The use (and abuse) of reference centiles with an application to weight gain in pregnancy

Mary Lou Thompson, L Paul Fatti

Centile charts are commonly used in many areas of health research and practice, e.g. growth charts for children, Doppler ultrasonography in pregnancy and assessment of cholesterol levels at different ages. Yet there are a number of aspects of both their construction and application that are problematic and it is some of these issues that will be raised in this paper.

The objective of the paper is to outline, in a non-technical way, some of the issues that need to be considered by the practitioner in estimating and using reference centile charts, but which frequently are either not known or ignored. These include: (*i*) the choice of reference population; (*ii*) how to estimate centiles; (*iii*) formally incorporating previous measurements on an individual, e.g. the interpretation of a child's weight that is on the 50th percentile for its age will be different if it has been moving along the 90th percentile at previous ages than if it has consistently been on the 50th percentile; and (*iv*) evaluation of centile charts used as a screen for problems.

The concepts are introduced using an aspect of a study conducted at Tygerberg Hospital where centile charts for maternal weight gain in pregnancy were developed and assessed for their usefulness in detecting light-for-gestational-age (LiGA) births. The reference centile charts for maternal weight show poor discriminating ability between LiGA and normal births. These results support arguments in favour of abandoning the routine weighing of pregnant women.

S Afr Med J 1997; **87**: 206-213.

Centile charts are commonly used in many areas of health research and practice, e.g. growth charts for children, Doppler ultrasonography in pregnancy and assessment of cholesterol levels at different ages.[1-3] Usually, there is a particular measure of interest, e.g. height in children's growth charts, and associated with this a temporal variable such as age. For each age, say, any particular centile represents the value below which that given percentage of the reference population falls.

While such centile charts are in common use, there are a number of aspects of both their construction and application that are problematic, and it is some of these issues that will be raised in this paper. A new approach to centile charts, which addresses some of the problems raised, will also be presented by way of illustration.

Weight gain in pregnancy

The concepts presented in this paper will be illustrated by means of an application to maternal weight gain in pregnancy. The issue of whether weight gain in pregnancy is a useful screen for problem pregnancies is the subject of ongoing debate. The 1950 edition of *Williams Obstetrics*[4] suggested that a weight gain of no more than 20 pounds during pregnancy was desirable. Since 1971, the updated version of the same publication[5] observes that 'rigid calorific restriction may be dangerous to the fetus'. In a study of British general practitioners,[6] it was established that over 95% of those surveyed *do* routinely weigh their pregnant patients, yet there was little consensus as to *why* they did so. Dommisse,[7] commenting on inappropriate antenatal care, recommends that 'repeated weighing of patients is of little value and could probably be discontinued'.

A study was undertaken at Tygerberg Hospital in the Western Cape to assess formally the usefulness of weight gain in identifying light for gestational age (LiGA) births during pregnancy in a local population. This forms part of a broader study to assess the usefulness of antenatal monitoring of weight and symphysis-fundal height in screening for a number of pregnancy outcomes, and is reported in more detail elsewhere.[8] Some of the results will be presented here to illustrate the concepts as they are developed. These results also serve to add support to the view that routine measurement of weight during pregnancy should be abandoned. The primary aim of this paper, however, is to focus attention on issues surrounding the use of centile charts generally, and the weight gain example is used here primarily as a vehicle towards that end.

Choice of reference population and reference sample

In developing a new centile chart, a sample is generally chosen from a reference population and the values of the variable of interest within the sample (e.g. maternal weight) are used to estimate the centiles at each time point (e.g. gestational age). So the very first question to ask when centile charts are derived should be: what is the appropriate reference population?

In part, the answer to this question depends on the use to which the centile charts will be put. They may be intended simply to describe the characteristics of the particular population or for comparison with other populations. More frequently, however, they will be used to define 'normality' in some sense, so that individuals who fall above or below certain (usually arbitrarily) specified centiles will be regarded as 'abnormal'. In this context, the reference centiles are

Department of Statistical Sciences, University of Cape Town, and Centre for Epidemiological Research in Southern Africa, Medical Research Council, Tygerberg, W Cape

Mary Lou Thompson, BSc Hons, PhD

Department of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg

L Paul Fatti, BSc Hons, MSc, PhD

being used as standards. For instance, babies that fall below the 10th percentile of birth weight for their gestational age are classified as LiGA.

