

Approaches for Constructing Age-Related Reference Intervals and Centile Charts for Fetal Size

M. Hynek^{1,2}

¹Gennet, Center for Fetal Medicine, Prague, Czech Republic

²Department of Gynecology and Obstetrics, Thomayer University Hospital, Prague, Czech Republic

Supervisor: Prof. RNDr. Jana Zvárová, DrSc.

Summary

The assessment of the fetal size based on ultrasound biometry is important for proper pregnancy management. The means for evaluations of these measurements represent age-related reference intervals and centile charts which are used to detect the extreme values, possibly indicating pathology. The key features concerning the design of studies used for the construction of these intervals are given. The aim of this paper is to review possible statistical approaches, pointing out the methodology, goodness of fit, advantages and limitations.

Keywords: Fetus, gestational age, ultrasound, biometry, pathology, reference interval, centile, centile chart, normal distribution, normalizing transformation, quantile regression

1. Introduction

The assessment of fetal size is of a crucial importance for the proper pregnancy management. It is implicitly based on ultrasound measurements of embryonic and fetal biometrical parameters (e.g. crown-rump length, head circumference, femur length, etc.), most of which almost invariably increase with the gestational age (GA) [1]. The means for evaluation of these measurements are age-related reference intervals (RIs) and centile charts, allowing interpretation of obtained fetal body measurement in comparison with the body size of fetuses in the reference population [2].

The RI (often misleadingly called 'normal range') represents the interval between a pair of predetermined extreme centiles (commonly the 5th and 95th for a 90% interval) of a size variable, denoted y , at a given GA (denoted t). Centile charts plot the value of y corresponding to one or more

centiles against relevant GA over a range of GAs.

In the field of fetal size, measured values of y which lie outside the RI are regarded as extreme and may indicate the presence of underlying pathology (e.g. morphological or chromosomal abnormality [3], intrauterine growth restriction [4], macrosomia [5]). Otherwise, more informatively, the value's exact centile position relative to the reference population is estimated from the knowledge of the distribution of y at a given GA. The proximity of the centile position to 0% or 100% is a measure of how extreme the observation is compared to the reference population [6].

Variety of strategies for constructing of RIs and centile charts has been published. The choice of appropriate methodology in the field of fetal biometry is especially crucial as inaccurate centiles may lead to incorrect conclusions regarding fetal development. The aim of this article is to scrutinize possible statistical methods for age-related RIs and centile charts for fetal size.

2. Study design

Prior to statistical methods we consider the key features of the study design.

Size and growth. It is important to bear in mind the difference between two notions, which are commonly confused: size and growth. The fetal size represents a single measurement point, while fetal growth refers to a dynamic process [7], [8]. These two concepts contain completely diverse information. The fetal size can be normal despite abnormal growth, whereas normal fetal growth can be associated with an abnormal size. It is necessary to retain the clear distinction between the size and

growth and to realize that the appropriate study design for these outcomes is different.

Cross-sectional and longitudinal studies.

The same as above applies for the difference between cross-sectional and longitudinal studies [1], [4]. In the former, each fetus contributes just once to a reference sample, whereas in the latter each fetus is measured several times. Therefore, cross-sectional data give information only on the size and can be used to compare the size of a fetus (of known GA) on a single occasion with reference data.

Longitudinal data may be used to produce RIs for the fetal size as well as growth. The analysis of longitudinal data requires a different approach. Serial measurements on an individual fetus are highly correlated so that the effective sample size in such an approach is likely to be nearer to the number of fetuses than to the total number of observations [1]. Thus, statistical methodology must be accordingly adjusted. Further information can be found in the literature [9], [10]. The following review will deal with the constructing RIs for the fetal size based on the cross-sectional data, as they are more common in the field of fetal biometry, maybe because of easier data collection compared to longitudinal studies.

Sample selection and data collection.

The choice of an appropriate sample is of a great importance. It is preferable to collect data specifically for the purpose of developing RIs, with each fetus being included only once [1]. The aim is to obtain as unselected and representative group as possible with no prior selection for examination as data collected at a scan done for any clinical indication may seriously bias the results.

Typically, the status of pregnancy is determined retrospectively, and there is scope for controversy in deciding which pregnancy should be regarded as abnormal and excluded from the reference sample. Altman and Chitty [1] recommend using for the exclusion only information available at the time of the ultrasound measurement, with the exceptions of subsequently found serious congenital anomaly or condition affecting fetal growth, such as maternal diabetes or renal disease. There is desirable to have approximately equal numbers of measurements at each week of gestation [1].

Sample size. It is hard to specify the appropriate sample size for developing centile charts. The larger the sample size the greater precision centiles will have. The interest is usually focused on the tails of the distribution. Royston [11] published that the standard error (SE) of the 100 α th centile ($c_{100\alpha}$) can be approximately expressed as a multiple of the age-specific standard deviation (SD) as

$$SE(c_{100\alpha}) \approx SD \sqrt{\frac{1 + \frac{1}{2} z_{1-\alpha/2}^2}{n}} \quad (1a),$$

where $z_{1-\alpha/2}$ is the appropriate value from the standard normal distribution and n is the number of cases [11]. Therefore for the sample size n we obtain

$$n \approx \frac{1 + \frac{1}{2} z_{1-\alpha/2}^2}{\left(\frac{SE(c_{100\alpha})}{SD}\right)^2} \quad (1b).$$

If, for example, we require SE for a 90% reference interval to be 5% of SD, the resulting value of n would be $n \approx (1 + 0,5 \cdot 1,6452)/0,052 \approx 941$, for SE corresponding to 10% of SD $n \approx 235$. Besides, more extreme centiles are less precisely estimated. It is clear that basically several hundred observations are necessary to get reasonable estimates of extreme centiles.

3. Statistical methods

The number of possible statistical methods for the calculation of RIs and centile charts has mushroomed in the last two decades.

The basic requirements for the method are to produce centiles which change smoothly with GA and provide good fit to the data [1]. The requirement stated in the nineties, to maintain as simple as possible [1], is becoming less important due to the broad availability of a computer technology. Finally, it is desirable that model allows the calculation of the relevant centile position and Z-score for any further measurement [12]. Z-scores have been used increasingly in recent years, and are the WHO-recommended system for comparing individual anthropometric measurement with the reference population [13]. They have also proved to be a powerful quality-control tool, allowing the sonographers to choose appropriate reference charts [14] and to audit their daily practice [15].

