ANNALS OF HUMAN BIOLOGY, 2002, VOL. 29, NO. 2, 202–218

Short-term and long-term variability of standard deviation scores for size in children

A. SHEEHY†, T. GASSER†, R. LARGO‡ and L. MOLINARI‡

† Abteilung Biostatistik, ISPM, University of Zurich, Zurich, Switzerland
‡ Abteilung Wachstum und Entwicklung, Universitätskinderspital Zürich, Zurich, Switzerland

Received 1 November 2000; in revised form 29 March 2001; accepted 26 April 2001

Summary.
Primary objective: To quantify long-term and short-term variability in the standard deviation scores (SDS's) for six skeletal size variables and body mass index (BMI) in children and to compare average values of these quantities for boys with those of girls and to make comparisons across variables.
Methods and procedures: The analysis is based on measurements made regularly for 120 boys and 112 girls from 1 month until 20 years for seven variables (standing height, sitting height, leg height, arm length, biiliac width, bihumeral width and BMI) as part of the first Zurich longitudinal growth study. Variation in these scores due to variablity in the timing of the pubertal spurt (PS) is separated out by rescaling the age axis on an individual basis and comparing children with the same developmental age rather than the same chronological age. For a given child, the relationship between the value of its SDS and age is modelled as the sum of an arbitrary (child dependent) smooth function plus an error term. The long-term variability for that child is defined to be the mean square of the departures of this smooth function from its mean level while the short-term variability is defined to be the variance of the error term.
Main outcomes and results: Girls' SDS scores have significantly more long-term variability than those of boys, while there is no significant difference between the sexes for short-term variability. Bihumeral width, BMI and sitting height have significantly more long-term variation than the other variables. Bihumeral width and BMI have the largest short-term variability and standing height has the smallest. Correlations between long-term variability and adult size and timing and intensity of the PS were small.
Conclusions: A useful way of assessing long-term and short-term variability of SDS's, which is widely applicable has been described and applied to data relating to the growth of children. The results of this analysis are intriguing. Why is the underlying growth process of girls more variable than that of boys? Differences across skeletal parameters are also interesting and deserve further consideration.

1. Introduction

That children do not stay in the same size percentile from birth to adulthood is common knowledge. Moreover, the fact that for a given child an early (late) pubertal spurt (PS) leads to a sharp increase (decrease) in size percentiles during the period of that child's PS is also obvious. In order to deepen our understanding of how children grow, it is of interest to be able to quantify the variability in size percentiles for a given child, having separated out the component due to deviations in the timing of the PS from the average. Because it is easier to model the relationship between standard deviation scores (SDS's) and age, we will in fact consider variability of SDS's rather than that of size percentiles: since sizes at a given age are approximately normally distributed,writing Φ, for the cumulative distribution function of the standard normal distribution, we have Φ^{-1}(standard deviation score) = size percentile, and large variability in the percentiles implies large variability in the SDS's and vice versa.

Annals of Human Biology ISSN 0301–4460 print/ISSN 1464–5033 online
http://www.tandf.co.uk/journals
DOI: 10.1080/0301446011007569 3

In this paper we examine the variability of SDS's from 1 month until 20 years for seven variables (standing height, sitting height, leg height, arm length, biiliac width, bihumeral width and bodymass index (BMI)). The analysis is based on measurements made regularly for 120 boys and 112 girls in the first Zurich longitudinal growth study. We are particularly interested in examining the intra-individual variability of these scores over time and in comparing variability across variables and between sexes. For the purpose of computing the SDS's, by rescaling the age axis on an individual basis, we group together children who have the same developmental age rather than the same chronological age. (The details are discussed in the methods section) We assume that the pattern of change for the SDS's can be modelled by a smooth curve, about which there are short-term fluctuations due to seasonal differences in growth rates, illness or other stresses, measurement error and other factors. We quantify the variation of the curve about its mean value over all ages by means of a 'long-term variability index (LTV)' and we quantify the short-term variations by means of a 'short-term variability index (STV)'.

Long-term variability is closely related to the notion of 'tracking' (Foulkes and Davis 1981, Goldstein 1981, McMahan 1981, Ware and Wu 1981). There are many definitions of tracking. Loosely put, one can say that the phenomenon of tracking is observed when the percentile of an individual's measurements with respect to a changing population distribution over time, stays constant. When data are normally distributed this corresponds to constancy of SDS's over age. Equivalently, in the terminology we will use here, perfect tracking occurs when the long-term variability is zero; conversely, the greater the long-term variability, the less a child's growth can be said to track. Here, we use a new statistical method for quantifying long-term variability. Most of the existing methods for quantifying tracking behaviour assume a specific model for the data. Ours in contrast is fully nonparametric, which seems more appropriate for growth data.

We can summarize the values of the LTVs and the STVs for boys and for girls and thus make comparisons between sexes and in a similar way we can make comparisons across variables. It has been noted (Sheehy, Gasser, Molinari *et al.* 2000b) that the adult size of girls is more difficult to predict from childhood growth parameters than that of boys. Molinari, Gasser, Largo *et al.* (1995) showed that the correlation of size at prepubertal ages with adult size is smaller for girls than for boys. We would expect then that girls track less well than boys; on average they should have greater values of the LTV than boys. In Sheehy *et al.* (2000b) it was also shown that adult sitting height and adult bihumeral width are more difficult to predict from various parameters of growth than the four size variables standing height, leg height, arm length and biiliac width. We would therefore expect that sitting height and bihumeral width would track less well than the other four variables: on average they should have greater values of the LTV than for the other variables. The question of whether the size of the LTV (equivalently, the tracking behaviour) is related to adult size, intensity of the pubertal spurt (PS) or timing of the PS is also of interest. One could, for example, conjecture that a short growth period (and thus an early onset to the PS) leads to relatively poor tracking and thus a high value for the LTV.

We also considered the following question: when we consider the variance of the SDS's of the six size variables (a quantity we will refer to as the 'body proportions variability (BPV)') calculated for each child at different ages, is there a relationship between this quantity and age? One might expect, for example, that for a typical child, the values of the SDS's for each of the six size variables would converge to a common value with increasing age so that the BPV would be zero for such a child. More generally, we wondered if these variances (on average) were decreasing with increasing age. We will refer to this as the 'decreasing BPV hypothesis'.

2. Subjects and methods

2.1. Subjects

Within an internationally coordinated design (Falkner 1960), a study was initiated at the Kinderspital Zürich, in 1954. Participation of a representative sample of 413 Swiss children was sought. This was successful for 160 girls and 161 boys, as judged by participation in the first year. Between 1955 and 1976 a total of 23 girls and 24 boys were lost for various reasons. For a more detailed description of this study see Prader, Largo, Molinari and Issler (1988). Children who missed more than two visits or who missed two successive visits were excluded from the present analysis. The same holds for children who developed a disease known to affect growth. These criteria led to a final sample of 112 girls and 120 boys.

2.2. Measurements

The children were seen at 1, 3, 6, 9, 12, 18 and 24 months and then annually. Starting at age 9 for the girls and age 10 for the boys, measurements were taken half-yearly, until the annual increment in height was less than 0.5 cm per year. Yearly measurements were continued until the increment in height had become less than 0.5 cm in 2 years. However, no child was discharged before age 18. Time limits were ideally ± 2 days at 1 month, ± 1 week for 3–18 months and ± 2 weeks afterwards. However, the exact days of measurement were available and used in the analysis which allowed us to include measurements outside the above intervals.

