STATISTICS IN MEDICINE, VOL. 16, 333–345 (1997)

CONSTRUCTION OF MULTIVARIATE CENTILE CHARTS FOR LONGITUDINAL MEASUREMENTS

MARY LOU THOMPSON

Department of Statistical Sciences, University of Cape Town, Rondebosch 7700, South Africa and Centre for Epidemiological Research, Medical Research Council, P.O. Box 19070, Tygerberg 7505, South Africa

AND

L. PAUL FATTI

Department of Statistics & Actuarial Science, University of the Witwatersrand, PO WITS, South Africa

SUMMARY

Reference centiles are commonly used as a means of identifying whether some measure of interest on an individual lies within a 'normal' range. Individuals outside the range may be at risk for some problem outcome. The information needed to screen for problem outcomes is often multivariate in nature. We develop a multivariate approach to centile estimation which allows for updating of centiles based on the prior path of the individual as well as adjustments to the centiles according to individual covariate values. It is also possible to consider several variables jointly as a screen for problem outcomes.

1. INTRODUCTION

Reference centiles are used in many areas of health research and practice as a means of establishing a 'normal' range and of identifying individuals who lie outside this range and hence may be subject to increased risk of morbidity and/or mortality. Much attention has focused in the recent statistical literature on the estimation of such centiles.[1–4]

The issue that we wish to address in this paper and which is also discussed in a companion paper[5] is that the information needed to screen for some problem outcome is often multivariate in nature and that it might be advantageous if this multivariate information could be simultaneously incorporated into the centile estimation. One obvious source of information that is neglected in most centile charts is the prior path of the individual. It would be likely, for instance, that an individual whose path has been moving along the 90th percentile and then drops to the 70th percentile will have a different prognosis from someone who has been moving along the 10th percentile and remains there. Conditional charts based on regression on the most recent observation have been used in charts for children's growth[6,7] and this represents a special case of our approach.

There is also often more than one longitudinal variable which impacts on the outcome of interest and it would be useful if the information from such variables could be combined, for screening use. Furthermore, static characteristics of the individual, that is, that do not change with follow-up, may also have impact on the centiles. Essentially in this last instance, one is considering the possibility of different centile charts for different strata of the population. For

CCC 0277–6715/97/040333–13

Received December 1994
Revised March 1996

example, in the pregnancy study which we will consider later, there might be different maternal weight centile charts depending on height, age and parity. The idea of customizing cross-sectional charts has been used in antenatal growth charts.[8]

In a companion paper,[5] we have developed a Bayesian approach to conditional centiles. Here, we will be considering a multivariate normal approach, of which the Bayesian approach can be viewed as a special case. We apply the ideas to a pregnancy study, where weight and fundal height of the mother were followed through the pregnancy as well as various demographic features being noted and the goal was to screen for problem pregnancies.

2. DEVELOPMENT OF THE MULTIVARIATE NORMAL APPROACH

We assume throughout that we have observations $\mathbf{X}_t^{(i)}, t = t_{i1}, \ldots, t_{iT_i}$, on n individuals, $i = 1, \ldots, n$, from some reference population, where at each time point, t, $\mathbf{X}_t^{(i)}$ is a $p \times 1$ vector of measures on p variables of interest on the ith individual who is observed at T_i time points. The time points at which the individuals are observed may be different for different individuals and need not be equispaced. The modelling approach used here is based on the multivariate normal distribution. If the raw observations are not normal, as is frequently the case, it is assumed that they have been transformed to normality. The Box–Cox family of transformations can be used to this end.[9] In the example of application considered in Section 3, the required transformations were not time-dependent. More generally, however, estimation of the transformation as a smooth function of time would also need to be incorporated into the modelling (see Cole[2]), with a corresponding increase in the number of parameters to be estimated.

2.1. A single follow-up variable

We assume here that for each individual i, $i = 1, \ldots, n$, $X_t^{(i)}$ is normally distributed and that $\mathbf{X}^{(i)} = (X_{t_{i1}}^{(i)}, \ldots, X_{t_{iT_i}}^{(i)})$ is multivariate normal. $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ $(i \neq j)$ are assumed to be independent.

It is generally biologically reasonable to assume that the parameters of the distribution of $\mathbf{X}^{(i)}$ change smoothly with t. One thus proceeds by modelling the mean, variance and covariance of $\mathbf{X}^{(i)}$, as smooth functions of t and perhaps static covariates, $\mathbf{u}_i$, that is,

$$E(X_t^{(i)}) = \mu_t^{(i)} = f_1(t, \mathbf{u}_i) \tag{1}$$

$$\mathrm{SD}(X_t^{(i)}) = \sigma_t^{(i)} = f_2(t, \mathbf{u}_i). \tag{2}$$

The standardized process

$$Z_t^{(i)} = (X_t^{(i)} - \mu_t^{(i)})/\sigma_t^{(i)}$$

is assumed to be covariance stationary, that is,

$$\mathrm{cov}(Z_t^{(i)}, Z_{t+h}^{(i)}) = f_3(h, \mathbf{u}_i).$$

Hence,

$$\mathrm{cov}(X_t^{(i)}, X_{t+h}^{(i)}) = f_3(h, \mathbf{u}_i)\sigma_t^{(i)}\sigma_{t+h}^{(i)}. \tag{3}$$

The assumption of covariance stationarity is justified in the example of application that we consider below. More general time-dependent covariance structures could be incorporated within the framework that we suggest, but with a corresponding increase in the modelling complexity.

Let $\mathbf{\Sigma}^{(i)}$ be the $T_i \times T_i$ variance-covariance matrix of $\mathbf{X}^{(i)}$, with $\sigma_t^{(i)2}$ on the diagonal and the off-diagonals defined by (3).