Generally, methods are divided into two broad categories: parametric (based on modelling the distribution) and non-parametric (empirical).

3.1 Mean and SD model

In this most common parametric method, the basic assumption is that at each GA the measurement of interest has a normal distribution with a mean and SD that vary smoothly with GA. A desired centile curve is then calculated using formula

$$c_{100\alpha} = \mu + k \cdot SD \quad (2),$$

where k is the corresponding centile of the normal distribution (e.g. for 10th and 90th centile k is ± 1.28 , for 5th and 95th centile k is ± 1.645 , etc.) and μ and SD are, respectively, the mean and standard deviation at the required GA for the reference population. The approach is based upon least-squares regression analysis to model both the mean and the SD curves as polynomial functions of GA [1], [4], [11], [16].

First, the regression analysis is applied to the raw data to find a suitable curve for the mean (Figure 1). As for the choosing the degree of polynomial, Royston [11] and Royston and Wright [4] recommend the initial use of a cubic polynomial ($a + b.t + c.t^2 + d.t^3$). If the coefficient d is not significantly different from zero (d divided

by its SE and the result referred to the appropriate point of the Student t distribution with $n - 4$ degrees of freedom), a quadratic polynomial is then fitted and the same assessment applied for c coefficient. The process is repeated until no further removal of terms is possible. Altman and Chitty [7] suggest the linear-cubic model ($a + b.t + c.t^2$) as a good alternative for fetal size data. Fitting polynomial of a higher degree than cubic is not advocated as the fitted curve may exhibit unrealistic features such as waviness or sharp deviation at extreme ages [4]. In case coefficients for quartic or quintic polynomials are statistically significant, the use of fractional polynomials is recommended (see section 3.1.2) [17]. The choice of curve should be based not only on statistical significance, but also the aesthetical appearance and the quality of fit to the data should be taken into account.

Next step is to check the goodness of fit of the obtained polynomial by plotting the residuals (observed minus predicted values) against GA to assess if and how variability changes with age [7]. The normal quantile-quantile (Q-Q) plot of residuals can reveal any departure from normality (e.g. positive or negative skewness). In this case transformation of the data towards normality should be tried prior to the analysis. See section 3.1.1.

After the suitable mean model is established, the attention is turned to the modelling of the variability. Many previously published studies fit the model which does not allow the SD to change with gestation. However, in the field of fetal measurements the between-subject variability almost always shows an upward trend during pregnancy [7], so it is crucially important to consider the relation between the SD and age. The approach of Altman [16] is one of the most frequently used. Based on the assumption that variable of interest has a normal distribution at all ages, the residuals from the mean model should be then also normally distributed. Subsequently, the absolute values of residuals should have a half normal distribution.

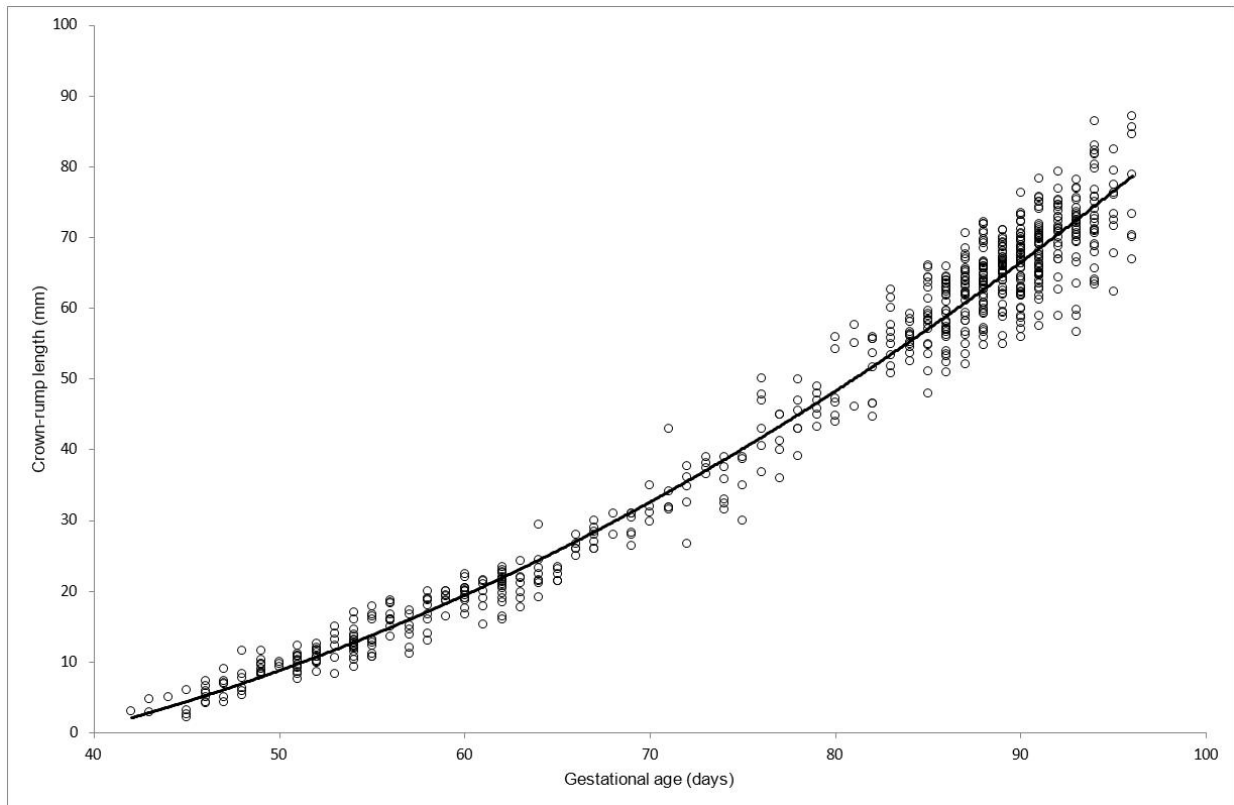


Fig. 1. Plot illustrating the relationship between fetal crown-rump length and gestational age during first trimester of pregnancy, showing the raw data and the fitted curve for the mean (—) with the quadratic polynomial $y = 0.0126 t^2 - 0.3177 t - 6.7365$. The sample consists of 664 fetuses with exactly known gestational age (in-vitro fertilization pregnancies).

