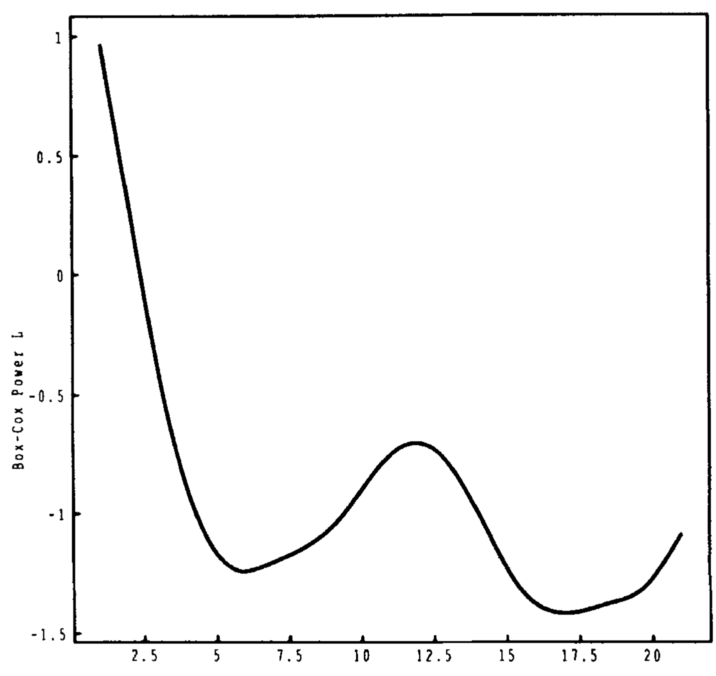


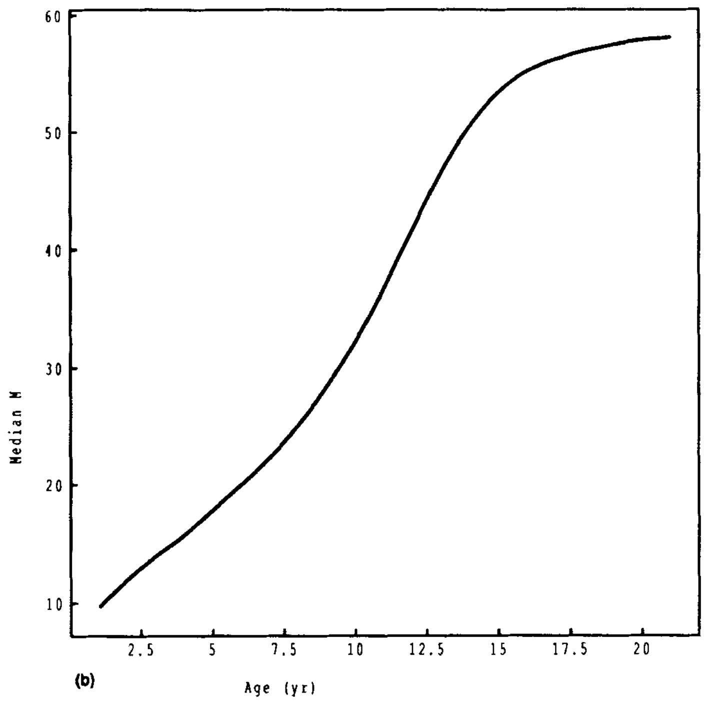
Figure 3. Plots of the smoothing parameter α versus the corresponding equivalent degrees of freedom (e.d.f.) for the LMS curves in Figure 1, for e.d.f. from 2 to 15