The joint likelihood of $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(n)}$ is then given by

$$L = \prod_{i=1}^{n} (2\pi\,|\boldsymbol{\Sigma}^{(i)}|)^{-(T_i/2)} \exp(-(\mathbf{X}^{(i)} - \boldsymbol{\mu}^{(i)})^t \boldsymbol{\Sigma}^{(i)^{-1}} (\mathbf{X}^{(i)} - \boldsymbol{\mu}^{(i)})/2) \tag{4}$$

where $\boldsymbol{\mu}^{(i)}$ and $\boldsymbol{\Sigma}^{(i)}$ involve the regression functions f_1, f_2 and f_3 above, the parameters of which can be estimated by the method of maximum likelihood.

The cross-sectional centiles of $X_t^{(i)}$ can then be estimated by substituting the estimated parameters into (1) and (2) above. Note that the incorporation of $\mathbf{u}_i$ results in cross-sectional centiles that are tailored to the characteristics of the individual. Examples of this will be provided in the application. The idea of customizing cross-sectional centiles has been used in, for example, antenatal growth charts.[8] The 10th centile for individual i (and individuals who share the characteristics $\mathbf{u}_i$) at time t would, for instance, be obtained from

$$Y_{10,t}^{(i)} = \mu_t^{(i)} - 1.28\,\sigma_t^{(i)}. \tag{5}$$

The centile estimate based on (5) above does not take into account the prior path of the individual leading up to time t. A straightforward application of multivariate normal theory leads to *conditional* centiles at time t, given the prior path of the ith individual, being estimated by substituting the maximum likelihood estimators into, for example

$$Y_{10,t}^{(i)} = \mu_t^{(i)} + \boldsymbol{\Sigma}_{tt\_}^{(i)} \boldsymbol{\Sigma}_{t\_t\_}^{(i)^{-1}} (\mathbf{X}_{t\_}^{(i)} - \boldsymbol{\mu}_{t\_}^{(i)}) - 1.28(\sigma_t^{(i)2} - \boldsymbol{\Sigma}_{tt\_}^{(i)} \boldsymbol{\Sigma}_{t\_t\_}^{(i)^{-1}} \boldsymbol{\Sigma}_{t\_t}^{(i)})^{0\cdot 5} \tag{6}$$

where $\boldsymbol{\Sigma}_{tt\_}^{(i)}$ is the variance-covariance matrix of $X_t^{(i)}$ and $\mathbf{X}_{t\_}^{(i)}$ where $\mathbf{X}_{t\_}^{(i)} = \{\mathbf{X}_s^{(i)}, s < t\}$ and $\boldsymbol{\Sigma}_{t\_t\_}^{(i)}$ is the variance-covariance matrix of $\mathbf{X}_{t\_}^{(i)}$.

In conventional centile charts, a single chart is constructed and the path of each individual is plotted on the chart with the prior path being incorporated at most informally in screening. The method outlined above allows updated centiles to be estimated for each individual as their path progresses. A further advantage of the conditional approach is that the longitudinal coverage of the centile bounds can be evaluated explicitly as it is simply the cross-sectional coverage raised to the power of the number of observations on the individual. By coverage we mean here the probability that the record of an individual from the reference population lies within the centile bounds. The longitudinal coverage of conventional unconditional centiles is unclear but is unlikely to be the nominal cross-sectional coverage. One would generally use higher conditional centiles to correspond to the same overall coverage as the unconditional ones. It should be noted, however, that, strictly speaking, all statements about coverage apply to the true, *unknown* theoretical centiles (be they conditional or unconditional), of which the centiles developed in this paper are estimates.

If the raw data have been transformed to achieve normality, the centiles in the original scale can easily be estimated by back-transforming (6).

2.2. Multiple longitudinal variables

The ideas developed in Section 2.1 above carry over to the case of multiple longitudinal variables. Here we observe p variables, $\mathbf{X}_{jt}^{(i)}$ $j = 1, \ldots, p$, on the ith individual at each time point t, and the joint distribution of $\mathbf{X}^{(i)}$ (now a vector of length pT_i) is assumed to be multivariate normal, with

$$E(X_{jt}^{(i)}) = \mu_{jt}^{(i)} = f_{1j}(t, \mathbf{u}_i)$$

$$\mathrm{SD}(X_{jt}^{(i)}) = \sigma_{jt}^{(i)} = f_{2j}(t, \mathbf{u}_i)$$

$$\mathrm{cov}(Z_{jt}^{(i)}, Z_{j,t+h}^{(i)}) = f_{3j}(h, \mathbf{u}_i)$$

$$\mathrm{cov}(Z_{jt}^{(i)}, Z_{k,t+h}^{(i)}) = f_{4jk}(h, \mathbf{u}_i)\ j \neq k.$$

Hence

$$\operatorname{cov}(X_{jt}^{(i)}, X_{j,t+h}^{(i)}) = f_{3j}(h, \mathbf{u}_i)\sigma_{jt}^{(i)}\sigma_{j,t+h}^{(i)} \tag{7}$$

and

$$\operatorname{cov}(X_{jt}^{(i)}, X_{k,t+h}^{(i)}) = f_{4jk}(h, \mathbf{u}_i)\sigma_{jt}^{(i)}\sigma_{k,t+h}^{(i)}\, j \neq k. \tag{8}$$

Let $\mathbf{\Sigma}_{jj}^{(i)}$ be the $T_i \times T_i$ variance-covariance matrix of $\mathbf{X}_j^{(i)}$, the vector of T_i measurements on the jth variable for the ith individual, with $\sigma_{jt}^{(i)2}$ on the diagonal and the off-diagonals defined by (7). Let $\mathbf{\Sigma}_{jk}^{(i)}$ be the $T_i \times T_i$ variance-covariance matrix of $\mathbf{X}_j^{(i)}$ with $\mathbf{X}_k^{(i)}, j \neq k$, defined by (8). The overall variance-covariance matrix for $\mathbf{X}^{(i)}$ is then the $T_i p \times T_i p$ matrix $\mathbf{\Sigma}^{(i)}$, which has $\mathbf{\Sigma}_{jj}^{(i)}$ on the diagonal $T_i \times T_i$ blocks and $\mathbf{\Sigma}_{jk}^{(i)}$ on the off-diagonals.