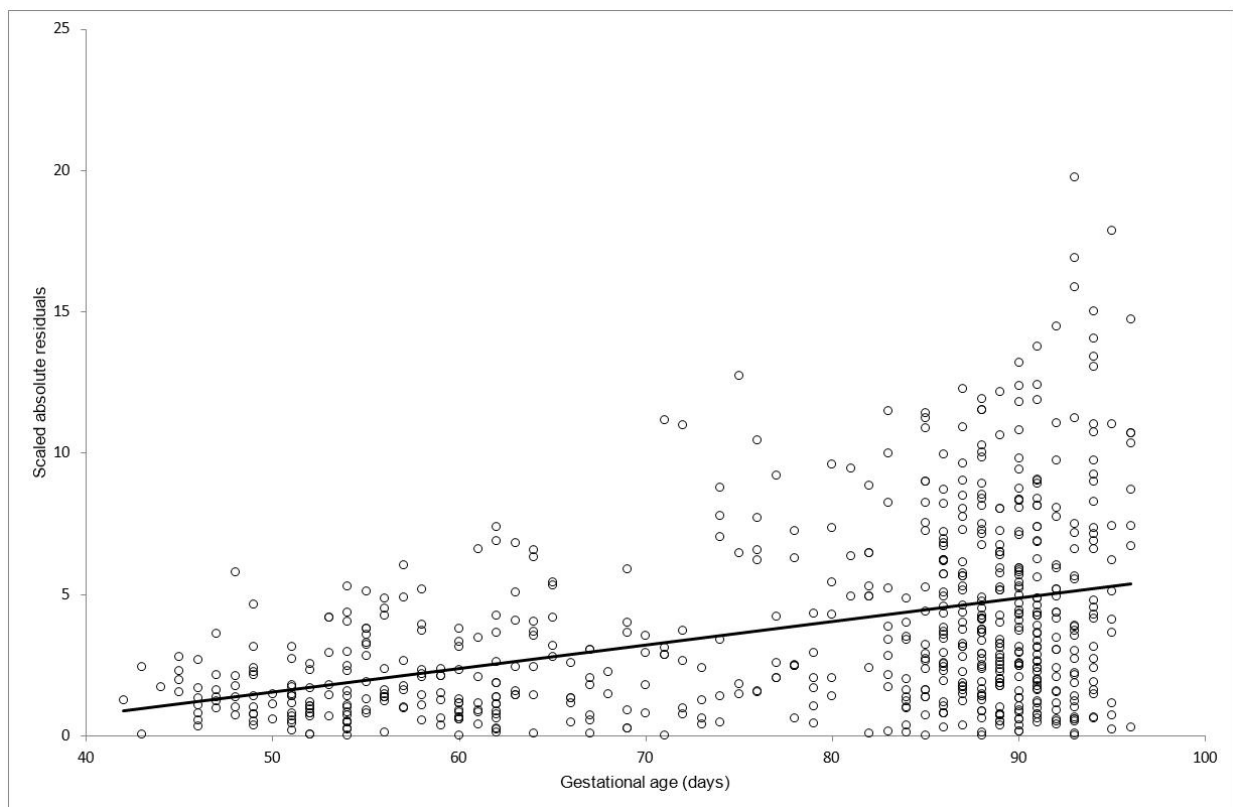


Fig. 2. Plot of scaled absolute residuals from the regression line shown in Figure 1 together with fitted linear regression line $y = 0.0833 t - 2.6173$.

If we have a variable X which follows a normal distribution with mean zero and variance 2, the absolute value $|X|$ follows a half-normal distribution which has mean $\sqrt{2/\pi}\sigma$ [18]. The mean of the absolute residuals multiplied by $\sqrt{(\pi/2)}$ is an estimate of the SD of the residuals. Thus if the SD is not reasonably constant over age, the predicted values from regression of the absolute residuals on age multiplied by $\sqrt{(\pi/2)}$ will give age-specific estimates of the SD of the residuals, and hence of y .

An identical approach published by Royston and Wright [4] comprises to produce 'scaled absolute residuals' (SARs) by multiplying the absolute residuals by $\sqrt{(\pi/2)}$. Subsequently, SARs are regressed on age in the same way as for the mean and again the predicted values from this curve estimate the SD of the residuals (Figure 2). It is unlikely that a curve more complex than quadratic is needed for a satisfactory fit for the SD [16]. Superimposing 1.645 SD on the residual plot is helpful in assessing how well the SD

has been modelled, as presumably 90% of the observed residuals should fall within these limits.

If absolute residuals show no trend with age, the SD is estimated as the SD of unscaled original residuals.

The regression analysis to estimate the mean ought strictly to take into account any change in SD with gestation. Thus the mean model could be subsequently refitted using the reciprocal of the square of the estimated age-specific SD as weights [19]. However, the effect of this refitting is almost always rather small [7].

The next very important step includes the evaluation of goodness of fit. A useful tool includes Z-scores which we calculate for the observed values y using formula

$$Z \text{ skór} = \frac{y - \mu_{GS}}{SD_{GS}} \quad (3),$$

where GA and SDGA are, respectively, the mean and SD given by the model for the GA at which the observation y is made. Then several modalities are advocated [7]: a plot of Z-scores against age to confirm that no untoward pattern is present and to check that the expected proportions of values fall between or outside the appropriate centiles (for example 90% of Z-scores should lie between $Z = 1.645$, 5% above and 5% below) (Figure 3). Further normal Q-Q plot of Z-scores can be used to check if they have close to normal distribution (Figure 4). This is signified by roughly straight line and can be formally confirmed using Shapiro-Wilk W [20] or Shapiro-Francia W' test [21]. However, in large samples slight deviation from normality may lead to statistically significant non-normality, so it is unwise to worry about moderately small p (such as $p = 0.01$) unless the normal plot shows clear deviation from a straight line [7].