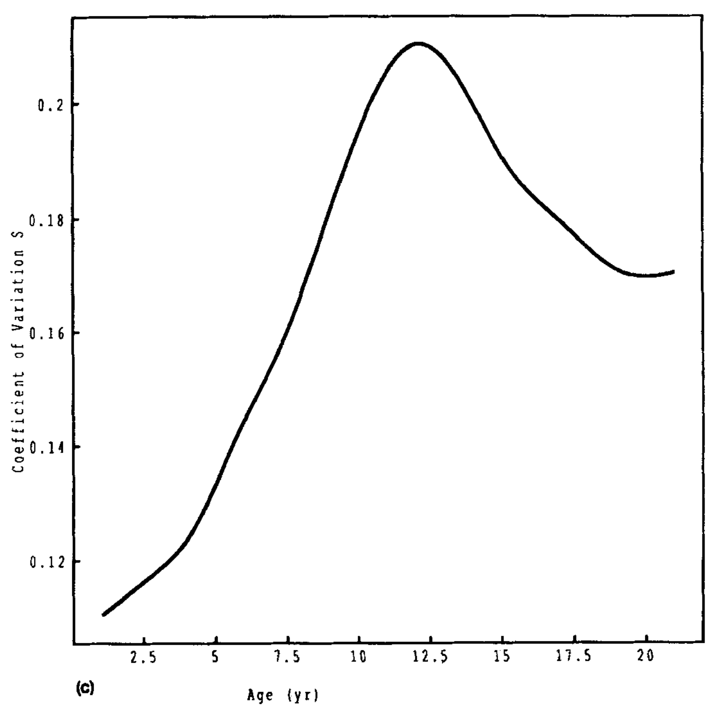
(a)

Figure 4. Weight in U.S. girls from age 1 to 21 years

(*a*) Box–Cox power (*L*) fitted by a spline curve with 7 equivalent degrees of freedom
(*b*) Median (*M*) fitted by a spline curve with 10 equivalent degrees of freedom
(*c*) Coefficient of variation (*S*) fitted by a spline curve with 7 equivalent degrees of freedom



(b)



(c)

Figure 4. (Continued)

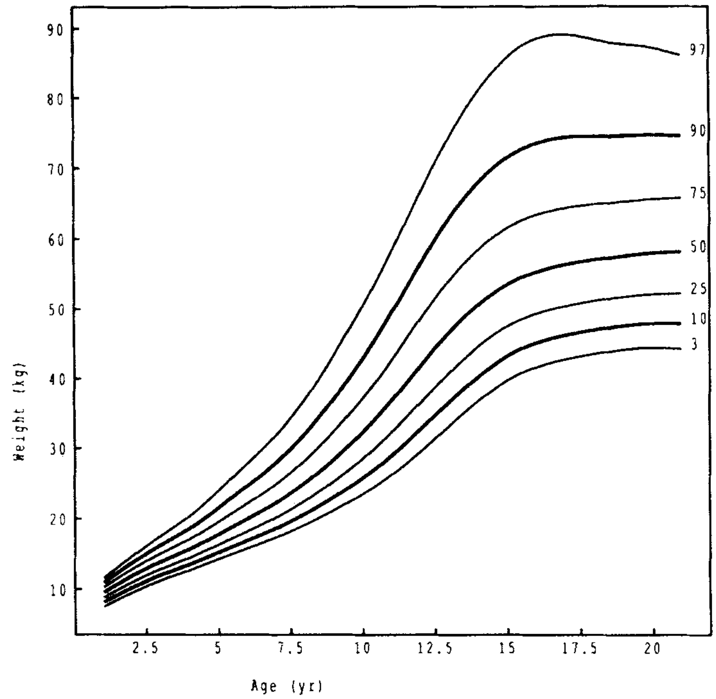


Figure 5. Seven centiles of weight in U.S. girls from age 1 to 21 years, based on the LMS curves in Figure 4

An important requirement of fitting centile curves is that they be properly calibrated, meaning that appropriate proportions of the sample fall between adjacent centiles at different ages. This is best done by expressing each measurement as an SD score (3), and if the fit is adequate they should be distributed as N(0, 1) throughout the age range.

Based on the LMS curves of Figures 4 (*a*)–(*c*) and (3), the mean SD score for U.S. girls' weight is 0·001 (SD 1·001). The distribution of SD scores is shown in Figure 6 using the model-free method of Healy;[1] seven empirical centile curves for SD score by age are derived by scatterplot smoothing using a bandwidth of 5 per cent. On the assumption of normality the expected centile curves are horizontal straight lines, which are also shown in Figure 6. It is clear that apart from random error the observed centiles are close to the expected values, with no systematic trends, and that the distribution is reasonably normal throughout the age range.

