Chapter 13

Growth References

Date: December 13, 2004

Stef van Buuren
TNO Prevention and Health
P.O. Box 2215
2301 CE LEIDEN
The Netherlands
Email: S.vanBuuren@pg.tno.nl
Voice: +31715181802, Fax: +31715181920

13.1 Motivation for growth references

13.1.1 What is a growth reference?

A growth reference describes the variation of an anthropometric measurement within a group of individuals. A reference is a tool for grouping and analyzing data and provides a common basis for comparing populations.[1]

A well known type of reference is the age-conditional growth diagram. The traditional Height-for-Age (HfA) diagram shows how height varies both within and across age. Figure 13.1 is the official Dutch diagram of height and weight for Dutch boys aged 1-21 years.[2],[3] In the HfA diagram, the vertical distance of shaded area in the graph delineates the variation in heights between -2 and +2 standard deviations (SD). The interval between the -2 and +2 SD curves contains about 95.4% of all individuals of the same age in the reference sample. The graph at the top displays the variation of weight as a function of height. For reasons that will be discussed in section 13.7, the shaded area is chosen here between the -1 and +1 SD curves.

-- INSERT FIGURE 13.1 ABOUT HERE --

Anthropometry is an extraordinary good tool for gauging health and well-being in both individuals and in populations.[4] Height and weight are cheap and easy to measure, and provide an almost universal appraisal for assessing children's well-being. Height is one of the very few positive health indicators. A secular shift in height is a sensitive indicator of socio-economic and socio-medical changes, and thus allows comparisons of the health status of different populations. In fact, secular shifts in height may provide a better and more relevant measure for the detection and evaluation of possible changes in living conditions than such vague concepts as 'income per capita' and 'national product'.[5] Weight is clearly relevant for evaluating both under-nutrition and obesity. Both are important problems on a global level. Besides height and weight, many other useful anthropometric measures exist, but the present chapter will concentrate on height, weight and derivates thereof.

13.1.2 Uses of growth references

Growth references are useful at both the population and the individual level. At the population level, growth references can assist in estimating prevalence, in determining causes of disease, in identifying groups at risk, in monitoring trends, and in evaluating the effects of interventions. On the individual level, growth references are indispensable tools for screening, diagnosis, monitoring and prognosis of growth-related diseases[6]. Table 13.1 provides an overview of typical questions that one might ask from growth references. In addition to these health-related applications, growth references are also needed in ergonomic design, for deriving safety guidelines, for establishing appropriate matches between garment size and the client population, for determining construction requirements, and so on.

-- INSERT TABLE 13.1 ABOUT HERE --

13.1.3 Standards and references

A subtle but important distinction exists between the concept of a 'reference' and a 'standard'. A growth reference simply describes the variation of some measure within a reference population, often conditional on age and sex. A growth standard, on the other hand, delineates the variation that is considered to be normal, optimal, or healthy. Standards embrace the notion of a norm or desirable target, and thus involve a value judgment.[1] The distinction between (normative) standards and (descriptive) references is often blurred because standards are frequently derived from references, but the conceptual distinction is important.

Clinical practice requires standards. A basic clinical question is: Is this child normal? Answering this question involves comparing the individual child to a standard. It is not always obvious which reference should be selected as a standard. The WHO recommends choosing 'references that resemble, as far as possible, true standards, so that the same deviation from the reference data has the same biological meaning'.[1] Breast-feeding is an example where this occurs. Breast-feeding is associated with positive health outcomes, and considerable evidence exists that growth of infants with exclusive breast-feeding is different from that of formula-fed infants.[7,8,9] This evidence motivated the development of separate references for exclusively breast-fed infants.

In many situations however, there is no cut-and-dried answer. Evidence is often scant or inconsistent, so it might be unclear what reference to use as a standard. Body mass index (BMI) provides an example where a (descriptive) reference might not be a good standard. Like in the rest of the industrialized world, the prevalence of overweight and obesity in The Netherlands substantially

increased between 1980 and 1997.[10,11] The Dutch references for BMI in 1997 accurately describe the growth expected in the 1997 population, but since a rise in BMI brings additional health risks, the 1997 references should not be used as some optimal standard. A better alternative is base any BMI standard on references from populations that existed before the recent obesity epidemic. Examples of such references include the IOTF international criteria for overweight and obesity[12] and the references for (serious) underweight in Dutch children.[13]

13.2 Issues in study design

References are constructed from data obtained from a sample of the reference population. Various choices in the study design affect the ability to construct appropriate references. In general, major choices in study design depend on the following questions:

- Will longitudinal references be needed?
- What is the statistical precision required at each age?
- What are the costs of recruitment and management of the reference sample?
- What resources are available for data collection and statistical analysis?
- In which settings will the references be used?
- When should the new references become available?