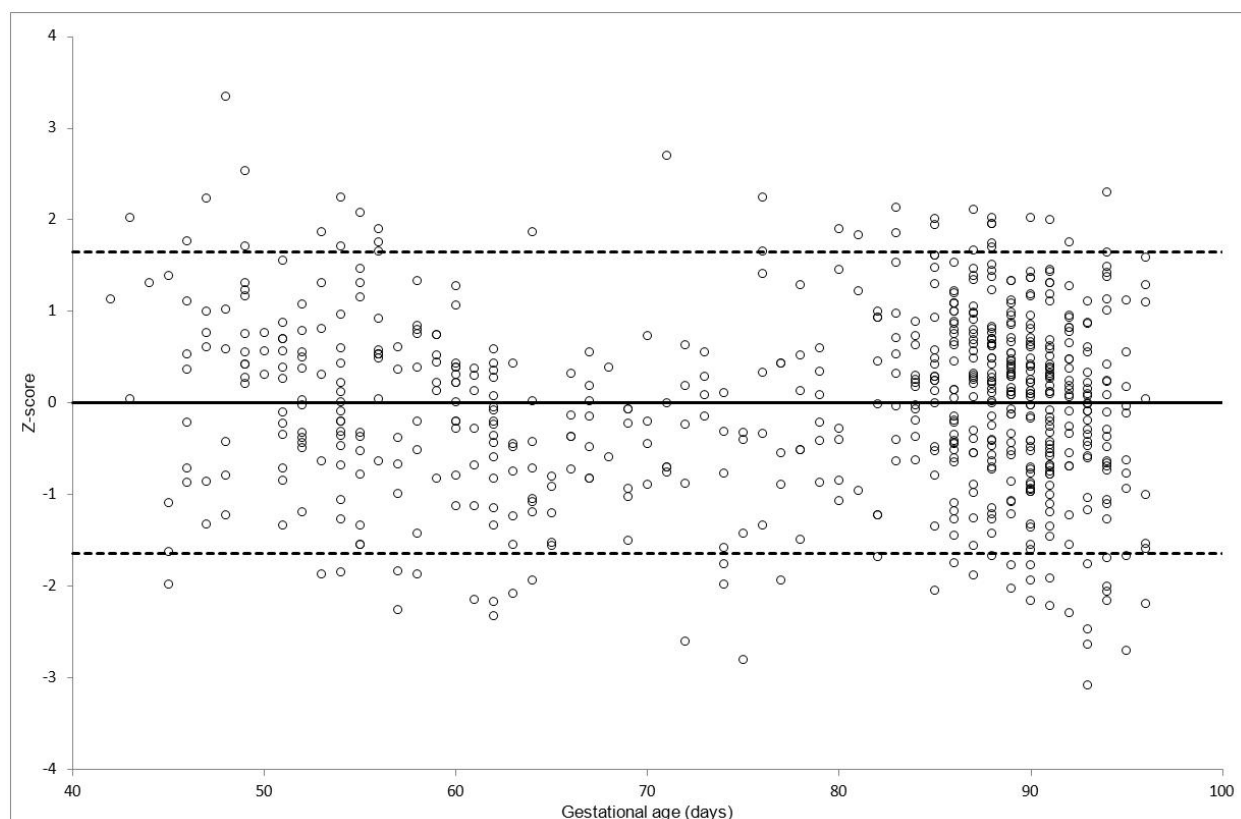


Fig. 3. Plot of Z-scores of observations with the expected 5th and 95th centiles 6.0% of observations fall below 5th centile, 5.3% above 95th centile and 88.7% between 5th and 95th centiles.

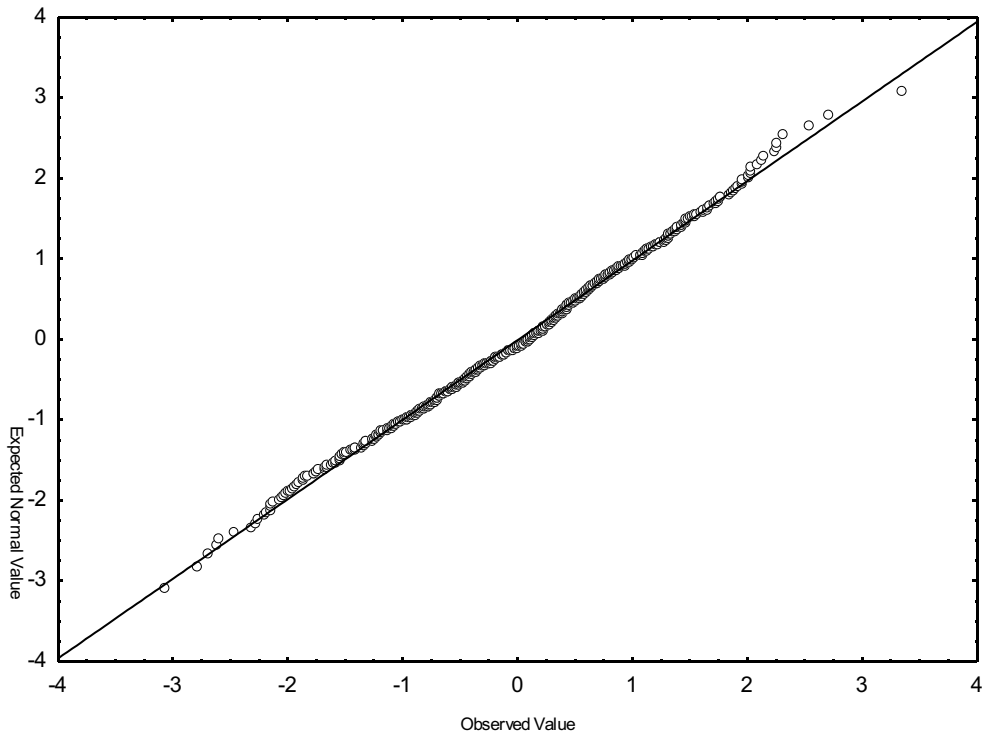


Fig. 4. Normal quantile-quantile plot of Z-scores of observations. Shapiro-Wilk W test = 0.997 ($p = 0.31$).

Eventually, a satisfactory model being determined, the centile curves for the desired RIs may be calculated by substituting the expression for the fitted

mean and SD into equation (2) and superimposed on a scatter diagram of observations as a final check to fit (Figure 5). The Z-score for any new measurement

can be calculated using equation (3) and its centile obtained from the inverse normal distribution function.