The joint likelihood of $\mathbf{X}^{(i)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(n)}$ is then given by

$$L = \prod_{i=1}^{n} (2\pi\,|\,\mathbf{\Sigma}^{(i)}\,|)^{-(pT_i/2)} \exp(-(\mathbf{X}^{(i)} - \boldsymbol{\mu}^{(i)})^t\, \mathbf{\Sigma}^{(i)^{-1}}(\mathbf{X}^{(i)} - \boldsymbol{\mu}^{(i)})/2) \tag{9}$$

where $\boldsymbol{\mu}^{(i)}$ and $\mathbf{\Sigma}^{(i)}$ now depend through f_1 to f_4 on the parameters, which can again be estimated by the method of maximum likelihood.

For each individual variable, j, the cross-sectional and conditional centiles of $X_{jt}^{(i)}$ can then be estimated by the obvious extension of the method described in Section 2.1. It is also possible to construct joint probability regions for the p variables simultaneously, using the fitted multivariate normal distribution. For example, an 80 per cent 'normal' reference region, which does not take account of the prior path, would be specified by the interior of the ellipsoid:

$$(\mathbf{X}_t^{(i)} - \boldsymbol{\mu}_t^{(i)})^t \mathbf{\Sigma}_t^{(i)-1}(\mathbf{X}_t^{(i)} - \boldsymbol{\mu}_t^{(i)}) = \chi_p^2(0{\cdot}80)$$

where $\mathbf{\Sigma}_t^{(i)}$ is the variance-covariance matrix of $X_{jt}^{(i)}, j = 1, \ldots, p$.

A similar region which is adjusted for the prior path of the p variables is given by

$$(\mathbf{X}_t^{(i)} - \boldsymbol{\mu}_{t.1}^{(i)})^t \mathbf{\Sigma}_{t.1}^{(i)-1}(\mathbf{X}_t^{(i)} - \boldsymbol{\mu}_{t.1}^{(i)}) = \chi_p^2(0{\cdot}80)$$

where

$$\boldsymbol{\mu}_{t.1}^{(i)} = \boldsymbol{\mu}_t^{(i)} + \mathbf{\Sigma}_{tt\_}^{(i)}\mathbf{\Sigma}_{t\_t\_}^{(i)-1}(\mathbf{X}_{t\_}^{(i)} - \boldsymbol{\mu}_{t\_}^{(i)})$$

$$\mathbf{\Sigma}_{t.1}^{(i)} = \mathbf{\Sigma}_t^{(i)} - \mathbf{\Sigma}_{tt\_}^{(i)}\mathbf{\Sigma}_{t\_t\_}^{(i)-1}\mathbf{\Sigma}_{t\_t}^{(i)}$$

and $\mathbf{X}_{t\_}^{(i)} = \{\mathbf{X}_{js}^{(i)}, j = 1, 2, \ldots, p, s < t\}$, $\boldsymbol{\mu}_{t\_}^{(i)} = \{\boldsymbol{\mu}_{js}^{(i)}, j = 1, 2, \ldots, p, s < t\}$, $\mathbf{\Sigma}_{tt\_}^{(i)}$ is the covariance matrix of $\mathbf{X}_t^{(i)}$ with $\mathbf{X}_{t\_}^{(i)}$ and $\mathbf{\Sigma}_{t\_t\_}^{(i)}$ is the variance-covariance matrix of $\mathbf{X}_{t\_}^{(i)}$.

Alternatively, one could combine the multiple variables into a single chart by considering the principal components of $\mathbf{X}_t^{(i)}$. Once the joint multivariate normal has been fitted, the conditional centiles of a principal component can be calculated using standard multivariate normal results. This approach will be illustrated in the example of application of the methodology which follows.

3. EXAMPLE OF APPLICATION

In a developing world setting where high technology equipment is not routinely available for antenatal assessment, it is particularly important to explore the possibility of monitoring pregnancy using simpler, more readily available measurements. Two such measurements are maternal weight and symphysis fundal height (SFH). Currently at Tygerberg Hospital in the Western Cape, both variables are measured at each antenatal clinic visit and plotted on separate centile charts.[10,11] Neither measurement is necessarily monotone increasing through the

pregnancy paths of individual women. Maternal weight may decrease from one antenatal visit to the next whilst still being associated with a 'normal' birth and a decrease in SFH can be the result of a decrease in liquor or a change in the lie of the fetus.

Part of our interest here was to explore possible ways of combining the two separate charts into a single screening tool as well as incorporating the prior path and covariate information into an adaptive chart. Also, ultimately, we wanted to establish whether these measures, either singly or jointly, were effective as a screen for problem pregnancies. The issue of whether weight and fundal height should be routinely recorded at each antenatal clinic visit is a contentious one[12–14] and one future aspect of this study will be to provide further insights on this issue. Is the routine measuring being undertaken in busy antenatal clinics serving a useful purpose?

One of the problem outcomes for which it might be desirable to screen is the birth of a small for gestational age (SGA) infant. In this study SGA is defined as being below the 10th percentile of the Dunn international reference chart.[15] Being born SGA is an established risk factor for increased morbidity and mortality.[16] The increased perinatal death rate of these babies is an immediate concern and spastic cerebral palsy may manifest in early childhood.[17] Recent attention has also focused on the long-term consequences in adulthood.[18] Early detection of poor intra-uterine growth may allow for interventions such as low dose aspirin therapy.[19]

A study was undertaken in the Western Cape, South Africa, where the antenatal records of 700 women who had 'normal' deliveries were drawn retrospectively from the Tygerberg Hospital records and used to define a reference population. Here 'normal' delivery was defined as:

(i) the subject gave birth after 37 weeks, that is, the baby was not preterm;
(ii) the birthweight of the infant was $\geqslant 2500$ g, that is, the baby was not low birthweight;
(iii) the baby was not small for gestational age, where this is defined as falling below the 10th percentile of the birthweight chart constructed by Dunn;[15]
(iv) there were no other complications (such as diabetes, pre-eclampsia).