Measurement procedures for the collection of variables considered here were as follows: Standing and sitting height were taken with a Harpenden stadiometer and leg height was taken to be the difference of these measurements. Bihumeral and biiliac width were measured to the nearest millimetre with callipers. Arm length was measured with a tape and the result was rounded to whole centimetres. Unfortunately, bihumeral width and not biacromial width was measured; therefore, when skeletal width is of interest, the measurements are contaminated by growth of muscle and fat tissue.

For a subset of 127 children, bihumeral width was not measured on two or more visits between the ages of birth and two years (in fact for 112 of them, it was never measured in this time). This was due to an administration error and these measurements can be considered to be missing at random. For this reason indices for bihumeral width will be calculated only on the remaining subset of 47 girls and 58 boys. The BMI of a child is defined to be its weight (in kilograms) divided by the square of its height (in metres squared). BMI values are well known to have a skewed distribution and for this reason we work here with the transformed values $-1/\mathrm{BMI}$ (see Gasser, Ziegler, Prader *et al.* 1994)

2.3. Methods

The computation of SDS's around puberty is problematic due to the variation in the timing of the PS from child to child: clearly a girl of aged 12 who has already achieved the maximum growth velocity associated with PS should not be compared with a girl of the same age whose growth velocity has not yet begun the steep increase associated with puberty. For this reason we have used 'longitudinal' ages: the age scale for each child is smoothly transformed by means of a (child specific) function so that in essence all children of the same sex have their PS at the same time. In this way, a child's size at a given age is compared with the size of other children at the same developmental age rather than the same chronological age.

What form should such a transformation take? It seems clear that, certainly until age 3, comparing children of the same chronological age for the purpose of computing SDS's is reasonable. Furthermore, subjects at age 20 may also be safely compared with each other. Using velocity and acceleration curves based on kernel smoothers (Gasser, Müller, Köhler *et al.* 1984), the age, T_{end}, at which the maximum absolute deceleration in growth after the PS occurred, was estimated. The age of a given child at his point of maximum absolute deceleration (T_{end}) in growth was matched with the average of these ages over all children of the same sex ($\overline{T}_{\text{end}}$). Ages between 3 and T_{end} were matched with ages in the interval 3 to $\overline{T}_{\text{end}}$ using a linear function, and ages between T_{end} and 20 were matched with ages in the interval $\overline{T}_{\text{end}}$ and 20. Notice that this transformation is variable dependent: the age at which T_{end} occurs is not the same for all variables. For the variable BMI we used the value of T_{end} for standing height. In figure 1 an example is given of such a transformation for standing height for a girl for whom $T_{\text{end}} = 12$, 1.5 years earlier than the value of $\overline{T}_{\text{end}}$ for standing height for girls.

Using the method of kernel smoothing (Gasser *et al.* 1984) we can estimate the size of a child at any age between 0 and 20 years. If $D_i(t)$ denotes the estimated size of the ith girl at age t for leg height, for example, and we write f_i for the corresponding transformation mapping chronological age into longitudinal age, then

$$\text{SDS (girl } i \text{ at 'longitudinal' age 8)} = \frac{D(f_i^{-1}(8)) - \text{mean}\{D(f_j^{-1}(8)) : j = 1, \dots, 112\}}{\text{SD}\{D(f_j^{-1}(8)) : j = 1, \dots, 112\}}$$

where mean $\{D(f_j^{-1}(8)) : j = 1, \dots, 112\}$ denotes the mean of the values in the braces and $\text{SD}\{D(f_j^{-1}(8)) : j = 1, \dots, 112\}$ denotes their standard deviation. Figure 2 compares the SDSs for leg height for a child based on chronological ages (chronological SDSs) and the SDSs based on longitudinal ages (longitudinal SDSs). The plot corresponds to a child whose puberty finished 2.65 years earlier than average and for this reason there is a sharp peak in the chronological SDSs at about 11 years. This peak has largely been removed in the plot for the longitudinal SDSs. An examination of a large number of plots as in figure 2 above led us to conclude that this method of age adjustment was adequate. We first considered using the Age at Peak growth Velocity (APV) in place of T_{end}. However, those children having very young APVs had sharp jumps in the size of their resulting longitudinal SDSs at ages around the average APV, indicating that for these children this adjustment was not appropriate.

The longitudinal SDSs are calculated for each child on a grid of ages. (For the girls at ages 12 and 18 months, then yearly until age 9, half-yearly between the ages of 9 and 16 and then yearly until age 20. For the boys again at the ages 12 and 18 months, then yearly until age 10, half-yearly between the ages of 9 and 17 and then yearly until age 20.) The grids were chosen to correspond to the ages at which most visits were planned, which in turn were chosen so that more frequent measurements would be made in age intervals during which growth velocity was high.

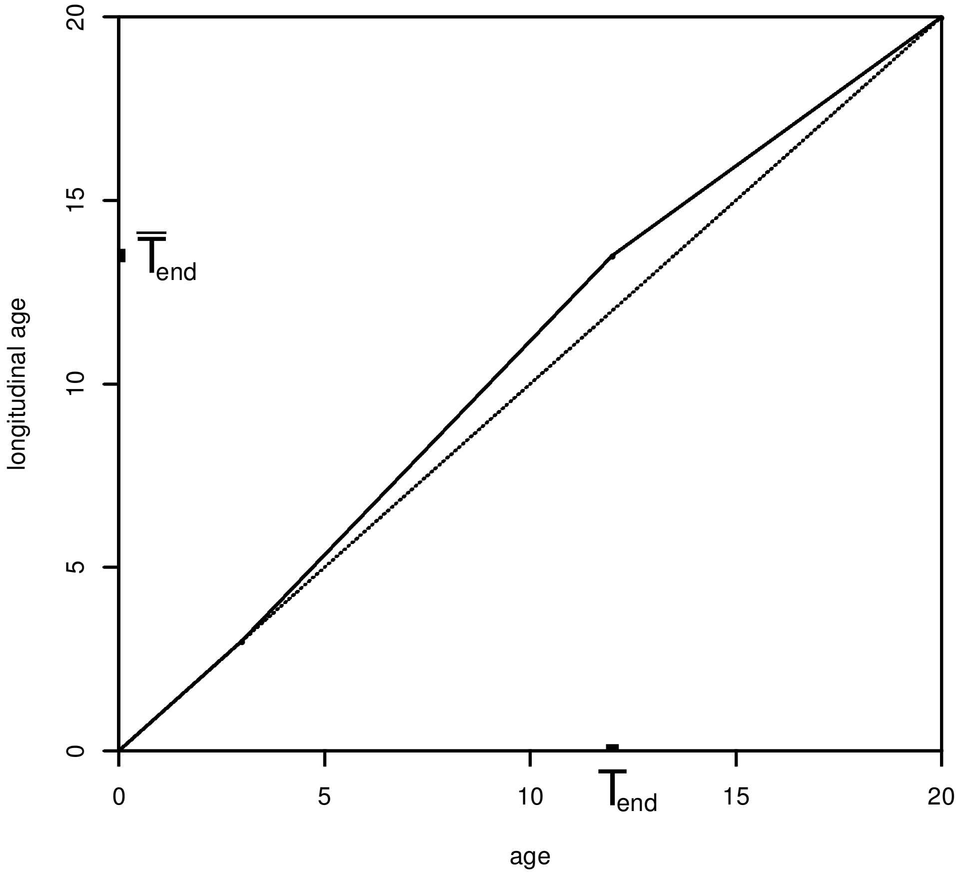

Figure 1. Longitudinal age versus age for a girl's standing height for whom $T_{end} = 12$. For standing height for girls $\overline{T}_{end} = 13.51$. The broken (45°) line corresponds to the transformation for a child for whom $T_{end} = \bar{T}_{end}$.