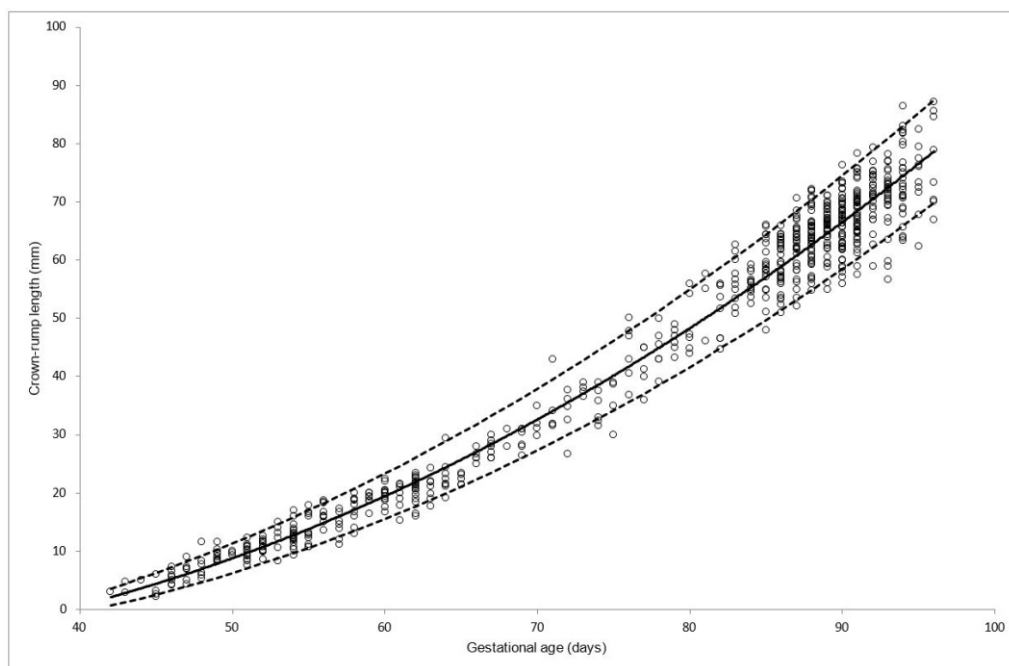


Fig. 5. Final model of fetal crown-rump length in relation to gestational age, showing raw data together with reference curves for the mean (—), 5th and 95th centiles (-----).

3.1.1 Logarithmic transformation

The method just mentioned requires the assumption that at each GA the data come from a population with normal distribution. However, this is not always the case, as many fetal measurements often present skewness in the distribution (mostly positive, with the right tail of the distribution longer than the left). The logarithmic transformation can help to overcome the issue. The solution at the same time stabilizes the variance in case the SD increases rapidly with age [11].

Royston [11] suggests if residuals from the initial model show a positive skew to perform a logarithmic transformation on the original value y and refit the model on $\log(y)$. If residuals from refitted model are once again skewed, then it is recommended to try a shifted logarithmic transformation of the form $\log(y + C)$, with $C > 0$ for residuals negatively skewed and $C < 0$ for positively skewed. The best value of C is that which maximizes p -value for the normality test of the residuals.

Once the model is finalized it is important to back-transform the curves using antilog (and to subtract C , respectively).

3.1.2 Fractional polynomials

Conventional polynomials suffer from several well-known disadvantages. Low order polynomials offer only a few curve shapes and thus do not always fit the data well, whereas high order ones may fit badly at the extremes of the observed range. Further, they do not have asymptotes and cannot fit data where limiting behaviour is expected [22]. Royston and Altman [17] proposed an extended family of curves, called fractional polynomials (FPs), whose power terms are restricted to a small predefined set of integer and non-integer values. Conventional polynomials are a subset of this family. If conventional polynomial is of a form

$$a + b.t + c.t^2 + d.t^3 + \dots \tag{4}$$

FPs are defined as

$$a + b.t^{p_1} + c.t^{p_2} + d.t^{p_3} + \dots \tag{5}$$

where p_1, p_2, \dots are chosen from the set $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$. Conventionally, the power 0 represents natural logarithmic transformation, so that t^0 equals to $\log_e(t)$. Another extension involves 'repeated powers'. In this case the second term is multiplied by $\log_e(t)$. For example, an FP of degree 3 with powers (0, 2, 2) is then of a form

$$a + b.\log_e(t) + c.t^2 + d.t^2.\log_e(t) \tag{6}$$

An FP of first degree is of the form $a + b.t^p$. For a given data set, the best value of p is found by fitting eight separate linear regressions using t^2, t^1, \dots, t^p and selecting the value of p which gives the best fit. Regarding FPs of the second degree ($a + b.t^{p_1} + c.t^{p_2}$), using the standard set detailed above would involve fitting a model for each of 36 permissible combinations of powers. The model with the lowest residual SD is chosen as the best.

The use of FPs can often give a better fit even with fewer terms compared to conventional polynomials.

3.2 LMS method

The LMS method, introduced by Cole [23], [24] and further refined by Cole and Green [25], is an extremely flexible and widely applicable semi-parametric method which can produce smooth centile curves even when the data appear to have a complex shape. Furthermore, time-varying skewness, which cannot be taken into account with classical logarithmic transformation, is easily dealt with. The method assumes that the use of suitable power transformation can remove skewness and normalize the data. Such a family of transformations is that proposed by Box and Cox [26], with the optimal power λ at a given age calculated from the data to completely remove the skewness.

The distribution of the variable of interest y changes smoothly with age and is completely summarized by three parameters λ (Box-Cox power, related to skewness), μ (median) and σ (coefficient of variation), the initials of which (L, M and S) give the name to the method.

Suppose that y has median μ and that y^λ (or $\log_e(y)$ if $\lambda = 0$) is normally distributed. Then we consider transformed variable x

$$x = \begin{cases} \left(\frac{y}{\mu}\right)^\lambda - 1 & \text{pro } \lambda \neq 0 \\ \log\left(\frac{y}{\mu}\right) & \text{pro } \lambda = 0 \end{cases} \tag{7}$$

based on the Box-Cox transformation [26]. This transformation maps the median of y to $x = 0$ and is continuous at $\lambda = 0$. For $\lambda = 1$ the SD of x is exactly the coefficient of variation (CV) of y , and this remains approximately true for all close to one [25]. The optimal value of λ is that minimizes the SD of x .