This choice of reference sample is a debatable one. We acknowledge that there are arguments in favour of fewer exclusions, but our idea was to attempt to characterize the *path* of pregnancies whose passage had been in some sense ideal, with the hope of being able to better identity problem pregnancies via deviations from this normal path.

For each woman, her height, age and parity were recorded and her weight and fundal height were followed up at each antenatal clinic visit. It must be emphasized that these are actual clinic records and not from a controlled study, it being the context that actually exists in practice which we are ultimately interested in assessing.

3.1. Construction of centiles for maternal weight during pregnancy

The distribution of the variable weight at each week of pregnancy (gestational age) was non-normal, but it was established that the inverse transformation of weight yielded a distribution which was normal at each gestational age. The Shapiro–Wilk test was used to assess normality.[20] The development that follows is concerned with the modelling of inverse weight, which will be denoted by X_1.

Exploratory analyses (see Figures 1(*a*)–(*c*)) suggested the following models for the parameters of the distribution of X_1. It should be noted in Figures 1(*a*) and (*b*) that the number of observations at each gestational age varies from fewer than 20 in weeks 14, 15 and 42 to over 400 in weeks 36–38.

$$E(X_{1t}^{(i)}) = \mu_{1t}^{(i)} = \alpha_1 + \alpha_2 t + \alpha_3 \,\mathrm{HT}_i + \alpha_4 \,\mathrm{AGE}_i + \alpha_5 \,\mathrm{AGE}_i \, t + \alpha_6 \,\mathrm{PARITY}_i.$$

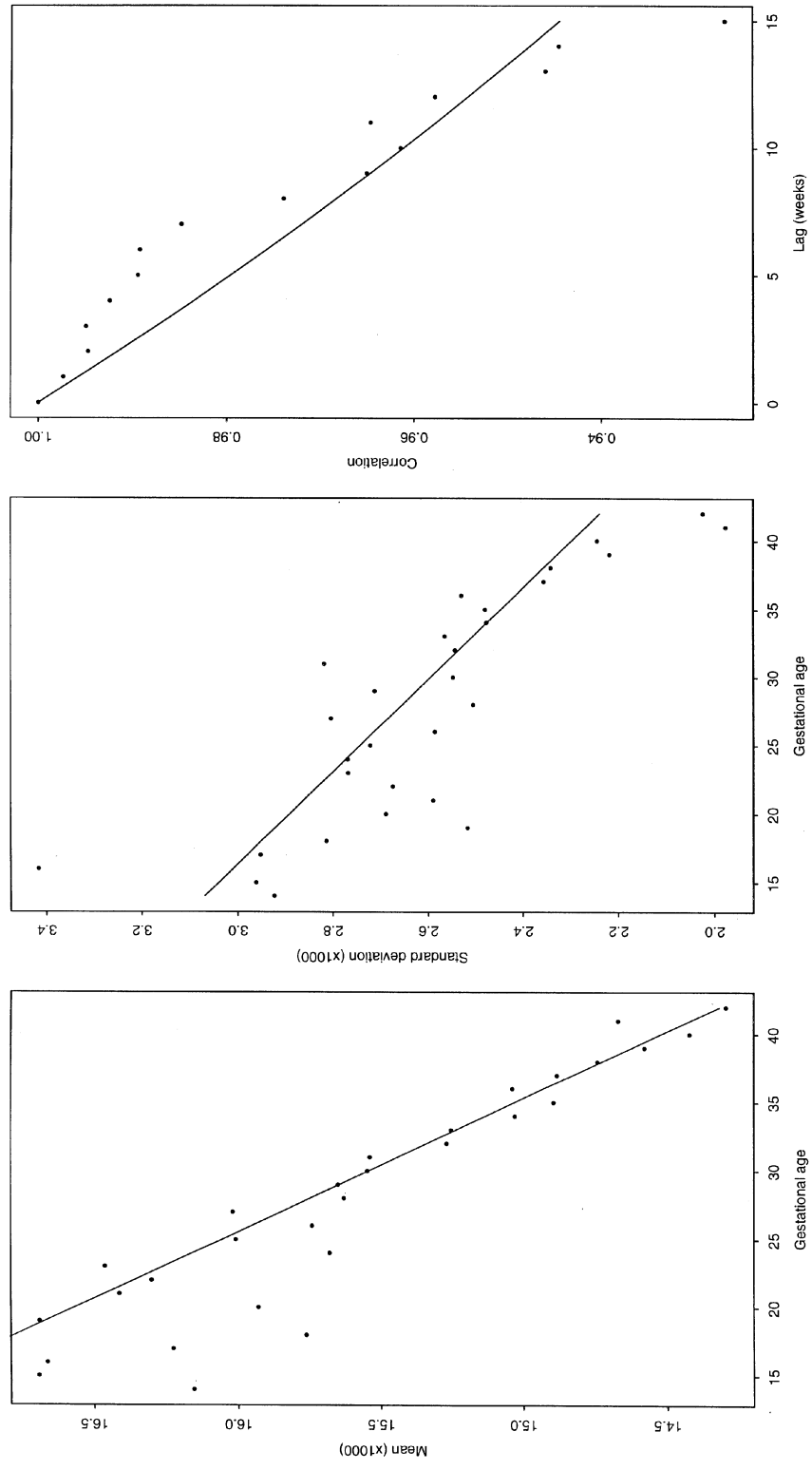

Figure 1. Exploratory modelling for parameters of inverse weight: (*a*) mean; (*b*) standard deviation; (*c*) correlation