In either case, the question of exclusion criteria arises. Should the sample be randomly selected from a frame that includes *all* individuals in the reference population or should there be exclusions? In the case of centiles for birth weight by gestational age, for instance, should one exclude babies born with congenital abnormalities? If one is looking at weight gain in pregnancy, should one exclude women with diabetes mellitus? Each of these choices has an impact on the interpretation of the resulting centile values and consideration needs to be given to these issues before the reference population and hence the sample are specified. If, as is typically the case, the centile estimation is based on a sample that in some sense reflects the 'optimal', then, by definition, 10% of the 'optimal' population will fall below the 10th percentile, whereas if centile estimation is based on an unselected sample, then the centiles reflect the population overall and to define those below the 10th percentile as 'not optimal' may well have more meaning.

If the reference centiles *are* to be used as a standard this will also have implications for the appropriate choice of reference population. Are international reference centiles appropriate for global use? Even in the developed world there are differences in growth paths of individuals that cannot be attributed solely to nutritional status.[9,10] It is certainly useful to have international references for purposes of general comparison, but for monitoring and setting targets locally it may be of value also to have local references.[11]

A further issue that needs consideration is whether each individual in the sample should contribute measurements at more than one time point. A sample based on many repeated observations on a few individuals may create the illusion of a large sample, but, in fact, the serial measurements on each individual will be highly correlated, reducing the effective sample size. Serial measurements are, of course, essential if one is studying *growth* rather than *size*.

The main point being made here is that care needs to be exercised when choosing the sample on which the centile estimates are based and that this choice has implications for the validity of the resulting charts and their implementation in different contexts.

In the study of weight gain in pregnancy, information was collected from Tygerberg Hospital records on the weight at each antenatal clinic visit of 700 women whose pregnancies satisfied the following criteria: (*i*) the subject gave birth after 37 weeks' gestation; (*ii*) the birth weight of the infant was $\geq$ 2 500 g; (*iii*) the baby was not LiGA; and (*iv*) there were no other complications, such as diabetes or pre-eclampsia.

What reference population does a centile chart based on this sample represent? It will reflect the weight gain patterns of that segment of the population served by Tygerberg Hospital that had 'normal' pregnancies in the sense defined above. So one would expect, then, for instance, 10% of these 'normal' pregnancies to fall below the 10th percentile at a particular gestational age. These exclusion criteria are clearly debatable and valid arguments can be raised to the effect that there should have been more limited (or no) exclusions in selecting this reference sample.

A further sample of 100 records was drawn of women who had given birth to LiGA babies. These records were not included in the centile estimation but were used subsequently to assess the usefulness of the weight gain centile charts to screen for LiGA births.

Estimation of centile charts

Once the reference population has been specified and the sample selected, the technical problem is the estimation of the centiles of interest (e.g. 10th, 50th, 90th) from the sample at each time point (e.g. gestational age). This is an area which has received considerable attention in the statistical literature in recent years, but non-statistician users of these charts also need to be aware that a wide variety of methods have been developed for estimating centile charts, ranging from non-parametric estimation (in which no particular distributional shape — such as the Normal — is assumed) at each time point followed by smoothing across time,[12-14] to various parametric approaches where smoothly changing probability distributions are fitted at each point in time and the centile estimates are derived from the probability distribution itself.[15-17]

An advantage of the parametric approach is that all values in the sample are incorporated in the centile estimation, whereas in non-parametric centile estimation only a neighbouring few observations contribute towards each centile estimate. Consequently, parametric estimates are generally more stable. The parametric approach, however, requires assumptions about the distributions of the variable under consideration at each time point, and the validity of the results is dependent on the appropriateness of the chosen model. No matter which approach is used, it must be kept in mind that the chart is an *estimate* of the centiles of the reference population and the precision with which these centiles have been estimated will also depend on the sample size at each point in time.

Most charts for weight gain in pregnancy cite the mean weight gain per week, without bounds around the mean to define a reference range.[18] The first local chart for weight gain in pregnancy was also based on data from Tygerberg Hospital[19] and the current study is an extension of those results. In both studies a parametric approach was used to estimate the centiles at each gestational age; the actual methodology is described in detail elsewhere.[8,16] The resultant centile chart (10th, 50th, 90th percentiles), for a woman who booked in the 18th week of pregnancy at which point she weighed 56.7 kg, is shown in Fig. 1. (The chart can be adapted for any booking weight and week.)