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

$$Z = \frac{x}{\sigma} = \begin{cases} \frac{\left(\frac{y}{\mu}\right)^\lambda - 1}{\lambda\sigma} & \text{pro } \lambda \neq 0 \\ \frac{\log\left(\frac{y}{\mu}\right)}{\sigma} & \text{pro } \lambda = 0 \end{cases} \tag{8}$$

and is assumed to take a standard normal distribution. Assume that the distribution of y varies with t (GA), and that λ, μ and σ at t are read off the smooth curve $L(t), M(t)$ and $S(t)$. Then

$$Z = \begin{cases} \frac{\left(\frac{y}{M(t)}\right)^{L(t)} - 1}{L(t)S(t)} & \text{pro } L(t) \neq 0 \\ \frac{\log\left(\frac{y}{M(t)}\right)}{S(t)} & \text{pro } L(t) = 0 \end{cases} \tag{9}$$

Rearranging (9) shows that centile $c_{100\alpha}$ of y at t is given by

$$c_{100\alpha}(t) = \begin{cases} M(t) \left[1 + L(t) S(t) z_\alpha \right]^{1/L(t)} \\ M(t) e^{S(t) z_\alpha} \end{cases}$$

$$\begin{aligned} \text{pro } L(t) &\neq 0 \\ \text{pro } L(t) &= 0 \end{aligned} \quad (10),$$

where z_α is the normal equivalent deviate of size α . This shows that if L , M and S curves are smooth, then so are the centile curves. The equation (10) allows us to calculate and draw any desired centile curve and the equation (9) allows converting any individual measurement into Z-score.

The original method [23], [24] involved splitting the data into age groups, as narrow as possible, but at the same time with adequate numbers of measurements (at least 100 for the best result), then estimating L , M and S for each group using maximum likelihood and finally constructing a smooth curve by the means of cubic splines, kernel methods, polynomials or simply drawing by eye. The choice of age cut-offs between groups is arbitrary and thus could influence the final result.

Cole and Green [25] added a non-parametric aspect to the original LMS method by using penalized likelihood function. The curves $L(t)$, $M(t)$ and $S(t)$ are estimated by maximizing the penalized likelihood

$$\begin{aligned} \ell - \frac{1}{2} \alpha_\lambda \int \{L''(t)\}^2 dt - \\ - \frac{1}{2} \alpha_\mu \int \{M''(t)\}^2 dt - \\ - \frac{1}{2} \alpha_\sigma \int \{S''(t)\}^2 dt \end{aligned} \quad (11),$$

where α_λ , α_μ and α_σ are smoothing parameters and log-likelihood function ℓ is given by

$$\begin{aligned} \ell = \ell(L, M, S) = \\ = \sum_{i=1}^n \left(L(t_i) \log \frac{y}{M(t_i)} - \log S(t_i) - \frac{1}{2} z_i^2 \right) \end{aligned} \quad (12).$$

Three integrals provide roughness penalties for the curves $L(t)$, $M(t)$ and $S(t)$, leading to natural cubic splines with knots at each distinct value of t . The advantage of this approach is that the subjective grouping step is removed, the entire dataset is treated as a single entity and the curve fitting across the age is controlled directly by smoothing parameters. No age cut-offs need to be specified, and L , M and S values at each age are used in turn to calculate the other two parameters. The only arbitrariness is the choice of α_λ , α_μ and α_σ or in practice equivalent degrees of freedom (EDF), calculated for each fitted curve as a function of these smoothing parameters, which give a more usable measure of the extent of the smoothing. The EDF of each L , M and S curves is a measure of complexity (e.g. EDF = 1 means a constant, EDF = 2 corresponds to the straight line, EDF = 3 almost a quadratic curve, and EDF ≥ 4 indicates progressively more complex curve shapes). The choice must strike balance between the fidelity to the data and the smoothness. Low values of EDF may lead to the oversmoothed curves whereas high values to the undersmoothed ones. The disadvantage is that the choosing of EDFs is somewhat subjective. However, as a very rough guideline, we may use a procedure for assessing goodness of fit by comparing the difference in deviance ($-2\log(\text{penalized likelihood})$) between two models where the total number of EDFs differed by e to a χ_e^2 -distribution [6]. The other procedure, which helps in finding proper smoothing values for EDFs, represents a worm plot, a very sensitive diagnostic tool proposed by Buuren and Frederiks [27]. The worm plot consists of a collection of detrended Q-Q plots of residuals (detrended means that each empirical quantile is subtracted from its corresponding unit normal quantile

[28]), which are split according to age. A model that fits data well is characterized by a flat worm-like string. A particular change in shape and location of worm corresponds to the type of misfit (e.g. in mean, variance, skewness, kurtosis).

3.3 LMSP method

The assumption of normality following the Box-Cox transformation may be violated by the presence of kurtosis, for which the transformation does not adjust.

The LMSP method of Rigby and Stasinopoulos [29] can be understood as a generalization of the LMS approach. It uses a Box-Cox power exponential distribution to try to overcome the issue of kurtosis. The distribution has four parameters denoted as μ , σ , ν and τ and , which may be interpreted as relating to location (median), scale (approximately CV), skewness (transformation to symmetry) and kurtosis- (power exponential parameter), respectively. The advantage is that this distribution provides a flexible model for both skewness and kurtosis (allowing either for platykurtosis or leptokurtosis).

The model is fitted by maximizing a penalized likelihood. Centile estimation proceeds in a manner not dissimilar to that of the LMS method.

3.4 HRY method

Healy, Rabash and Young [30] (hence HRY) proposed a non-parametrical procedure based on the technique of Cleveland [31] for a smoothing a scatter plot by using locally weighted regression. This approach makes no assumption about the nature of the distribution of measurements at a given age and centiles are estimated empirically. At the same time it is expected that both centiles themselves and the intervals between centile at a fixed age should behave smoothly. This is done by constrain that the spacings between centiles can be expressed as a low-order polynomial in the underlying Z-score.

The first stage involves obtaining the selected raw centiles. The data are sorted into ascending age order. The first k measurements are then selected, where k is a fraction of the total number available, typically 5-10%. A regression of these k measurements on age is fitted, the desired centiles are obtained from the ranked residuals, using interpolation when necessary and their values are plotted against the median age value of k points. This procedure has used points 1 to k . It is repeated successively using points 2 to $k + 1$, 3 to $k + 3$, ... until the whole span of ages has been covered.

The initial centile curves will be very irregular, so the second stage is needed to smooth them. A smooth curve for the i th centile can be estimated by the polynomial of degree p

$$y_i = a_{0i} + a_{1i}t + a_{2i}t^2 + \dots + a_{pi}t^p \quad (13).$$

The coefficients a_{ji} for a fixed j are then modelled as a polynomial in z_i , where z_i is the normal equivalent deviate of the i th centile, so that

$$a_{ji} = b_{j0} + b_{j1}z_i + b_{j2}z_i^2 + \dots + b_{jq_j}z_i^{q_j} \quad (14),$$

where the degree q_j of the polynomial may differ from one value of j to another. This restricts the distance between centiles and prevents the resulting curve to cross. Combining equations (13) and (14) gives a linear model which simultaneously fits all the centiles estimated in stage one, and this can be fitted by least squares.