Our principal motivation in writing this paper was to assess the tracking behaviour in the growth of children. Segal and Tager (1993) provide a nice summary of existing methods for quantifying tracking behaviour and discuss problems associated with indices of tracking (some of which we have here: for example, the dependence on the time interval over which the index is calculated and problems with the interpretation of the size of the index). Ware and Wu (1981) consider the problem of tracking in the framework of prediction of future values from past values. They assume a multivariate normal distribution for the measurements. McMahan (1981) also considers the case where the (repeated) measures can be modelled (conditional on the covariates) assuming a linear model and a normal distribution. Foulkes and Davis (1981) define an index of tracking based on the probability that the fitted SDS curves of two randomly selected individuals will not cross each other. Goldstein (1981) uses the standard deviation of the SDS's for a given child across age as a measure of tracking. We will now describe the motivation and definition of the LTV, which we use to measure tracking behaviour and explain why this index seems more appropriate for our purposes than the aforementioned indices.

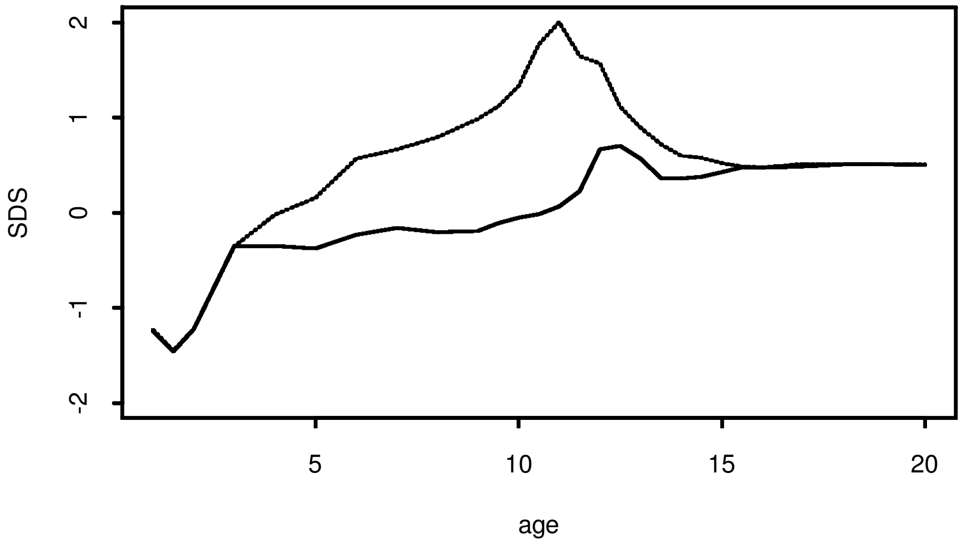

Figure 2. SDS's versus age for a child whose PS ends 2.65 years earlier than average. The broken line represents chronological SDS's, while the solid line represents longitudinal SDS's.

In what follows we will work only with longitudinal SDS's. In figures 3 and 4 examples are given of plots of SDS's against longitudinal age for two children. We will refer to such plots as 'observed tracking curves'. This is done for each of the six variables sitting height, leg height, arm length, biiliac width, bihumeral width and BMI. In figure 3, the SDS's for leg height and sitting height increase gradually from about 0.5 at age 1 to about 1.5 at age 7 and then remain relatively constant. The SDSs for arm length are consistently lower and drop to about 0.25 at age 15 years and remain fairly constant from thereon. SDS's for bihumeral and biiliac width also increase from age 1 until about age 7 and vary about 1 from then on. Notice the large amount of short-term variation in the bihumeral width values: this is partly due to the fact that SDS's for bihumeral width are calculated with respect to a much smaller sample than all the other variables but is also probably due to the fact that bihumeral width (as is also the case with BMI) is strongly affected by exogenous factors. It is also interesting to note that at adulthood, while sitting and leg height have SDS's of about 1.5, arm length has an SDS of about 0.25, and biiliac and bihumeral width have SDS's of about 1, indicating that the morphology of this adult varies somewhat from the average.

The child in figure 4 has an SDS which appear to converge to approximately the same adult value of -1 across age for leg and sitting height and to the value of 0 for bihumeral and biiliac width. These two examples are not atypical and a quick examination of a sample of such plots reveals, not suprisingly, that almost anything is possible! The methods of Ware and Wu (1981) and McMahan (1981) do not seem to be appropriate here. The method of Goldstein (1981) is appealing because it does not require any distributional or model assumptions about the data. However, variation due to short-term fluctuations is not separated from variation due to long-term variation. The method that we use here does this.

We write $Z_{ik}(t_j)$ for the SDS of the ith boy (or girl) at the longitudinal age t_j, for the kth variable. We assume that the $Z_{ik}(\cdot)$s can be modelled as follows:

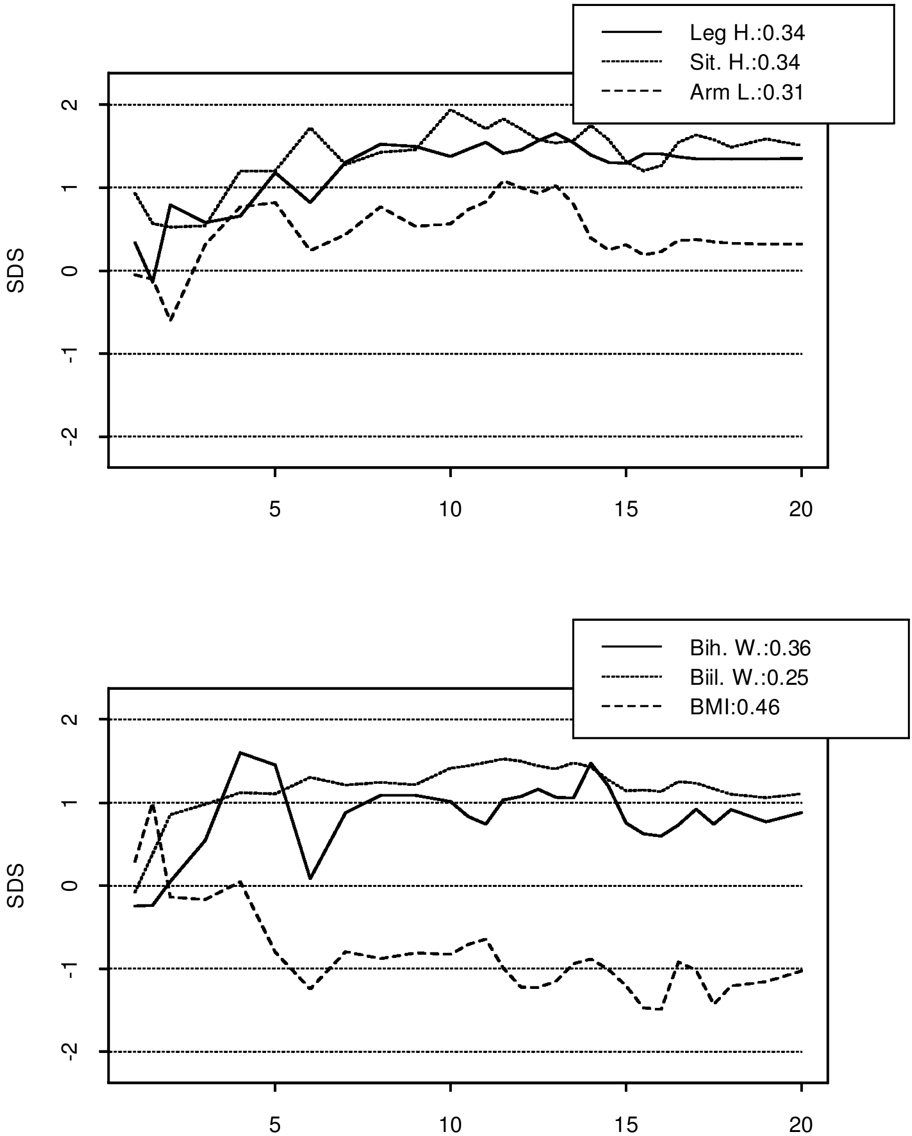

Figure 3. Longitudinal SDSs for a boy. Also given are the values of the square root of the LTV for each of the six variables.

$$Z_{ik}(t_j) = u_{ik}(t_j) + \delta_{ik}(t_j)$$

where u_{ik} is some smooth trend and the $\delta_{ik}(t_j)$s are independent random variables with mean 0 and variance σ_{ik}^2, not depending on age. The function u_{ik} models the long-term changes in SDSs from birth to 20 years and the subscript *ik* indicates that its form can vary from child to child and from one variable to another. The $\delta_{ik}(t_j)$s represent the short-term fluctuations. The STV estimates the value σ_{ik}^2. We write

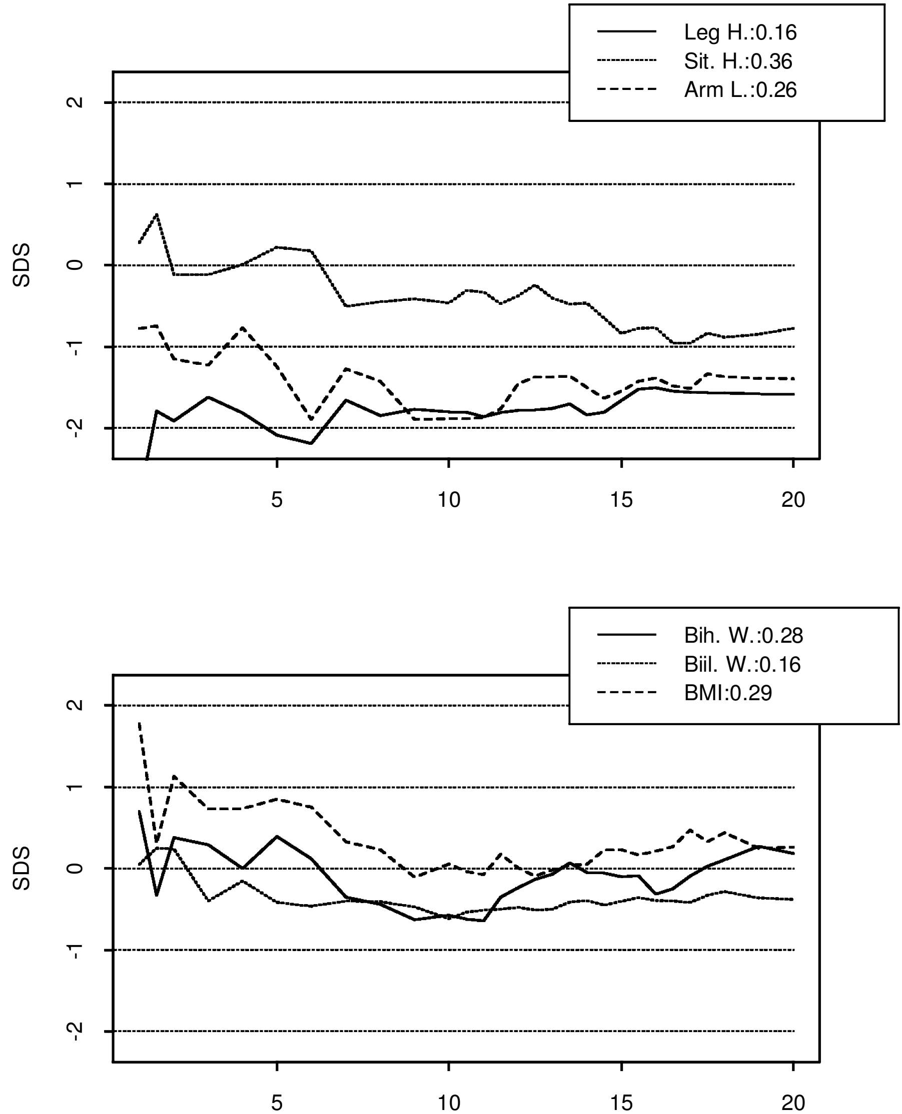

Figure 4. Longitudinal SDSs for a boy. Also given are the values of the square root of the LTV index for each of the six variables.

$$D^2_{0,ik}(1,20) = \frac{1}{20-1}\int_1^{20}\left(u_{ik}(t) - \frac{1}{20-1}\int_1^{20} u_{ik}(s)\mathrm{d}s\right)^2 \mathrm{d}t,$$

the variation in the regression curve between the ages of 1 and 20 years for the SDS's of the ith child and the kth variable; the LTV estimates this quantity and will be used to assess tracking behaviour. Large values of the LTV correspond to poor tracking behaviour. The following remarks are intended to facilitate interpretation values of the LTV:

- When a child stays at a constant SDS between the ages of 1 and 20 this quantity is zero.

- In general, one could consider the variation in the curve u_{ik} over an arbitrary interval $[a, b]$ by replacing the limits on the integrals above with the numbers a and b and the constant $1/(20-1)$ by $1/(b-a)$. When comparing values of the LTV across intervals of different length one can compare values of the square root of the LTV divided by the length of the interval: this quantity measures (heuristically) the size of the changes in the SDS's per year.
- If the function u_{ik} is well approximated by a straight line, then the LTV is approximately equal to a constant times the square of the slope of this line. This constant is equal to $(20-1)^2/12$.
- If for the ith child the function u_{ik} is a constant c times the function u_{jk} associated with the jth child (if $c > 1$ this implies greater long-term fluctuation for the ith child), then the LTV for the ith child will be c^2 times the LTV for the jth child.