Centile charts as a screen

As discussed above, centile charts may be used simply for descriptive purposes, to compare one population with another or to characterise a particular population. Frequently, however, charts are intended for use as a screen for individuals who fall above or below certain (generally arbitrary) levels and are therefore regarded as being at higher risk. Babies born below the 10th percentile of birth weight for their gestational age are LiGA; similarly, children are classified as having low weight-for-age and so on.

Essentially the reference centile chart is being used here as a screening tool to flag potential problems. As with any other screening method, it should then be assessed in terms of its

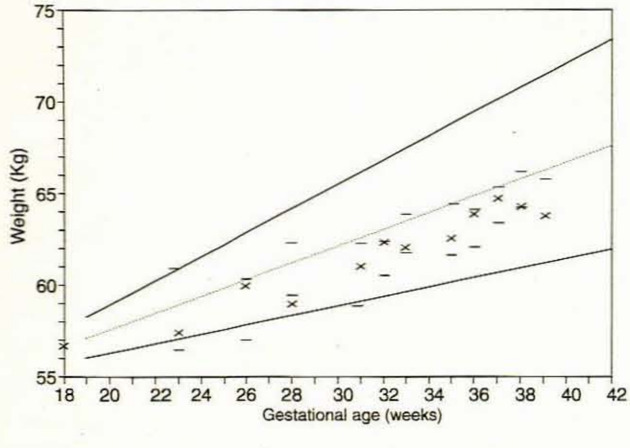

Fig. 1. Centile charts for maternal weight gain.

sensitivity, specificity, and positive and negative predictive values (PPV and NPV) (Table I).[20] To do this requires one first to specify the outcome that one is attempting to detect. This is often surprisingly difficult. In the study of British general practitioners[6] mentioned earlier, although 95% of respondents routinely weighed their pregnant patients, there was little consensus as to why they did so. Then one needs to establish whether the screening method is any good at detecting the outcome of interest. This means going beyond simply establishing that an association exists between the outcome and the variable measured on the chart. There can be a significant association without its being of useful predictive value. There is, for instance, a well-established association between maternal weight gained in pregnancy and birth weight,[21-22] but the question of whether monitoring weight gain during pregnancy is a useful screen for problem pregnancies remains. There may well be a lot of diligent measuring going on which actually has little value in terms of improving outcome.

Table I. Definitions of terms

Term	Definition
Sensitivity	Proportion of those truly 'diseased' who screen positive
Specificity	Proportion of those truly 'disease'-free who screen negative
Positive predictive value	Proportion of those screened positive, who truly have the 'disease'
Negative predictive value	Proportion of those screened negative, who are truly 'disease'-free
Accuracy	Proportion of those screened, who are correctly identified
Centile coverage	Proportion of the reference population that lies within the centiles

In addition, the usefulness of screening is often assessed at just one (often arbitrary) centile cut-off, e.g. the definition of LiGA babies as those below the 10th percentile. Each percentile cut-off will have a different associated sensitivity and specificity for identifying the problem outcome of interest and these can be displayed collectively by means of a receiver operating characteristic (ROC)[23] curve, on the basis of which one can make an informed (rather than arbitrary) choice as to the most useful (if any) percentile cut-off for each particular purpose.

To clarify these ideas, consider the weight gain in pregnancy example and assume that the problem outcome that is being screened for is a LiGA birth. As one uses a higher centile cut-off to flag potentially problematic pregnancies, the sensitivity of the screen will increase (one will flag a higher percentage of truly LiGA births) but the specificity will decrease (one will also flag a higher percentage of non-LiGA births). If the 10th percentile of the centile chart for weight gain is used as a screen for LiGA in the current study, it is found to have sensitivity and specificity of 16% and 79% respectively. This is comparable to the results of Dawes and Grudzinskas[24] for a population of pregnant women at Oxford. One could, instead, flag all women whose weight path falls below, say, the 25th percentile at some point. This would result in a screen that has sensitivity and specificity of 44% and 60%. The values for different cut-offs are presented in the (dashed) ROC curve shown in Fig. 2. The closer the ROC curve is to the upper left-hand corner, the better. An idealised curve is shown on the same graph — this has a sensitivity of 90% associated with 65% specificity and sensitivity of 80% associated with 80% specificity. The diagonal from bottom left to top right represents the 'no knowledge' situation and values below this represent 'worse than guessing'. One sees that the ROC curve associated with use of maternal weight as a screen for LiGA births is very close to the diagonal and, at times, below it, indicating that one would do better to leave the decision to chance than to use this form of weight monitoring as a screening method!