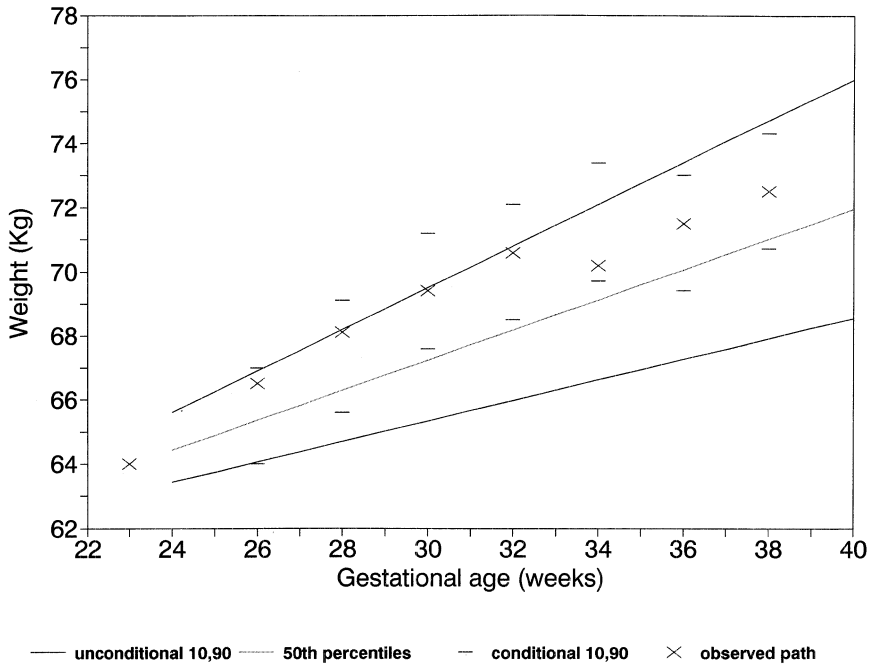

Figure 2. Centiles for maternal weight
(Note that the conditional centiles are based on the observed path as plotted, up to the epoch previous to the current point in time)

This model adjusts the overall location of the chart according to the individual's height (HT), age (AGE) and parity (PARITY) and also allows the slope over time to change with age. The variable parity is zero when parity is zero and 1 otherwise:

$$\mathrm{SD}(X_{1t}^{(i)}) = \sigma_{1t}^{(i)} = \alpha_7 + \alpha_8 t$$

$$\mathrm{cov}(Z_{1t}^{(i)}, Z_{1,t+h}^{(i)}) = 1 - \alpha_9[1 - \exp(-h/\alpha_{10})].$$

The functional form for the covariance structure is that of an exponential variogram model[21] and is chosen so as to ensure positive definiteness.

The multivariate normal likelihood for the reference sample using the above parameterization was fitted using the IMSL routine DUMINF.[22] The parameter estimates are given below:

$$\mu_{1t}^{(i)} = 0{\cdot}0430 - 1{\cdot}81 \times 10^{-4} t - 1{\cdot}18 \times 10^{-4}\,\mathrm{HT}_i - 2{\cdot}31 \times 10^{-4}\,\mathrm{AGE}_i + 3{\cdot}15 \times 10^{-6}\,\mathrm{AGE}_i\, t - 2{\cdot}35 \times 10^{-4}\,\mathrm{PARITY}_i$$

$$\sigma_{1t}^{(i)} = 0{\cdot}00298 - 2{\cdot}39 \times 10^{-5} t$$

$$\mathrm{corr}(X_{1t}^{(i)}, X_{1,t+h}^{(i)}) = 1 - 0{\cdot}216[1 - \exp(-h/50{\cdot}2)].$$

Whilst all available data was used in the above maximum likelihood estimation of the model parameters, the fitted models (and resulting centiles) were only implemented from week 18 of pregnancy onwards, where we had sufficient data to be confident of the fit of the models.

Figure 2 shows the maternal weight chart (10, 50, 90 per cent) that was derived from an earlier study[3,10] (this is essentially a chart conditional on the first observed weight), with

the conditional centiles from the current modelling superimposed. The unconditional centiles for weight are, as is to be expected, very wide and not particularly illuminating. It is rather the *gain* in weight that one is wanting to monitor. Also shown is the actual weight path of a 23 year old, nulliparous woman, height 161 cm who *did* give birth to an SGA infant and whose record commences in week 23 of her pregnancy at which point she weighted 64 kg. This woman is initially on a 'high' weight gain path and the conditional centiles adapt accordingly. The conditional centiles reflect the 'normal' range of weight, given the woman's prior path. The woman's weight path at no point lies outside the cross-sectional bounds or conditional bounds. It should be noted that one would, in practice, use wider conditional centile bounds to achieve the same overall coverage as the unconditional chart. The coverage of the conditional chart is multiplicative, and, for a fixed number of clinic visits, the coverage is hence clearly defined. As noted earlier, it must be kept in mind that these comments on coverage apply to the true centiles, of which those developed here are estimates. It is not possible to make clear-cut coverage statements for the estimated centiles.

3.2. Construction of centiles for fundal height during pregnancy

The distribution of the variable fundal height at each gestational age was also non-normal and it was established that the logarithmic transformation of fundal height yielded a distribution which was normal at each gestational age. The development that follows is then concerned with the modelling of the log transformed variable, which will be denoted by X_2. It should be noted that, whilst for both weight and fundal height the same transformation to normality could be used at each gestational age, this is a fortunate feature of these data and not an inherent constraint of the approach which can accommodate smoothly varying transformations over time.

Exploratory analyses (see Figures 3(*a*)–(*c*)) suggested the following models for the parameters of X_2. It should again be noted in Figures 3(*a*) and (*b*) that the number of observations at each gestational age varies from fewer than 20 in weeks 14, 15 and 42 to over 400 in weeks 36–38.

$$E(X_{2t}^{(i)}) = \mu_{2t}^{(i)} = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 \,\mathrm{HT}_i + \beta_5 \,\mathrm{AGE}_i.$$

This model adjusts the overall location of the chart according to the height and age of the individual:

$$\mathrm{SD}(X_{2t}^{(i)}) = \sigma_{2t}^{(i)} = \beta_6 + \beta_7 t + \beta_8 t^2$$

$$\mathrm{cov}(Z_{2t}^{(i)}, Z_{2,t+h}^{(i)}) = \exp(-h^{\beta_9}/\beta_{10}).$$

The functional form for the covariance structure is that of the so-called stable covariance model[23] and is again chosen so as to ensure positive definiteness.