DISCUSSION

Fitting smooth centile curves has always been something of a subjective exercise, or even a black art.[2] The difficulty lies in deciding whether a bump or dip observed on a centile curve at a particular age is a real feature of the data, or whether it is simply sampling error. The LMS method as originally described went some way towards avoiding this problem, in that shapes of the centiles are determined by three essentially uncorrelated curves, the L, M and S curves. The first defines the skewness of the distribution at each age, the second the median and the third the coefficient of variation. Thus if a bump shows itself on the median curve, just one decision has to

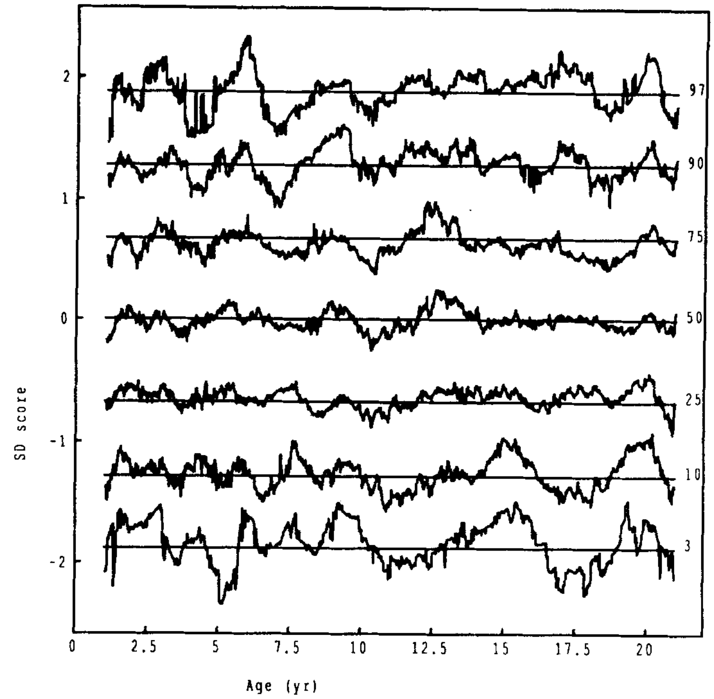


Figure 6. Seven observed and expected centiles of U.S. girls' weight SD scores, obtained from the standard of Figure 5. The observed centiles are derived by the scatterplot smoothing method of Healy *et al.*,[1] using a bandwidth of 200 points (5 per cent)

be made about its importance, whereas with seven centile curves, each being smoothed independently, the same decision has to be made repeatedly.

The examples shown here demonstrate that the distributions of Gambian triceps skinfold and U.S. body weight are both skew, with values of the Box–Cox power on the L curve falling well below zero (considerably more skew than a log transformation), and more importantly, that they vary appreciably with age. The S curve also provides useful information about the changing coefficient of variation across the age range.

The current paper describes an improved approach to the LMS method, which removes one arbitrary element from the fitting process and focuses on the smoothing. As originally described the method involved splitting the data into age groups, estimating L, M and S for each group separately, and then smoothing the group values across age. The choice of age cut-offs between groups was arbitrary, and in theory could have influenced the final result. In addition, the value for the power L read off the smoothed curve was not the value used to calculate the M and S values at the same age, so that the process did not iterate to convergence. Green[3] highlighted the problem, and pointed out that penalized likelihood could solve it.

In practice penalized likelihood provides an elegant solution – the smoothing of the three curves becomes an integral part of the likelihood maximization, with the roughness penalties incorporated with the likelihood (6). No age cut-offs need be specified, and the L, M and S values at each age are used in turn to calculate the other two parameters. Thus the only arbitrariness in the whole procedure is the choice of the three smoothing parameters α.

Even this arbitrariness could be reduced by developing a formal inference procedure for assessing goodness of fit. Hastie and Tibshirani[14] and Yandell and Green[16] have discussed inference in other non-parametric regression models using methods analogous to the F test. It would be possible to assess goodness of fit here by regarding twice the increase in log-likelihood for each unit increase in e.d.f. as χ^2 distributed, but there is no mathematical justification for this.