In addition to sensitivity and specificity, which are independent of prevalence, the positive and negative predictive values (PPV and NPV) of any screening procedure (which do depend on prevalence) need to be evaluated at each potential cut-off. In the weight gain example, if the prevalence of LiGA is 10%, the PPV and NPV at the 10th centile cut-off are 8% and 89% respectively and at the 25th centile cut-off, 11% and 91%. If the prevalence of LIGA were 30% (say in a hypertensive population), the PPV and

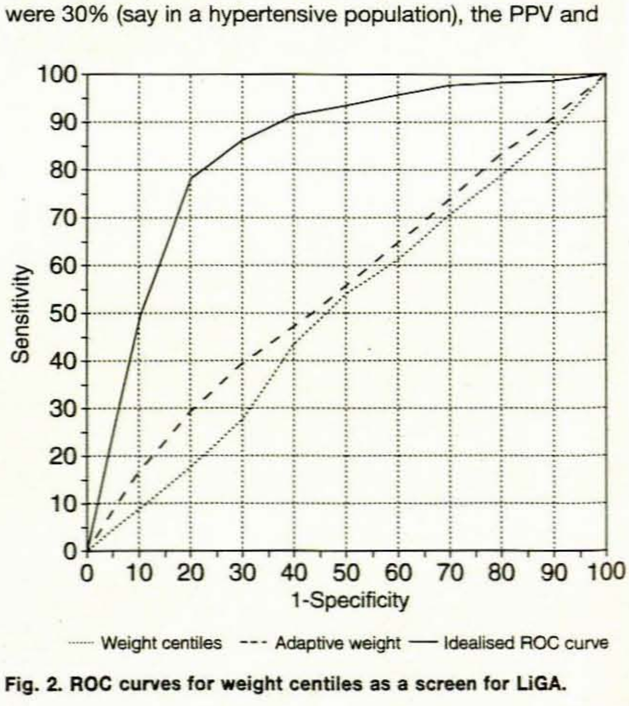

Fig. 2. ROC curves for weight centiles as a screen for LiGA.

NPV at the 10th centile cut-off would be 25% and 69% respectively, and at the 25th centile cut-off 32% and 71%.

Longitudinal versus cross-sectional charts

Say one is considering a chart with 10%, 50% and 90% centiles. If this chart is used just once at a particular time point (as birth weight by gestational age), then it is clear that, e.g. 10% of babies from the reference population will fall below the 10th percentile. If, however, one monitors an individual longitudinally, it is no longer clear what percentage of the reference ('ideal') population would fall outside the bounds at some point in their follow-up path. We saw above, for instance, that when the 10% cut-off was used for maternal weight, 21% of the 'normal' sample had records below the 10th centile at some point during their pregnancy. (This corresponds to 79% specificity.)

Information on the prior path

As stated by Ruel and Habicht,[25] 'Growth charts are almost universally used as if only the latest measurement was relevant.' The prognosis of an individual who has been moving along the 90th percentile and then falls to the 50th is likely to be different from that of someone who is on the 10th percentile and continues to remain there. Centile charts, even if used longitudinally, are generally interpreted cross-sectionally and the prior path of the individual is hence ignored, or at best incorporated informally. This is the issue of size, e.g. height, versus growth. Conventional centile charts provide information on size at a particular age but not information on growth; yet growth is most commonly the real focus of interest.

The NCHS growth charts for children,[1] for instance, reflect the 'normal' range of sizes at each particular age, but do not reflect the 'normal' growth (i.e. change in size) from one age to another, even though children are followed up across ages on these charts. The charts developed by Cameron[26] reflect growth rather than size and were derived from a prospective study of British children. With serial monitoring on a chart of size (say weight for age) one is able to observe, for example, that a child's weight is close to the 90th percentile at each time of measurement, but not how typical or atypical its *growth* is from one time point to the next relative to that of other children who are also on the 90th percentile. What is the 'normal' range of weight at 2 years for children who were on the 90th percentile at 1 year? This is a question that can only be answered by a true *growth* chart rather than one which merely reflects *size* at each time point.