The multivariate normal likelihood for the reference sample using the above parameterization was again fitted using the IMSL routine DUMINF.[22] The parameter estimates are given below:

$$\mu_{2t}^{(i)} = 1{\cdot}41 + 0{\cdot}0986\, t - 0{\cdot}00108\, t^2 - 0{\cdot}000364\,\mathrm{HT}_i + 0{\cdot}00167\,\mathrm{AGE}_i$$

$$\sigma_{2t}^{(i)} = 0{\cdot}411 - 0{\cdot}0210 t + 0{\cdot}000303 t^2$$

$$\mathrm{corr}(X_{2t}^{(i)}, X_{2,t+h}^{(i)}) = \exp(-h^{0{\cdot}604}/3{\cdot}23).$$

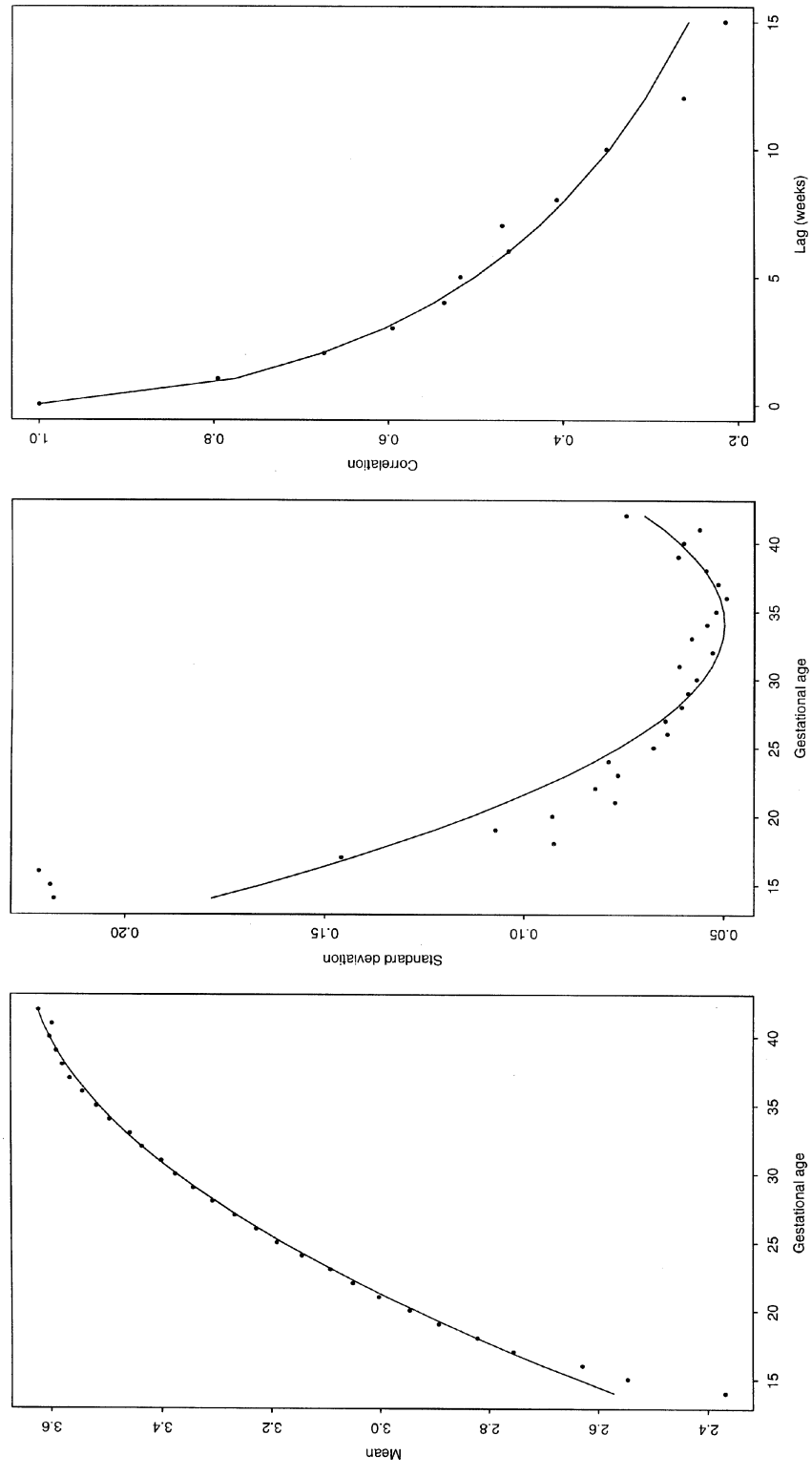

Figure 3. Exploratory modelling for parameters of log(SFH): (*a*) mean; (*b*) standard deviation; (*c*) correlation

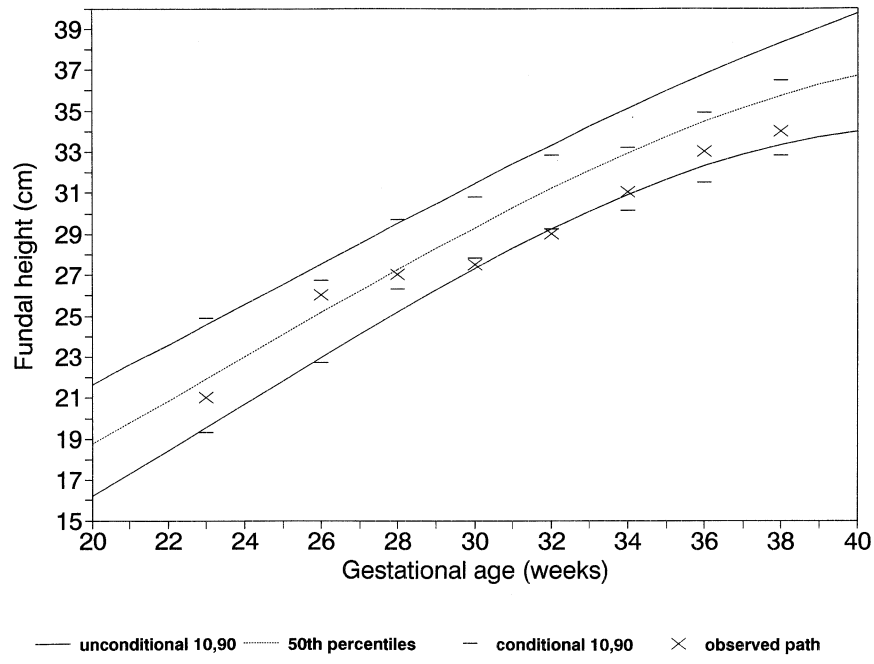

Figure 4. Centiles for symphysis fundal height
(Note that the conditional centiles are based on the observed path as plotted, up to the epoch previous to the current point in time)