The benefit of the LMS method is shown by the second example. This dataset was also used by Cole (Reference 8; Figures 2–5), and the comparison of the curves fitted by age group, as there, and as fitted here is instructive. The main differences between the two sets of curves are minor edge effects in the L and S curves, but overall it is reassuring that the relatively crude method used by Cole[8] does not seriously distort the underlying distribution. An advantage of the penalized likelihood approach is that it avoids identifying individual age groups as good fits or outliers, and instead treats the entire dataset as a single entity.

The examples illustrate two practical problems of the method, which are quite general and likely to affect all smoothing techniques. The first is the presence of edge effects – in both examples, but particularly the Gambian women, the LMS curves turn up sharply at the top end of the age range, due to the relatively small numbers at this age. The second concern is the non-uniform smoothing seen in Figures 2 (*a*) and 2 (*b*), where the adults are less well smoothed than the children, due to the relatively small proportion of adults (30 per cent) in the sample. Both problems are ones of sampling – the age distribution should be essentially uniform throughout the age range, but perhaps with increased density at the extremes to minimise edge effects, or at other ages where the LMS curves change rapidly.[2]

The ability to express the smoothing parameters in terms of equivalent degrees of freedom[14] is a powerful feature of the method, as the α are in unfamiliar units and suitable default values are by no means obvious. Working with e.d.f. allows the required smoothness to be specified at the outset, and a given e.d.f. implies the same smoothness whatever dataset it is applied to. The smallest possible e.d.f. is 2, corresponding to a straight line. In this limiting case, the spline curve is the same shape as the polynomial with the same degrees of freedom. From Figures 1 (*a*)–(*c*) it is clear that a curve with 3 e.d.f. is similar to a quadratic, but as the e.d.f. increase the corresponding families of spline curves and polynomials become increasingly different. This emphasizes the benefit of using splines rather than polynomials – the possible range of shapes that can be fitted is much greater.

The essential linearity of the relationship between log e.d.f. and log α, illustrated in Figure 3, is useful for streamlining the iteration process. The exact relationship between e.d.f. and α (7) is complex, non-linear and data-dependent, so that the empirical linearity between them, given by $\alpha = C\ \mathrm{edf}^k$, is a bonus for the fitting process. An initial choice of α leads to its corresponding e.d.f., which estimates C in the equation. This can then be used to extrapolate to the required e.d.f.

A useful feature of the spline curves in the LMS method is that they can be used either to smooth the data or to investigate underlying structure. Sufficient has been said about the first application, but Figure 1 (*a*) provides an illustration of the second. The trough in the L curve during puberty is steep and narrow when fitted with 12 or 15 e.d.f., but as the e.d.f. are reduced so the trough becomes wider and shallower. The benefit of fitting the curve with 15 e.d.f. is that the underlying shape of the L curve is clearly seen. Apart from the trough, the curve is essentially constant throughout adulthood, except for the rise at age 50.

This illustrates a second feature of spline curve fitting. Looking at Figure 1(*a*), there is a temptation to model the L curve as a steep and narrow trough followed by a flat section. However, this would be an over-interpretation, as the later section only becomes flat by smoothing it, and this smoothing widens and flattens the trough at the same time. The spline curve provides an objective measure of the amount of weight to be attached to different features of the curve, and balances between them.

The U.S. girls' body weight example shows that the e.d.f. required for the M curve are greater than for the L and S curves. This is not surprising, as in general the median of the distribution is much better specified than the coefficient of variation or the skewness. It might be thought that in turn, the coefficient of variation ought to be better specified than the skewness, but in practice the two curves require much the same degree of smoothing, as seen in Figures 4(*a*) and 4(*c*) where they both have 7 e.d.f. The Gambian skinfolds by contrast use the same e.d.f. for all three curves, but this is somewhat different owing to the wider age range.

In conclusion, the LMS method fitted by penalized likelihood provides a convenient 'black box' for the fitting of smooth reference centile curves. It also highlights features of the underlying distribution as the covariate changes, and provides an objective tool to determine their relative importance.