By incorporating information on the prior path in the form of conditional centiles[8,27] it is possible to incorporate both size and growth centiles in the same chart. *Conditional* centiles are adaptive, adjusting their level according to the past measurements on each individual. The horizontal bars superimposed on Fig. 1 are the 10th and 90th percentiles of weight gain, conditional on the prior path of the woman whose record is also shown. This is the actual weight gain path of a woman who gave birth to a LiGA baby. The woman first visits the clinic in the 18th week of pregnancy at which point her weight is 56.7 kg, she next returns to the clinic 5 weeks later in week 23 of her pregnancy when her weight has only increased to 57.5 kg, close to the 10th percentile of the unconditional chart for weight gain. The conditional percentiles adjust for the fact that this woman is on a low weight gain path, whereas conventional charts look at the weight at a particular gestational age, ignoring the prior path. The woman next visits the clinic in week 26, when she weighs 60 kg. By this point, the conditional centiles reflect the growth (from 57.5 g) that would be expected, given her prior path, and are correspondingly adjusted downward relative to the unconditional chart. The woman's record at no point lies outside the conventional unconditional chart, but her weight gain path is below the 10th percentile of the conditional chart in the 28th week of pregnancy.

Conditional charts used for screening should also be assessed in terms of their predictive value. The actual specificity for a given centile cut-off will be lower than in the unconditional charts. Here the sensitivity and specificity following a 10% cut-off are 34% and 77%, respectively and, following a 25% cut-off, 58% and 48%. These correspond to PPV and NPV for the 10th centile cut-off of 14% and 91% (10% prevalence of LiGA) and 39% and 64% (30% prevalence of LiGA), and for the 25th centile cut-off of 11% and 91% (10% prevalence of LiGA) and 32% and 73% (30% prevalence of LiGA). These results are shown in Table II for a hypothetical population of 1 000 pregnant women. The resulting ROC curve is also shown in Fig. 2 and can be seen to be a small improvement on that of unconditional weight gain.

Table II. Predictive value of weight gain centiles

(a) Prevalence = .1

		LiGA	AGA				LiGA	AGA	
Screen	+	34	207	241	Screen	+	58	468	526
	−	66	693	759		−	42	432	474
		100	900	1 000			100	900	1 000

Sensitivity = 34%
Specificity = 77%
Accuracy = 73%

Sensitivity = 58%
Specificity = 48%
Accuracy = 49%

(b) Prevalence = .3

		LiGA	AGA				LiGA	AGA	
Screen	+	102	161	263	Screen	+	174	364	538
	−	198	539	737		−	126	336	462
		300	700	1 000			300	700	1 000

Sensitivity = 34%
Specificity = 77%
Accuracy = 64%

Sensitivity = 58%
Specificity = 48%
Accuracy = 51%

Discussion

A number of issues relating to the estimation and use of centile charts have been raised here which will hopefully lead to both improved estimation of such charts in the future and to more informed use. In particular, the distinction between *size* and *growth* needs to be kept in mind when serial monitoring is intended. In such instances conditional charts offer a means of monitoring both aspects simultaneously.

In addition, the importance of formally assessing any chart that is being used as a screen has been emphasised. This requires one first explicitly to specify the problem outcome that is to be screened for. ROC curves enable one to assess the screening characteristics of a centile chart over the full nge of cut-offs. Beyond this, one needs to consider the ccuracy of a particular cut-off (the percentage of correct agnoses) and the costs involved (not necessarily purely nancial) of the false positives and false negatives. The onditional weight screen at the 10% cut-off, for instance, ith a 10% prevalence of LiGA births, would have 72.7% ccuracy, i.e. 72.7% of pregnancies followed up in this way ould be correctly diagnosed. Of the misdiagnoses (27.3%), 6% would be LiGA births (66% of LiGA births would be issed) and 20.7% normal births (misdiagnosed as LiGA) able II).

Despite the improvement in screening quality obtained by corporation of information on the prior path, neither chart ppears to offer a screening tool for LiGA births that arrants the diligent measuring of weight that is going on in usy clinics. There may well be other problem outcomes ssociated with pregnancy for which weight gain does ovide a useful screen (perhaps as an early indicator of ypertension), but at least in the context of screening for GA births no strong evidence emerges for the continuation these routine measurements and this analysis hopefully ovides a telling example of the value of formal assessment the usefulness of routine measurement.