Once again, whilst all available data was used in the above maximum likelihood estimation of the model parameters, the fitted models (and resulting centiles) were only implemented from week 18 of pregnancy onwards, where we had sufficient data to be confident of the fit of the models. The fit of the model for the mean is not entirely satisfactory for the earlier gestational ages and could be improved by use of, say, a fractional polynomial model.[24]

A cross-sectional chart (10, 50, 90 per cent) for the fundal height of the same woman considered in Figure 2 is given in Figure 4. Superimposed on this figure is the actual observed fundal height path of this woman and the corresponding conditional centiles. It can once again clearly be seen how the conditional centiles adjust to the woman's prior information and so provide a narrower confidence band. The woman's record lies outside the conditional centiles in week 30 and week 32 and she would also be flagged by the unconditional centiles in week 32. The sudden flattening of this woman's SFH path is a typical indication of growth retardation and is picked up more quickly by the conditional chart.

3.3. Construction of joint multivariate centiles of weight and fundal height

In the joint modelling of (transformed) weight and fundal height, we established that a single parameter would suffice to model the correlation of weight with fundal height:

$$\mathrm{cov}(Z_{1t}^{(i)}, Z_{2,t+h}^{(i)}) = \gamma$$

that is, it was assumed that the correlation between weight and fundal height was independent of lag.

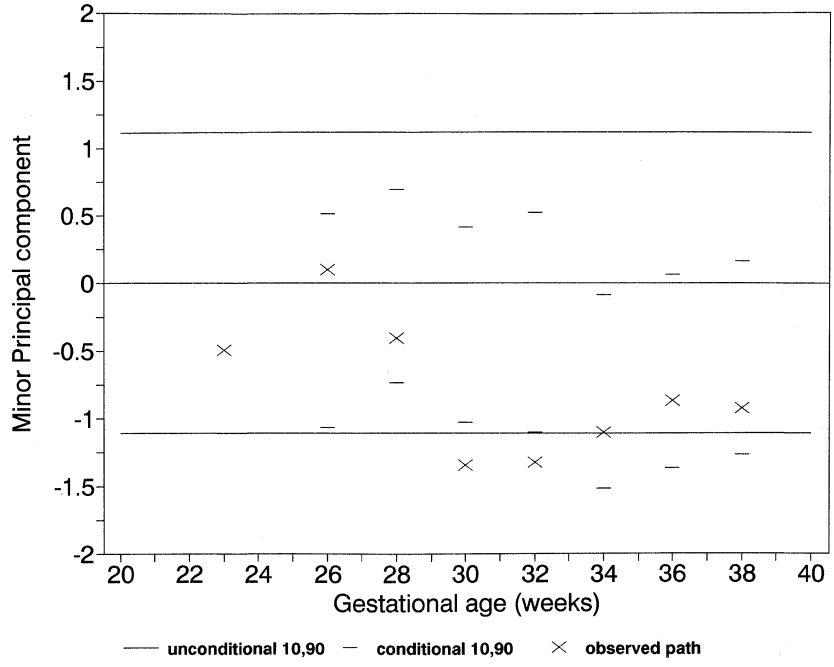

Figure 5. Centiles for minor principal component of (transformed) weight and fundal height
(Note that the conditional centiles are based on the observed path as plotted, up to the epoch previous to the current point in time)

The parameters of the joint model were again estimated by maximum likelihood and did not differ substantially from those fitted to the variables individually. The fitted correlation between (transformed) weight and fundal height was $-0{\cdot}246$ (the exploratory estimates, based on the correlation between (transformed) weight and SFH were, for example, $-0{\cdot}28$ at lag 0, $-0{\cdot}32$ at lag 6 and $-0{\cdot}24$ at lag 20). This implies a positive association on the raw scale, as would be expected.

The major and minor principal components were calculated as follows:

$$P_t^{(i)} = \frac{1}{\sqrt{2}}\left(\frac{X_{1t}^{(i)} - \mu_{1t}^{(i)}}{\sigma_{1t}^{(i)}} \pm \frac{X_{2t}^{(i)} - \mu_{2t}^{(i)}}{\sigma_{2t}^{(i)}}\right).$$

The unconditional cross-sectional distribution of $P_t^{(i)}$ is $N(0, 1 \pm \gamma)$ and the conditional centiles can be calculated from standard multivariate normal results. Figure 5 shows the path of the minor principal component for the same woman who was considered in Figures 2 and 4. Interestingly, both the unconditional and conditional centile charts identify this woman's path as potentially problematic in weeks 30 and 32.

4. DISCUSSION

While centile charts are sometimes constructed merely to establish ranges of typical values for academic interest, more frequently they are used as a means of identifying problems, particularly when the path of an individual is followed longitudinally. The determinants of problem outcomes

are generally complex and it might be envisaged that one would have a more sensitive screening tool if information from several variables which are believed to contribute towards the problem outcome are viewed jointly. It is also most plausible that the position of an individual's path relative to past values carries information as to the 'normality' of path.

The methodology that we have developed allows one to incorporate both prior information on a particular variable as well as measurements on several variables in following the path(s) of an individual. In the case of pregnancy, the aim is to facilitate early warning of problem outcomes such as SGA births, perinatal morbidity and hypertensive pregnancies. Particularly in the developing world context, access to high technology equipment such as ultrasound is frequently not available and it would be extremely helpful if the monitoring of easy to measure variables such as weight and fundal height would prove to be useful as screening tools for problem pregnancies.