APPENDIX

To deal with possible ties in the $\{t_i\}$, let $\{T_j\}$ be the m distinct ordered values of the $\{t_i\}$. The vectors of values of $L(T)$, $M(T)$ and $S(T)$ for the current iteration are then represented by λ, μ and σ. It can be shown that the corresponding value of $\int \{L''(T)\}^2 \mathrm{d}T$ is given by $\lambda^{\mathrm{T}} K \lambda$, and similarly for μ and σ, where K is an $m \times m$ square matrix dependent only on the $\{T_j\}$.[13]

Given λ, μ and σ at the current iteration, an improved set of estimates is derived by Fisher scoring. The updated estimates λ^*, μ^* and σ^* are solutions to the scoring equations:

$$\begin{bmatrix} W_\lambda + \alpha_\lambda K & W_{\lambda\mu} & W_{\lambda\sigma} \\ W_{\mu\lambda} & W_\mu + \alpha_\mu K & W_{\mu\sigma} \\ W_{\sigma\lambda} & W_{\sigma\mu} & W_\sigma + \alpha_\sigma K \end{bmatrix} \begin{bmatrix} \lambda^* - \lambda \\ \mu^* - \mu \\ \sigma^* - \sigma \end{bmatrix} = \begin{bmatrix} u_\lambda - \alpha_\lambda K \lambda \\ u_\mu - \alpha_\mu K \mu \\ u_\sigma - \alpha_\sigma K \sigma \end{bmatrix} \tag{8}$$

where

$$u_\lambda = \frac{\partial \ell}{\partial \lambda} = \frac{z}{\lambda}\left(z - \frac{\log(y/\mu)}{\sigma}\right) - \log(y/\mu)(z^2 - 1)$$

$$u_\mu = \frac{\partial \ell}{\partial \mu} = \frac{z}{\mu\sigma} + \frac{\lambda(z^2 - 1)}{\mu}$$

$$u_\sigma = \frac{\partial \ell}{\partial \sigma} = \frac{(z^2 - 1)}{\sigma}$$

$$W_\lambda = -E^*\left(\frac{\partial^2 \ell}{\partial \lambda \lambda^{\mathrm{T}}}\right) = \operatorname{diag}\left(\frac{7\sigma^2}{4}\right)$$

$$W_\mu = -E\left(\frac{\partial^2 \ell}{\partial \mu \mu^{\mathrm{T}}}\right) = \operatorname{diag}\left(\frac{(1 + 2\lambda^2\sigma^2)}{\mu^2\sigma^2}\right)$$

$$W_\sigma = -E\left(\frac{\partial^2 \ell}{\partial \sigma \sigma^{\mathrm{T}}}\right) = \operatorname{diag}\left(\frac{2}{\sigma^2}\right)$$

$$W_{\lambda\mu} = -E\left(\frac{\partial^2 \ell}{\partial \lambda \mu^{\mathrm{T}}}\right) = \operatorname{diag}\left(\frac{-1}{2\mu}\right)$$

$$W_{\lambda\sigma} = -E\left(\frac{\partial^2 \ell}{\partial \lambda \sigma^{\mathrm{T}}}\right) = \mathrm{diag}\,(\lambda\sigma)$$

$$W_{\mu\sigma} = -E\left(\frac{\partial^2 \ell}{\partial \mu \sigma^{\mathrm{T}}}\right) = \mathrm{diag}\left(\frac{2\lambda}{\mu\sigma}\right).$$

Note that the i suffices for y and z, and the j suffices for λ, μ and σ, have been dropped. Each of the final expressions above is interpreted in the following way: the jth element of each u vector and the (j, j)th element of each W matrix is the sum over those observations $\{i\}$ whose $\{t_i\}$ equals T_j. Derivatives of the penalized log-likelihood with respect to λ do not have finite expectations, so just the first three terms of a Taylor expansion of log (y/μ) are used to define the operator E^* for W_λ.