The STV is defined to be the nonparametric variance estimator described in Gasser, Sroka and Jennen-Steinmetz (1986). We now give a definition for the LTV, which estimates $D^2_{0,ik}$. Let $1 = t_1 < t_2 < \cdots < t_{n-1} < t_n = 20$ be the grid of points on which the longitudinal SDSs are calculated. Write $\Delta_i = (t_{i+1} - t_{i-1})/2$ for $i = 2, \ldots, n-1$ and $\Delta_1 = (t_2 - 1)/2$ and $\Delta_n = (20 - t_{n-1})/2$. Then for the ith child and the kth variable

$$\mathrm{LTV} = \sum_{j=1}^{n} \Delta_i (Z_{ik}(t_j) - \overline{Z_{ik}})^2 \Bigg/ \left(1 - \sum_{1}^{n} \Delta_i^2\right) - \mathrm{STV}$$

where $\overline{Z_{ik}} = \sum_{j=1}^{n} Z_{ik}(t_j)/n$. One can show that if the grid were equally spaced then Goldstein's growth constancy index (Goldstein 1981) for a group of children is roughly equal to one minus the average, taken over the group, of the sum of the quantities LTV and STV. It seems clear though that one should look separately at these two sources of variation.

There are two main sources of error in this procedure: errors in the estimation of the parameter T_{end} and errors in the size of the SDS's due to the fact that they are estimated with respect to a sample rather than with respect to a population. The first kind of error will be reflected in deviations of the observed regression function from the true regression function and the second will be reflected in an increased variance for the error term. As we do not expect these errors to be either variable or sex dependent, while they may cause bias in the LTV's and the STV's they should not have an effect on comparisons made across variables or between sexes. One exception to this argument is bihumeral width: as explained in section 2.2, there is a smaller sample size for this variable and as a consequence the values of the STV will be more inflated than the other six variables considered here.

We use boxplots to display observed values of the LTV and the STV. The centre line through the box marks the median of these values. The upper and lower extremities of the box-correspond to the 75th and 25th percentiles, respectively, of the data. The 'whiskers', extending from the ends of the box, encompass values of the data which are not extreme, assuming the data to be normally distributed. Outlying points are marked with a circle.

As discussed in section 1, we are interested in differences in magnitude of the indices between boys and girls and across the six skeletal size variables. Because the values for bihumeral width are not complete for a large subset of children, we restrict ourselves to the remaining five skeletal size variables for the purpose of testing for statistical significance of differences. To do this we used a repeated measures analysis of variance (ANOVA). For each child there are five repeated measures (the five LTVs corresponding to the five skeletal size variables). A similar analysis is carried out for the STVs. The LTVs and the STVs have highly skewed distributions. We found that a cube-root transformation of the indices led to a reasonably symmetric distribution and this transformation was used for the ANOVA described above.

The BPV for a given child at a given longitudinal age is defined to be the variance of the SDSs of the variables standing, leg and sitting height, biiliac width and arm length. BMI is excluded from this analysis because it is not directly related to skeletal growth, and bihumeral width is excluded because of the high percentage of missing values.

3. Results

Figure 5 displays boxplots for values of the square root of the LTV for each of the seven variables, for boys and girls separately. For bihumeral width, leg height and BMI the median of the LTV is about the same for both boys and girls and for all other variables it is greater for girls. Overall, bihumeral width, BMI and sitting height have higher values for the LTV than biiliac width, standing, and leg height. The values for three children have been omitted because they were outliers: two for BMI and one for sitting height. Plots of the SDS's versus longitudinal age for these children gave no indication that this was due to errors in the age transformation.

In figure 6, we display boxplots of the STV. Differences between boys and girls with respect to the STV are small. Bihumeral width has, on average, by far the largest value for the STV, and standing height has the smallest, followed by biiliac width. The values of the STV for leg height, arm length, BMI and sitting height are, on average, all about the same and are intermediate in size.

As explained in the method section, we use a repeated measures ANOVA on the cube root of the LTVs to test for significant differences in sizes of the LTVs and STVs between boys and girls and across all skeletal size variables, except bihumeral width. Statistically significant differences between means of the transformed values correspond to statistically significant differences between medians of the original values. In figure 7, the mean values are plotted—for each of the sex $\times$ variable subgroups—for the transformed indices (for the sake of completeness we include all seven variables in this display though only five variables were involved in the ANOVA). Lines joining adjacent points are drawn in order to display the dependence of the sex differences on variable: parallel lines between two variables indicates that mean differences between boys and girls are the same for these two variables.

For the LTV, differences between boys and girls are found to be statistically significant ($p = 0.017$) as are differences across variables ($p < 0.0001$). However, the p-value for the interaction term sex $\times$ variable is not significant. For the STV, sex differences are not significant, and differences across variables are highly significant ($p < 0.0001$). Again, the interaction term sex $\times$ variable is not significant.

Table 1 gives rank correlations between the LTVs for the variables leg and sitting height, arm length and biiliac width. Height is not included because of its being the sum of leg and sitting height. The LTV for biiliac width shows moderate positive correlations with the other variables for the girls and negligible correlations for the

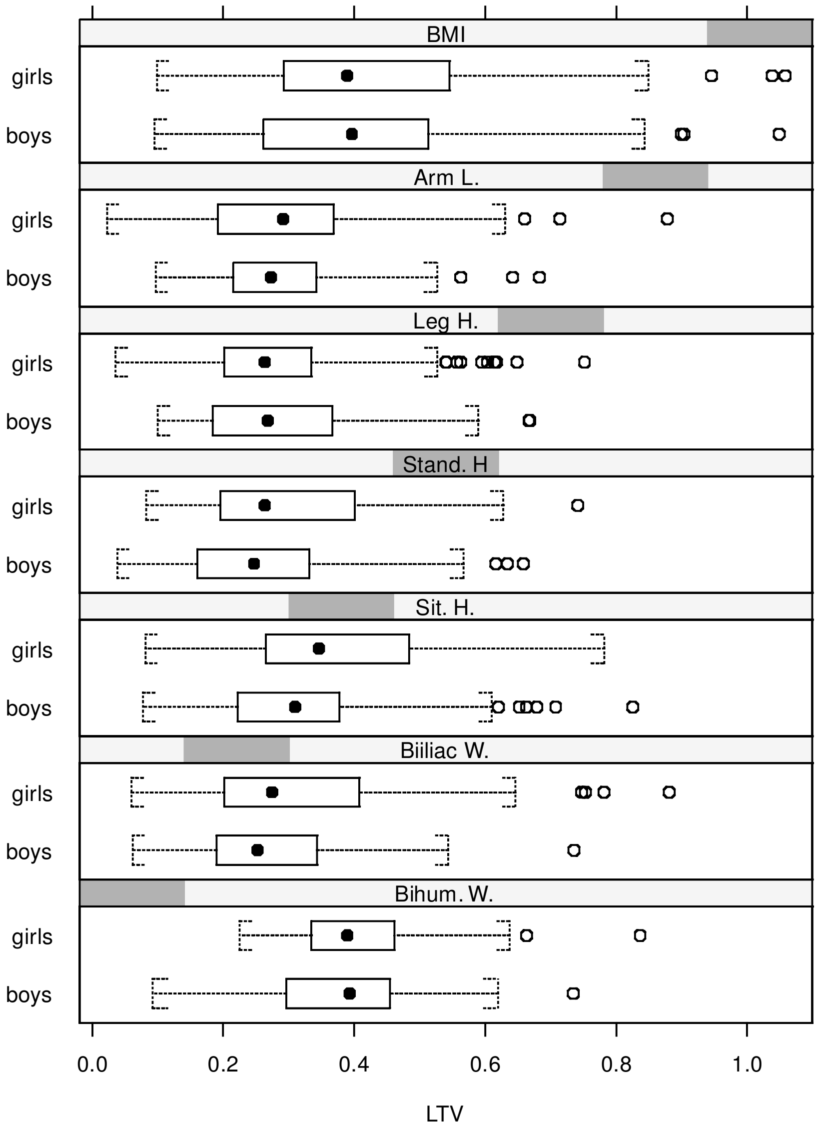

Figure 5. Boxplots of values of the square root of LTV for each of the variable × sex groups.