The method allows a considerable amount of flexibility. However, in practice a good deal of experimentation may be needed to choose good values for the adjustable parameters k , p and q_j . Authors recommend to judge the suitability of a particular set by counting the points falling between adjacent centiles and comparing these counts with expected values. In theory, for any observation a corresponding Z-score can be calculated by solving a polynomial equation, though the order of the polynomial may realistically prohibit this [32].

As certain variables observed over a wide age range may need more complex curve shapes than a single polynomial could provide. Pan et al. [33] suggested dividing the data into contiguous age groups, fitting polynomials within each group and smoothing points where these meet by using an extra polynomial term. As an alternative, Goldstein and Pan [34] proposed that these age groups be defined initially and that the fitted response be made smooth across the join points by equating derivatives of the curves on either side.

3.5 Non-parametric quantile regression

Other approaches for constructing the reference curve include a range of methods based on non-parametric quantile regression [35]. These methods allow quantiles to be estimated as a smooth function of covariates without imposing parametric distributional assumption, thus they are valuable in case that any transformation method is not able to achieve normality over the full range of relevant ages. The approach is robust to the presence of outliers and moreover, the procedure is unsupervised since all the smoothing parameters are determined adaptively.

Gannoun et al. [36] proposed three methods using kernel estimation, local constant kernel estimation and double kernel estimation of conditional quantile curves. Regarding the choice of kernels, authors use the univariate standard normal and uniform densities and also give some guideline for the selection of the bandwidths. Another example of using quantile regression presented Wei et al. [37]. For more details we refer to the original papers [36], [37].

Regarding the goodness of fit, an elegant way presents the worm plot, as shown by Buuren [38].

4. Discussion

For constructing age-related quantiles there are several different methods available, each of them having both advantages and limitations. Hence, it is unlikely that a single one would be appropriate in all circumstances.

The parametric 'mean and SD model' benefits from being relatively simple and easy to use, with the necessary techniques available in most basic statistical packages. It must be emphasized that the validity of reference ranges depends critically on the assumption of normality. The approach is able to cope with some heteroscedasticity by modelling the SD as age-varying and skewed data may sometimes be corrected by logarithmic or shifted logarithmic transformation. However, time-varying skewness and non-normal kurtosis cannot be easily accommodated. The resulting centile curves and Z-scores can be expressed as explicit formulae. The fact that the method suffers from the well-known limitations of a polynomial curve shape can be greatly improved by using the family of fractional polynomials.

In the field of fetal measurements this is the approach which has been repeatedly used so far as a wide range of fetal biometric measurements available from ultrasound scanning fulfil sufficiently the assumption of normality. Nevertheless, it is not always tenable and alternative techniques are required.

The LMS method with penalized likelihood is extremely flexible and widely applicable producing convincing centile curves irrespective of the complexity of the curve shape. It can easily cope with time-varying skewness, though some non-normal kurtosis may remain. Appealing by-product of the method is that $L(t)$, $M(t)$ and $S(t)$ curves may be of interest in their own as they facilitate further investigation in the underlying structure of the dataset. Although succinct formulae for centile curves are unobtainable, Z-score for any individual measurement can be easily calculated. There are, however, some drawbacks with the smoothing approach. These are the presence of 'edge effects' (spurious changes), if the data are sparse near the ends, or non-uniform smoothing, if there are unequal numbers of observations throughout the age range [25].

The LMS method has been increasingly used in recent years and it was the chosen procedure for creating the 2000 CDC Growth Charts for the United States [2]. The introduction of specially designed programs (LMSChartmaker by Cole and Pan [39]) and packages for general statistical programs (package lmsqreg by Carey [40] for R) made the method rather accessible.

A proposed extension to the LMS approach, the LMSP method is even more flexible as it takes into account the presence of kurtosis in the distribution.

The main advantage of the HRY method is that no assumption about the nature of distribution is made. The approach is flexible and capable of handling many patterns of growth, with the suggestions of Pan et al. [33] and Goldstein and Pan [34] making it even more so. Nevertheless, the choice of degrees of polynomials requires considerable experience and trial, and it is not always clear how to improve the fit. Although the formulae are available, the estimation of the Z-score and centile value for further observations is not simple, unless a very basic model has been fitted. The method is vulnerable to outliers and requires specially written software, as well. Non-parametric approaches based on quantile regression are very robust, extremely flexible, provide a much better fit to the data than other methods [38] and moreover, are fully unsupervised. Although more detailed insight into the issue of quantile regressions is not so easy and straightforward compared to previous approaches, the possibilities of applications are very wide. The method has potential of identifying the features in the data undetected by other methods and is rapidly entering mainstream statistics [38]. Another compelling motivation for the quantile regression approach is the ability to extend the conventional unconditional models depending only on the age to models that incorporate prior growth and other covariates. Further, the number of implementations in statistical software is growing (e.g. package quantreg by Koenker [41] for R).

However, some drawbacks exist. The approach lacks an explicit formula to convert measurement into quantile and

Z-score. The produced curves may be irregular near the extremes, and are generally less aesthetically pleasing [38].

5. Conclusion

Cole remarked [23] that fitting smooth centile curves has always been something of a black art. We have presented several different approaches how to deal with the constructing centile charts of the fetal size.

A closer understanding how the particular method works can help us to reveal its potential and advantages and lead us to choose the most appropriate one for our goal. On the other hand, knowledge of its limitations may prevent us from misinterpretation of the results and consequently from false clinical conclusions and suboptimal care.

The explicit formula that allows one to convert a measurement into quantile or Z-score was one of the requirements set forth by a WHO expert committee [13]. This would exclude the empirical methods such as the HRY method and quantile regression approach. Hence, the choice is basically left to the trade-off between the simplicity and usability of more limited parametric approaches, and the more flexible and applicable but less user-friendly models provided by the LMS, respectively LMSP methods. The potential of quantile regression models in the field of fetal medicine remains to be further explored.

Acknowledgements

The work was supported by the grant SVV-2010-265513.