In a further arm of this study, we intend to estimate the sensitivity and specificity of the conditional charts presented here in identifying various problem outcomes. This information will be particularly useful to planners in antenatal clinics in developing world settings. If these variables are not useful screening tools, much time can be saved in abandoning these routine measurements. More optimistically, if the charts provide a screening tool with satisfactory sensitivity and specificity we will have identified a useful and accessible way of early problem pregnancy diagnosis.

There are many other areas where conditional centile charts may prove useful. Monitoring of blood pressure during pregnancy is one example. There is a wide range of 'normal' blood pressures and so information on the prior path is likely to prove advantageous.

The principal component approach is attractive in that it reduces multivariate information to a single dimension. It is also potentially a more powerful method of diagnosis in that information from several variables is being used simultaneously. The drawback is that the centiles are no longer in units which are meaningful to the practitioner and hence their use becomes mechanistic, without the possibility for interaction. Once again, in any particular context, the usefulness of a principal component chart as a screen for problem outcomes would have to be assessed.

ACKNOWLEDGEMENTS

The authors wish to acknowledge support from grants from the South African Medical Research Council and the Foundation for Research Development. We are indebted to the meticulous data collection of Suzanne van Zyl and Anneke Theron, and also for fruitful discussions with Dr. Gerhard Theron and other members of the Medical Research Council's Perinatal Mortality Research Unit. We wish to acknowledge the valuable suggestions of two anonymous referees.

REFERENCES

1. Van't Hof, M. A., Wit, J. M. and Roede, M. J. 'A method to construct age references for skewed skinfold data, using Box–Cox transformations to normality', *Human Biology*, **57**, 131–139 (1985).
2. Cole, T. J. 'Fitting smoothed centile curves to reference data', *Journal of the Royal Statistical Society, Series A*, **151**, 385–418 (1988).
3. Thompson, M. L. and Theron, G. B. 'Maximum likelihood estimation of reference centiles', *Statistics in Medicine*, **9**, 539–548 (1990).
4. Altman, D. G. 'Construction of age-related reference centiles using absolute residuals', *Statistics in Medicine*, **12**, 917–924 (1993).
5. Fatti, L. P., Thompson, M. L. and Senaoana, E. M. Technical Report, Department of Statistics & Actuarial Science, University of the Witwatersrand, P.O. Wits, 2050, South Africa (1996).
6. Cole, T. J. 'The use and construction of anthropometric growth reference standards', *Nutrition Research Reviews*, **6**, 19–50 (1993).

7. Cole, T. J. 'Statistical constructs of human growth: new growth charts for old' *in* Ulijaszek, S. J. and Mascie-Taylor, C. G. N. (eds), *Anthropometry*: *the Individual and the Population*, Cambridge University Press, 1994, pp. 78–98.
8. Gardosi, J., Chang, A., Kalyan, B., Sahota, D. and Symonds, E. M. 'Customised antenatal growth charts', *Lancet*, **339**, 283–287 (1992).
9. Box, G. E. P. and Cox, D. R. 'An analysis of transformations', *Journal of the Royal Statistical Society*, *Series B*, **26**, 211–252 (1964).
10. Theron, G. B. and Thompson, M. L. 'A centile chart for weight gain in pregnancy for the urban population of the Western Cape, South Africa', *International Journal of Gynecology and Obstetrics*, **33**, 127–133 (1990).
11. Belizan, J. M., Villar, J., Nardin, J. C., Malamud, J. and Sainz de Vicuna, L. 'Diagnosis of intrauterine growth retardation by a simple clinical method: Measurement of uterine height', *American Journal of Obstetrics and Gynecology*, **131**, 643–646 (1978).
12. Taffel, S. M. and Keppel, K. G. 'Advice about weight gain during pregnancy and actual weight gain', *American Journal of Public Health*, **76**, 1396–1399 (1986).
13. Dawes, M. G. and Grudzinskas, J. G. 'Repeated measurement of maternal weight during pregnancy. Is this a useful practice?' *British Journal of Obstetrics and Gynaecology*, **98**, 189–194 (1991).
14. Rosenberg, K., Grant, J. M., Tweedie, I., Aitchison, T. and Gallagher, F. 'Measurement of fundal height as a screening test for fetal growth retardation', *British Journal of Obstetrics and Gynaecology*, **89**, 447–450 (1982).
15. Dunn, P. M. 'A perinatal growth chart for international reference', *Acta Paediatrica Scandinavica*, Supplement, **319**, 180–187 (1985).
16. Dobson, P. C., Able, D. A. and Beischer, N. G. 'Mortality and morbidity of fetal growth retardation', *Australian and New Zealand Journal of Obstetrics and Gynaecology*, **21**, 69–73 (1981).
17. Blair, E. and Stanley, F. 'Intrauterine growth and spastic cerebral palsy. I. Association with birth weight for gestational age', *American Journal of Obstetrics and Gynecology*, **162**, 229–237 (1990).
18. Hytten, E. 'Longterm consequences of fetal deprivation', *British Journal of Obstetrics and Gynaecology*, **97**, 665–666 (1990).
19. Trudinger, B. J., Cook, C. M., Thompson, R. S., Giles, W. B. and Connely, A. 'Low-dose aspirin therapy improves fetal weight in umbilical placental insufficiency', *American Journal of Obstetrics and Gynecology*, **160**, 681–685 (1988).
20. Royston, J. P. 'An extension of Shapiro and Wilk's W test for normality to large samples', *Applied Statistics*, **31**, 115–124 (1982).
21. Cressie, N. *Statistics for Spatial Data*, Wiley, New York, 1991.
22. *IMSL User's Manual*, Math/Library, Houston, Texas, 1987.
23. Yaglom, A. M. *Correlation Theory of Stationary and Related Random Functions*, Springer-Verlag, Berlin, 1986.
24. Royston, P. and Altman, D. G. 'Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling', *Applied Statistics*, **43**, 429–453 (1994).