boys. Correlations between variables of length are sizeable for both sexes, in particular when one considers that the LTV can only be determined with some statistical error and this inevitably leads to lower observed correlations.

We also calculated, on a variable by variable basis, rank correlations between the LTV and the age at which the peak growth velocity occurred (loosely speaking, the timing of the PS), between the LTV and adult size and between the LTV and the maximum growth acceleration associated with pubertal growth. All correlations were very small so that it may be concluded that there is no relationship between

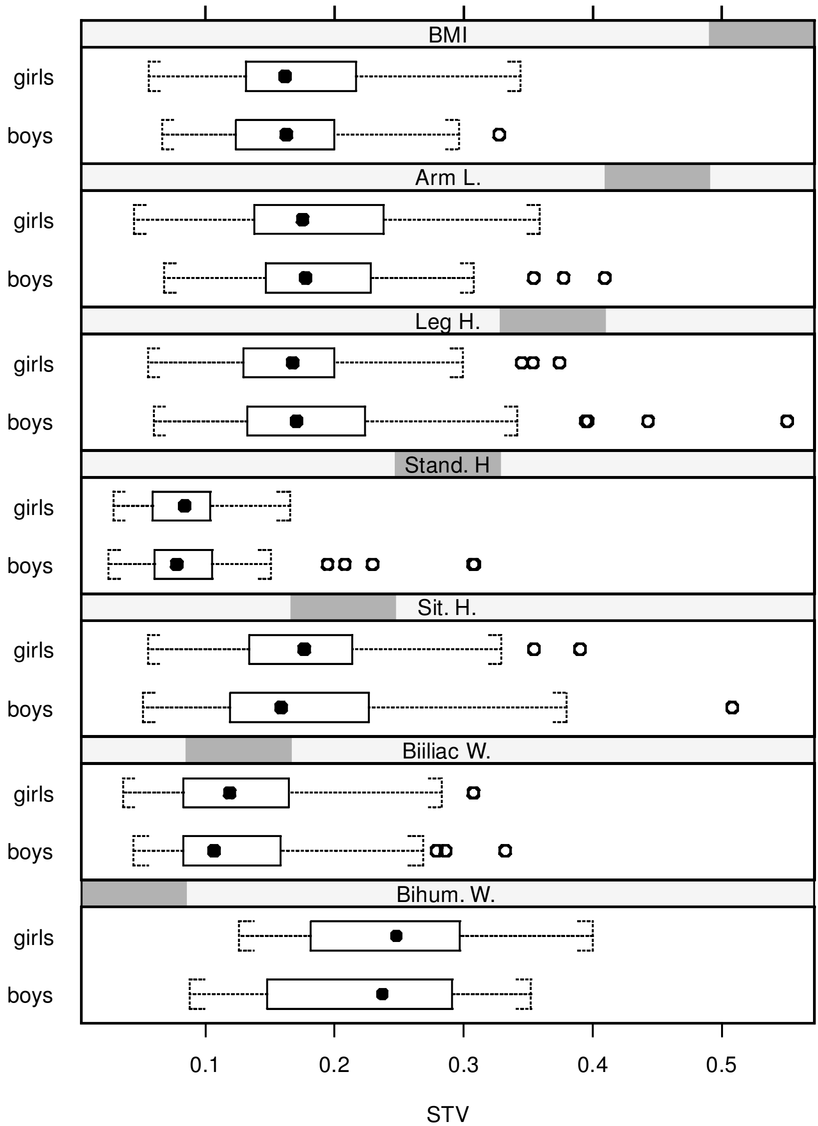

Figure 6. Boxplots of values of the square root of STV for each of the variable × sex groups.

tracking behaviour and timing or intensity of the PS, or between tracking behaviour and adult size: these three parameters being the most important for characterizing growth.

In an attempt to understand whether there was evidence to support the 'decreasing BPV hypothesis' as described in the Introduction,in figure 8 a plot is given of the (slightly smoothed) curve of the average, across children, of the square root of the BPV, for boys and girls separately (see methods section for a definition). There is a clear decrease in this value with increasing age, reaching a minimum at about the average age of onset for the PS. (In Sheehy, Gasser, Molinari *et al*. (2000a) it was

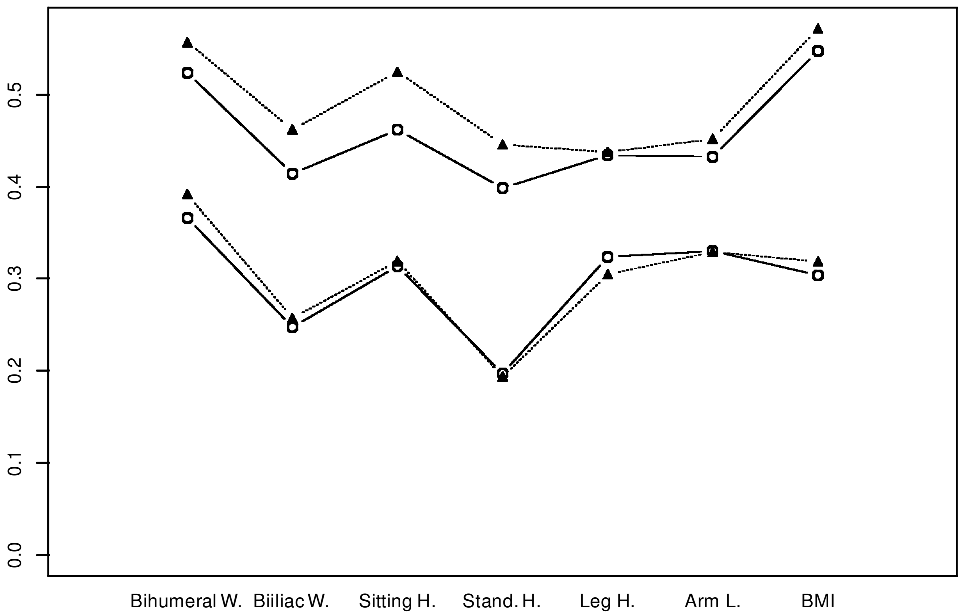

Figure 7. Means of the (transformed) indices for each of the seven variables (the LTV is above, the STV is below). The broken line represents girls and the solid line represents boys.