References

- [1] Altman D.G., Chitty L.S.: Design and analysis of studies to derive charts of fetal size. *Ultrasound Obstet Gynecol* 1993; 3: 378-384.
- [2] Kuczmarski R.J., Ogden C.L., Guo S.S.: 2000 CDC growth charts for the United States: Methods and development. Hyattsville: National Center for Health Statistics; 2002.
- [3] Snijders R.J., Sherrod C., Gosden C.M., Nicolaidis KH. Fetal growth retardation: associated malformations and chromosomal abnormalities. *Am J Obstet Gynecol* 1993; 168: 547-555.
- [4] Royston P., Wright E.M.: How to construct 'normal ranges' for fetal variables.

- [5] Sokol R.J., Chik L., Dombrowski M.P., Zador I.E.: Correctly indentifying the macrosomic fetus: improving ultrasonography-based prediction. *Am J Obstet Gynecol* 2000; 182: 1489-1495.
- [6] Wright E.M., Royston P.: A comparison of statistical methods for age-related reference intervals. *J R Statist Soc A* 1997; 160: 47-69.
- [7] Altman D.G., Chitty L.S.: Charts of fetal size: 1. Methodology. *Br J Obstet Gynecol* 1994; 101: 29-34.
- [8] Salomon L.J.: Early fetal growth: concepts and pitfalls. *Ultrasound Obstet Gynecol* 2010; 35: 385-389.
- [9] Royston P., Altman D.G.: Design and analysis of longitudinal studies of fetal size. *Ultrasound Obstet Gynecol* 1995; 6: 307-312.
- [10] Royston P.: Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal studies. *Stat Med* 1995; 14: 1417-1436.
- [11] Royston P.: Constructing time-specific reference ranges. *Stat Med* 1991; 10: 675-690.
- [12] Waterlow J.C., Buzina R., Keller W., Lane J.M., Nichaman M.Z., Tanner J.M.: The presentation and use of height and weight data for comparing nutritional status of groups of children under the age of 10 years. *WHO Bulletin* 1977; 55: 489-498.
- [13] WHO. Physical status: the use and interpretation of anthropometry. Geneva: World Health Organization; 1995.
- [14] Salomon L.J., Bernard J.P., Duyme M., Buvat I., Ville Y.: The impact on choice of reference charts and equations on the assessment of fetal biometry. *Ultrasound Obstet Gynecol* 2005; 25: 559-565.
- [15] Salomon L.J., Bernard J.P., Ville Y.: Analysis of Z-score distribution for the quality control of fetal ultrasound measurements at 20-24 weeks. *Ultrasound Obstet Gynecol* 2005; 26: 750-754.
- [16] Altman D.G.: Construction of age-related reference centiles using absolute residuals. *Stat Med* 1993; 12: 917-924.
- [17] Royston P., Altman D.G.: Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Statist* 1994; 43:429-467.
- [18] Bland J.M.: The half-normal distribution method for measurement error: two case studies. Talk presented in the Statistics and Econometrics seminar series, University of York, February 2005. <http://www-users.york.ac.uk/~mb55/talks/halfnor.pdf> [Accessed 9 October 2010]
- [19] Aitkin M.A.: Modelling variance heterogeneity in normal regression using GLIM. *Appl Statist* 1987; 36: 332-339.
- [20] Shapiro S.S., Wilk M.B.: An analysis of variance test for Normality (complete samples). *Biometrika* 1965; 52: 591-611.

- [21] Shapiro S.S., Francia R.S.: An approximate analysis of variance test for normality. *J Am Statist Assoc* 1972; 67: 215-216.
- [22] McCullagh P., Nelder J.A.: *Generalized linear models* (2nd ed). London: Chapman&Hall/CRC; 1989.
- [23] Cole T.J.: Fitting smoothed centile curves to reference data. *J R Statist Soc A* 1988; 151: 385-418.
- [24] Cole T.J.: The LMS method for constructing normalized growth standards. *Eur J Clin Nutr* 1990; 44: 45-60.
- [25] Cole T.J., Green P.J.: Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; 11: 1305-1319.
- [26] Box G.E.P., Cox D.R.: An analysis of transformation. *J R Statist Soc B* 1964; 26: 211-252.
- [27] Buuren S., Fredriks M.: Worm plot: a simple diagnostic device for modeling growth reference curves. *Stat Med* 2001; 20: 1259-1277.
- [28] Wilk M.B., Gnanadesikan R.: Probability plotting methods for the analysis of data. *Biometrika* 1968; 55: 1-17.
- [29] Rigby R.A., Stasinopoulos D.M.: Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. *Stat Med* 2004; 23: 3053-3076.
- [30] Healy M.J.R., Rabash J., Yang M.: Distribution-free estimation of age-related centiles. *Ann Hum Biol* 1988; 15: 17-22.
- [31] Cleveland W.S.: Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979; 74: 829-836.
- [32] Silverwood R.J., Cole T.J.: Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size. *Ultrasound Obstet Gynecol* 2007; 29: 6-13.
- [33] Pan H.Q., Goldstein H., Yang Q.: Nonparametric estimation of age-related centiles over wide age ranges. *Ann Hum Biol* 1990; 17: 475-481.
- [34] Goldstein H., Pan H.: Percentile smoothing using piecewise polynomials, with covariates. *Biometrics* 1992; 48: 1057-1068.
- [35] Koenker R.: *Quantile regression*. Cambridge: Cambridge University Press; 2005.
- [36] Gannoun A., Girard S., Guinot C., Saracco J.: Reference curves based on non-parametric quantile regression. *Stat Med* 2002; 21: 3119-3135.
- [37] Wei Y., Pere A., Koenker R., He X.: Quantile regression methods for reference growth charts. *Stat Med* 2006; 25: 1369/1382.
- [38] Buuren S.: Worm plot to diagnose fit in quantile regression. *Stat Med* 2007; 7: 363-376.
- [39] Cole T.J., Pan H.: LMSChartmaker. <http://homepage.mac.com/tjcole> [Accessed 13 September 2010].
- [40] Carey V.: LMSqreg: An R package for Cole-Green reference centile curves. <http://www.biostat.harvard.edu/~carey> [Accessed 13 September 2010]
- [41] Koenker R.: Quantreg: An R package for quantile regression and related methods. <http://cran.r-project.org> [Accessed 13 September 2010]

Contact**MUDr. Martin Hynek**

Gennet

Center for Fetal Medicine

Kostelní 9

170 00 Prague 7

Czech Republic

e-mail: martin.hynek@gmail.cz