To solve these updating equations, it is convenient to eliminate pairs of λ^*, μ^* and σ^* in turn, giving

$$\lambda^* = (W_\lambda + \alpha_\lambda K)^{-1}\{u_\lambda + W_\lambda \lambda - (\mu^* - \mu) W_{\lambda\mu} - (\sigma^* - \sigma) W_{\lambda\sigma}\}$$
$$\mu^* = (W_\mu + \alpha_\mu K)^{-1}\{u_\mu + W_\mu \mu - (\sigma^* - \sigma) W_{\mu\sigma} - (\lambda^* - \lambda) W_{\mu\lambda}\}$$
$$\sigma^* = (W_\sigma + \alpha_\sigma K)^{-1}\{u_\sigma + W_\sigma \sigma - (\lambda^* - \lambda) W_{\sigma\lambda} - (\mu^* - \mu) W_{\sigma\mu}\}. \tag{9}$$

But the operation of computing $(W + \alpha K)^{-1} Wy$ from y is precisely that of fitting a cubic spline to pseudo-data y. Thus the updating equations (8) are solved by an inner iteration cycling around the three expressions in (9), applying Reinsch's algorithm[13] to re-estimate each of the curves in turn. Once this process stabilizes, λ, μ and σ are replaced by λ^*, μ^* and σ^*, the u and W are recalculated (8) and the process repeats.

ACKNOWLEDGEMENTS

The authors thank Dr. Sally Poppitt for access to the Gambian data, and the National Center for Health Statistics for the U.S. data. We are also grateful to Professor Michael Healy for his help in deriving Figure 6.

REFERENCES

1. Healy, M. J. R., Rasbash, J. and Yang, M. 'Distribution-free estimation of age-related centiles', *Annals of Human Biology*, **15**, 17–22 (1988).
2. Cole, T. J. 'Fitting smoothed centile curves to reference data (with Discussion)', *Journal of the Royal Statistical Society, Series A*, **151**, 385–418 (1988).
3. Green, P. J. In the Discussion to Cole (1988), *Journal of the Royal Statistical Society, Series A*, **151**, 410–411 (1988).
4. Jones, M. C. In the Discussion to Cole (1988), *Journal of the Royal Statistical Society, Series A*, **151**, 412–413 (1988).
5. Jones, M. C. and Hall, P. 'Mean squared error properties of kernel estimates of regression quantiles', *Statistics and Probability Letters*, **10**, 283–289 (1990).
6. Pan, H. Q., Goldstein, H. and Yang, Q. 'Nonparametric estimation of age related centiles over wide age ranges', *Annals of Human Biology*, **17**, 475–481 (1990).
7. Thompson, M. L. and Theron, G. B. 'Maximum likelihood estimation of reference centiles', *Statistics in Medicine*, **9**, 539–548 (1990).
8. Cole, T. J. 'The LMS method for constructing normalized growth standards', *European Journal of Clinical Nutrition*, **44**, 45–60 (1990).
9. Rossiter, J. E. 'Calculating centile curves using kernel density estimation methods with application to infant kidney lengths', *Statistics in Medicine*, **11**, 1693–1701 (1991).
10. Box, G. E. P. and Cox, D. R. 'An analysis of transformations', *Journal of the Royal Statistical Society, Series B*, **26**, 211–252 (1964).

11. Green, P. J. 'Penalized likelihood for general semi-parametric regression models', *International Statistical Review*, **55**, 245–259 (1987).
12. Silverman, B. W. 'Some aspects of the spline smoothing approach to non-parametric regression curve fitting', *Journal of the Royal Statistical Society, Series B*, **47**, 1–52 (1985).
13. Reinsch, C. H. 'Smoothing by spline functions', *Numerische Mathematik*, **10**, 177–183 (1967).
14. Hastie, T. and Tibshirani, R. 'Generalized additive models: some applications', *Journal of the American Statistical Association*, **82**, 371–386 (1987).
15. Hamill, P. V. V., Drizd, T. A., Johnson, C. L., Reed, R. B. and Roche, A. F. *NCHS Growth Curves for Children Birth–18 years*. National Center for Health Statistics, Vital and Health Statistics Series 11: No 165, Washington DC, 1977.
16. Yandell, B. S. and Green, P. J. 'Semi-parametric generalised linear model diagnostics', *Proceedings of the Statistical Computing Section, American Statistical Association*, 48–53 (1986).