Table 1. Rank correlations between LTV's: boys upper triangle, girls lower triangle.

	Biiliac width	Sitting height	Leg height	Arm length
Biiliac width	–	0.11	0.16	−0.01
Sitting height	0.30	–	0.49	0.53
Leg height	0.24	0.67	–	0.18
Arm length	0.35	0.43	0.35	–

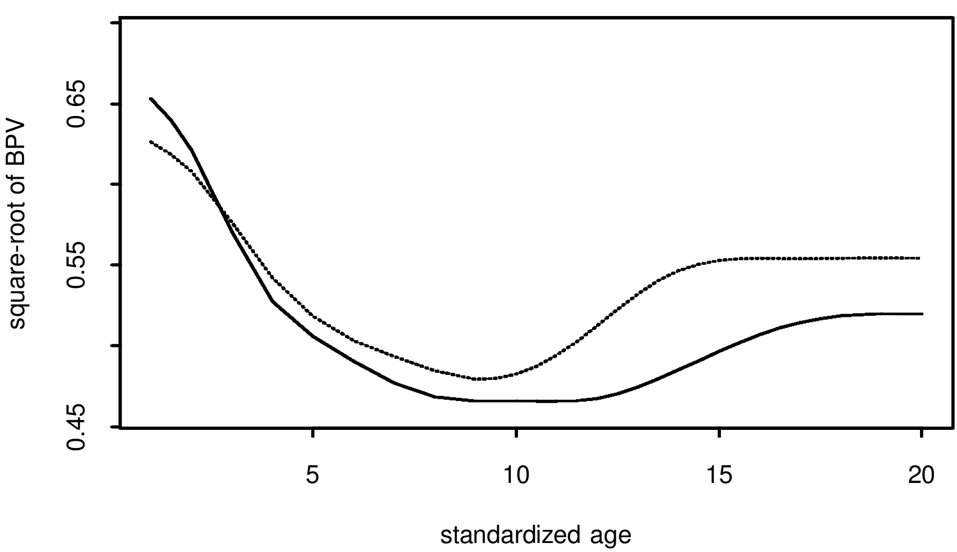

Figure 8. Average of the square root of BPV versus age. The broken line represents girls and the solid line represents boys.

shown that for standing height this average is equal to 11.1 years for boys and 9.8 years for girls.) Following this, there is an increase until a plateau is reached after the PS is over. Interestingly, this average is greater for girls than for boys at every age after about 3 years. This sex difference was not found to be statistically significant. For boys the average at age 1 year and the average at age 11 years were found to be highly statistically significantly different from each other ($p < 0.001$) and similarly, for the girls the average at age 1 year and the average at age 10 years were found to be highly statistically significantly different from each other ($p < 0.001$).

4. Discussion

The LTV measures the stability of the growth process of a given individual with respect to a sample. Because maturational tempo is a confounding factor for tracking of growth, we have used longitudinal ages which adjust for the inter-individual variation in the timing of the PS.

Boys show better tracking than girls (they have on average smaller values for the LTV) and this difference is statistically significant. While we have no truly biological explanation for this finding, it is in line with previous results (Sheehy *et al.* (2000b): girls show lower correlation of size attained with adult size across the whole age span, when compared to boys).

Variables differ substantially with respect to their tracking behaviour. As one would expect, BMI, has on average, the largest LTV values. This is to be expected given the fact that while BMI on the one hand has a small measurement error, on the other, it is strongly influenced by exogenous factors. Bihumeral width also has relatively large values for the LTV: this can be attributed to the fact that its size is affected by deposits of fat and muscle tissue and so, like the BMI, is strongly influenced by exogenous factors. There are also substantial differences between the remaining five 'hard' variables (height, leg and sitting height, arm length and biiliac width) which were found to be highly statistically significant when tested using ANOVA. In particular, sitting height has much poorer tracking than all the other 'hard' variables. This must in some way be attributable to a particular pattern of bone growth for the spine—though what that is, is unknown to us at present. This finding, is however, consistent with table 2 in Sheehy *et al.* (2000b), where the amount of variability explained by prepubertal increments was found to be least for sitting height.

What explains the fact that for a given variable, one child shows good tracking and another child poor tracking? We found that the LTV is uncorrelated with adult size, with the timing and with the intensity of the PS. Note that these are the most important characteristics of the growth process. We speculate that variation in the LTV has a large genetic component: due to the individual genetic programme, given two children with the same adult size, one child might gain more than the average in some prepubertal period, while the other achieves the same adult size by a modest increase prepubertally and a relatively large contribution due to an intense PS.

We have also checked the so called 'decreasing BPV hypothesis' (see the Introduction). The variation in proportions decreases after age 1 year and a minimum is reached at about the age of the onset of puberty. Thus, proportions become more homogeneous over the prepubertal period, indicating that the genetic programme of growth is specific for different parts of the body. During and after puberty, it increases again, and slightly more so for girls (we see again that girls are more variable than boys). We attribute this to the varying contributions of the PS to adult size in different variables: for girls, the larger increase in the BPV is probably due to the sizeable inter-individual variability in the contribution of the PS to biiliac width.

For the STV, average differences between boys and girls were not found to be statistically significant, however there are large differences between variables. It is not surprising that this average is highest for bihumeral width: the smaller sample with respect to which the SDS's are calculated is certainly a large factor here and as discussed in section 2.2, the fact that the measurement error is large, compared with the other variables, is also a factor. Recall that the STV for the ith child, estimates σ_i^2, the variance of the error term in the regression model for the SDS (we supress the dependence on variable here). We can write,

$$\sigma_i^2 = \sigma_i^2(\text{sampling}) + \sigma_i^2(\text{msmt error}) + \sigma_i^2(\text{pure STV})$$

where σ_i^2 (sampling) represents the variance in the error term due to estimating the SDS from the sample, σ_i^2 (msmt error) represents the variance in the error term due to measurement error and σ_i^2 (pure STV) represents the variance in the error term due to other factors. For a child whose true SDS is 0 we can (using simulations or using moments of the non-central t distribution) calculate σ_i^2 (sampling) $= 0.019$, for $n = 53$ (the number of boys in the sample having complete data for bihumeral width) and σ_i^2 (sampling) $= 0.008$, for $n = 120$ (the number of boys in the sample having complete data for all other variables). The median value for the STV for bihumeral width for boys is 0.056, while the median value for sitting height is 0.025. So for an average boy, for bihumeral width we can estimate that the part of σ_i^2 *not* due to sampling error is $0.056 - 0.019 = 0.037$, while for sitting height it is $0.025 - 0.008 = 0.017$. This argument makes it clear that the larger values of the STV for bihumeral width are not only due to the increased sampling error for this variable.

That the STV for the BMI is high is also not surprising—while BMI has a small measurement error it is clearly subject to much more short-term fluctuations than any of the other variables. That arm length is high may be partly attributable to the fact that measurements were always rounded to the nearest centimetre and so the measurement error is inevitably rather large. However, no such explanation is available for the fact that values for sitting and leg height are about the same size as those for arm length. That standing height is low can be explained by the fact that it is the sum of two variables: leg and sitting height. We have no explanation for why biiliac width has so much less short-term variability than arm length, sitting and leg height.