EFERENCES

1. National Center for Health Statistics (NCHS). *NCHS Growth Curves For Children: Birth-18 Years, United States* (Vital and Health Statistics, Series II, publication No. (PHS) 78-1650). Washington DC: US Department of Health, Education and Welfare, 1977.
2. Pattinson RC, Theron GB, Thompson ML, Lai Tung M. Doppler ultrasonography of the fetoplacental circulation — normal reference values. *S Afr Med J* 1989; **76:** 623-625.
3. Rossouw JE, Steyn K, Berger GMB, *et al*. Action limits for serum total cholesterol. *S Afr Med J* 1988; **73:** 693-700.
4. Eastman NJ. *Williams Obstetrics*. 10th ed. New York: Appleton-Century-Crofts, 1950: 314.
5. Hellman LM, Pritchard JA. *Williams Obstetrics*. 14th ed. New York: Appleton-Century-Crofts, 1971: 336.
6. Dawes MG, Green J, Ashurst H. Routine weighing in pregnancy. *BMJ* 1992; **304:** 487-489.
7. Dommisse J. Inappropriate antenatal care. *S Afr Med J* 1993; **83:** 460.
8. Thompson ML, Fatti LP. Construction of multivariate centile charts for longitudinal measurements. *Stat Med* 1997 (in press).
9. Roede MJ, Van Wieringen JC. Growth diagrams 1980. Netherlands third nation-wide survey. *Tijdschrift voor Sociale Gezondheidszorg* 1985; **63:** 1-34.
10. Tanner JM, Davies PSW. Clinical longitudinal standards for height and height velocity for North American children. *J Pediatr* 1985; **107**: 317-329.
11. Dunn PM. A perinatal growth chart for international reference. *Acta Paediatr Scand Suppl* 1985; **319:** 180-187.
12. Hamill PVV, Drizd TA, Johnson CL, Reed RB, Roche AF, Moore WM. Physical growth: National Center for Health Statistics percentiles. *Am J Clin Nutr* 1979; **32:** 607-629.
13. Ryan AS, Martinez GA. Physical growth of infants 7 to 13 months of age: results from a national survey. *Am J Phys Anthropol* 1987; **73:** 449-457.
14. Healy MJR, Rasbash J, Yang M. Distribution-free estimation of age-related centiles. *Ann Hum Biol* 1988; **15:** 17-22.
15. Cole TJ. Fitting smoothed centile curves to reference data. *J R Stat Soc Series A* 1988; **151:** 385-418.
16. Thompson ML, Theron GB. Maximum likelihood estimation of reference centiles. *Stat Med* 1990; **9:** 539-548.
17. Altman DG. Construction of age-related centiles using absolute residuals. *Stat Med* 1993; **12:** 917-924.
18. Rosso P. A new chart to monitor weight gain during pregnancy. *Am J Clin Nutr* 1985; **41:** 644.
19. Theron GB, Thompson ML. A centile chart for weight gain in pregnancy for the urban population of the Western Cape, South Africa. *Int J Gynecol Obstet* 1990; **33:** 127-133.
20. Mausner JS, Kramer S. *Epidemiology — An Introductory Text*. Philadelphia: WB Saunders, 1985: 214-223.
21. Abrams BF, Laros RK. Prepregnancy weight, weight gain and birth weight. *Am J Obstet Gynecol* 1986; **154:** 503-509.
22. Seddman DS, Ever-Hadani P, Gale R. The effect of maternal weight gain in pregnancy on birth weight. *Obstet Gynecol* 1989; **74:** 240-246.
23. Thompson ML, Zucchini W. On the statistical analysis of ROC curves. *Stat Med* 1989; **8:** 1277-1290.
24. Dawes MG, Grudzinskas JG. Repeated measurement of maternal weight during pregnancy. Is this a useful practice? *Br J Obstet Gynaecol* 1991; **98:** 189-194.
25. Ruel MT, Habicht JP. Response to 'The correct use of growth charts' (Letter). *Am J Public Health* 1992; **82:** 899.
26. Cameron N. Conditional standards for growth in height of British children from 5.0 to 15.99 years of age. *Ann Hum Biol* 1980; **7:** 331-337.
27. Cole TJ. Statistical constructs of human growth: new charts for old. In: Ulijaszek SJ, Mascie-Taylor CGN eds. *Anthropometry: the Individual and the Population*. Cambridge: Cambridge University Press, 1994: 78-98.

Accepted 10 Aug 1996.