Acknowledgements

Anne Sheehy was supported by a Marie Heim-Voegtlin Fellowship from the Swiss National Science Foundation. The support of the Swiss National Science Foundation for project no. 3200-045829.95/2 is also gratefully acknowledged.

References

Falkner, F., 1960, *Child Development: An International Method of Study* (Basel: Karger).

Foulkes, M., and Davis, C., 1981, An index of tracking for longitudinal data. *Biometrics*, **37,** 439–446.

Gasser, T., Müller, H., Köhler, W., Molinari, L., and Prader, A., 1984, Nonparametric regression analysis of growth curves. *Annals of Statistics*, **12,** 212–229.

Gasser, T., Sroka, L., and Jennen-Steinmetz, C., 1986, Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73,** 625–633.
Gasser, T., Ziegler, P., Prader, A., Molinari, L., and Largo, R., 1994, Measures of body mass and of obesity from infancy to adulthood and their appropriate transformation. *Annals of Human Biology*, **21,** 111–123.
Goldstein, H., 1981, Measuring the stability of individual growth patterns. *Annals of Human Biology*, **8,** 549–557.
McMahan, C., 1981, An index of tracking. *Biometrics*, **37,** 447–455.
Molinari, L., Gasser, T., Largo, R., and Prader, A., 1995, Child–adult correlations for anthropometric measurement. In *Essays on Auxology*, edited by R. Hauspie, G. Lindgren and F. Falkner (Welwyn Garden City, UK: Castlemead Publications), pp. 164–177.
Prader, A., Largo, R., Molinari, L., and Issler, C., 1988, Physical growth of Swiss children from birth to 20 years of age. *Helvetica Paediatrica Acta (Suppl.)* **52,** 1–125.
Segal, M., and Tager, I., 1993, Trees and tracking. *Statistics in Medicine*, **12,** 2153–2168.
Sheehy, A., Gasser, T., Molinari, L., and Largo, R., 2000a, An analysis of variance of the pubertal and midgrowth spurts for length and width. *Annals of Human Biology*, **26,** 309–331.
Sheehy, A., Gasser, T., Molinari, L., and Largo, R., 2000b, Contributions of growth phases to adult size. *Annals of Human Biology*, **27,** 281–298.
Ware, J., and Wu, M., 1981, Tracking: prediction of future values from serial measurements. *Biometrics*, **37,** 427–437.

Address for correspondence: Anne Sheehy, Abteiling Biostatistik, ISPM, University of Zurich, Zurich, Switzerland.

Zusammenfassung. *Ziel:* Charakterisierung der long-term und short-term Variablität der Standard Deviation Scores (SDS) bei sechs Skelettmaßen und beim Body Mass Index (BMI) von Kindern und Vergleich der Mittelwerte dieser Variablen zwischen Jungen und Mädchen sowie Vergleich der Variablen untereinander.
Methoden und statistische Verfahren: Die Analyse basiert auf den Messungen von sieben Variablen (Körperhöhe, Sitzhöhe, Beinlänge, Armlänge, Schulterbreite, Bihumeralbreite und BMI), die regelmäßig an 120 Jungen und 112 Mädchen im Alter von einem Monat bis 20 Jahre im Rahmen der Ersten Züricher Longitudinalen Wachstumsstudie durchgeführt wurden. Die Variation des Scores, die durch die Variablität beim zeitlichen Auftreten des puberalen Wachstumsspurts (PS) bedingt ist, wird durch die Umskalierung der Altersachse auf individueller Basis eliminiert, so dass Kinder mit gleichem Entwicklungsalter anstatt gleichem chronologischen Alter verglichen werden. Für ein bestimmtes Kind wird das Verhältnis zwischen dem Wert seines SDS und dem Alter als Summe aus einer willkürlichen (kinderabhängigen) geglätten Funktion und einem Fehlerwert gebildet. Die long-term Variabilität für dieses Kind wird durch die mittlere quadratische Abweichung von der geglätteten Funktion definiert, während die short-term Variabilität durch die Varianz des Fehlerwertes definiert ist.
Ergebnisse: Die SDS-Werte der Mädchen weisen eine signifikant größere long-term Variabilität auf als die der Jungen, während es keinen signifikanten Unterschied zwischen den Geschlechtern bei der short-term Variabilität gibt. Bihumeralbreite, BMI und Sitzhöhe weisen eine signifikant größere long-term Variabilität als die anderen Variablen auf. Bihumeralbreite und BMI haben die größte short-term Variabilität und Körperhöhe die kleinste. Die Korrelationen zwischen long-term Variabilität und Erwachsengröße und zeitlichem Auftreten und Intensität des PS waren niedrig.
Schlissfolgerungen: Eine hilfreiche Methode zur Bestimmung der long-term und short-term Variabilität von SDS-Werten, welche breit anwendbar ist wird vorgestellt und für Wachstumsdaten von Kindern angewandt. Die Ergebnisse dieser Analyse sind faszinierend. Warum verläuft der Wachstumprozeß der Mädchen variabler als der der Jungen? Die Unterschiede zwischen den Skelettmaßen sind ebenfalls interessant und verdienen weitere Aufmerksamkeit.

Résumé. *Objectif premier:* Qualifier la variabilité à long et court terme des scores d'écart-type (SET) de six variables de format squelettique et de l'indice de masse corporelle (IMC) d'enfants et comparer leurs valeurs moyennes chez les garçons et chez les filles, ainsi que comparer les variables entre-elles.
Méthodes et procédures: L'analyse qui s'inscrit dans la première Etude Longitudinale de Croissance de Zurich, est fondée sur des mesures effectuées régulièrement sur 120 garçons et 112 filles, de 1 mois à 20 ans pour sept variables : stature, taille-assis, longueur de la jambe, longueur du bras, largeur biiliaque, largeur bihumérale et indice de masse corporelle (IMC). La variation de ces scores due à la variabilité de la chronologie de la poussée pubertaire (PP) est singularisée par le rééchelonnement de l'axe des âges sur une base individuelle et en comparant les enfants d'âge de développement semblable plutôt que de même âge chronologique. Pour un enfant donné, la relation entre la valeur de son SET et l'âge est modélisée comme la somme d'une fonction régulière arbitraire (dépendant de l'enfant) plus un coefficient d'erreur. La variabilité à long terme pour cet enfant est définie comme le carré moyen des écarts de cette fonction régulière à partir de son niveau moyen, tandis que la variabilité à court terme est définie comme la variance du terme d'erreur.

Principaux résultats: Les SET des filles ont une variabilité à long terme significativement plus grande que celle des garçons, tandis qu'on n'observe pas de différences sexuelle de variabilité à court terme. La largeur bihumérale, l'IMC et la taille assis ont des variations à long terme significativement plus grandes que les autres variables. La largeur bihumérale et l'IMC ont les variabilités à court terme les plus grandes et la stature a la plus faible. Les corrélations entre variabilité à long et court terme, taille adulte, chronologie et intensité de la PP sont faibles.
Conclusions: On a décrit et appliqué à des données de croissance une méthode d'application générale pour rendre compte de manière utile, de la variabilité à long et à court terme des SET.