13.2.1 Cross-sectional and longitudinal studies

A cross-sectional growth study is a study in which children of different ages are measured once at the same point of time. Cross-sectional studies are relatively easy to carry out. References obtained from cross-sectional studies are perfectly suited for evaluating the status of a child encountered in a situation where only one measurement is made, such as a screening study. Repeated cross-sectional studies provide very sensitive information about changes in the health status of the population. Secular changes of growth are among the most striking biological consequences of social and economic development.[5]

Cross-sectional studies, however, do not provide adequate information for following growth of an individual child over time, in particular at ages where the growth rate is high, as in infancy and during puberty.[4,14] The reason is that references derived from cross-sectional studies do not allow for the child's tempo of growth. Though it is possible to infer the average rate of growth from the difference of average size at increasing age, a cross-sectional study cannot be used to estimate the variability between children in the amount of growth over time. To do so requires a longitudinal study, i.e., a study where children are measured at two or more occasions. In longitudinal data, the growth rate

between the occasions is directly observed for each child. This information can be used to construct longitudinal references that portray the variability in growth between children over time.

Longitudinal studies are generally more difficult to manage, but may be efficient if the amount of effort to identify and locate the individuals in the reference sample is large.[15] A longitudinal study takes more time, so if important secular trends occur, references obtained from long studies may be outdated by the time they are published. On the other hand, it may be more efficient to use a longitudinal design if the costs of managing the cohort are small. Having multiple measurements per individual not only increases the precision of the growth references, but also allows for the construction of longitudinal references.

A mixed longitudinal design represents a compromise between a cross-sectional and a longitudinal design. In a mixed design, individuals are measured more than once, but not throughout the entire age range. Mixed longitudinal designs are quite popular. Examples of recent, large scale mixed longitudinal studies include the Euro-Growth study[16,17] and the WHO Multicentre Growth Reference Study.[18,19] The mixed longitudinal study combines the advantages of the cross-sectional and longitudinal designs, but may have less precision than an exclusively cross-sectional or longitudinal study. In addition, a mixed design is often more difficult to analyse.

13.2.2 Sample size

Sample size is the most important factor affecting the precision of the reference values. Other relevant determinants include the study design, the timing of measurement and the method of curve fitting.[15] The simplest advice is to have at least 200 individuals per age and sex group.[20] Assuming a normal distribution with a standard deviation of 7cm, the standard error of the mean is equal to $7/\sqrt{200} = 0.5$cm. Thus, a sample size of 200 is adequate to detect a secular shift in mean height of about 1.9 cm with a type I error rate of 5% and a power of 80%. The standard errors of a given percentile can be obtained by multiplying the standard error of the mean by a fixed constant that depends on the centile value only.[21] Under normality, these multiplication factors are 1.25 for P50, 1.74 for P90, 1.81 for P95, and 2.51 for P97. So for a sample size of 200, the standard error of 97th centile is equal to 2.5*0.5 = 1.2 cm.

Goldstein[22] argued that the assumption of normality ceases to provide good estimates for the extreme percentiles, even for variables like height. Using a simple ranking method free of the normality assumption leads to sample sizes that are about twice as high to achieve the same precision. Other

specialised ways of calculating sample size have been proposed, dealing with the slope of the median curve, the correlation between measurements, and the precision of the median curve overall.[15]

In longitudinal designs, the primary parameter of interest is the correlation between measurement occasions. The precision of the correlation coefficient depends on both sample size and the expected correlation coefficient. Table 13.2 contains the 95% confidence interval for the correlation for various samples sizes and correlations.

-- INSERT TABLE 13.2 ABOUT HERE --

The use of neighbouring information across age by smoothing methods can drastically increase precision of the overall percentile curves. One could potentially base sample size calculations on the precision of the smoothed curves, but this does not seem to have been practiced yet.

A related question is how the sample should spread across the age range. Traditionally, oversampling at the ages of high growth rates has been used in order to capture accurately the pattern of growth. Approximate sampling fractions per sex are known.[22] As an example, the Fourth Dutch Growth Study[2] constructed age groups such that the population for ages 0-2 years and around puberty were oversampled. In order to improve precision at the edges when using smoothing methods, enough children should be sampled at the extreme ages of the intended references. The WHO Multicentre Growth Study applied a fourfold increase of the sample size at birth, and extended the sample to month 71, whereas the final chart should stop at an earlier age.[19]

13.2.3 Inclusion criteria

Inclusion criteria articulate the reference population. For normative references, the population should live in a healthy environment and contain no overtly sick or very few clinically sick individuals. Eligibility criteria in the WHO Multicentre Growth Study include, amongst others, socio-economic status that does not constrain growth, low mobility of the population, mothers willing to follow feeding recommendations, term birth, single birth, no maternal smoking, and birth weight > 1500 gram.[19]

Inclusion criteria like these take care that children in the reference sample are raised healthy environment. Note however their use makes the reference unsuitable as a descriptive reference. Children not living in healthy environment are systematically missing from the sample, leading to a picture that is too positive. If descriptive references are wanted also (for example in order to monitor

progress on the population level) one could measure all children meeting in the population, and temporarily set aside the data from ineligible children when calculating normative references.

13.3 Distance references

13.3.1 Uses of distance references

The diagrams in Figure 13.1 summarise the distribution of attained size in a known reference population at different ages. A chart of attained size is known as a distance chart. Distance charts are often constructed from cross-sectional data. A distance chart can be legitimately used:

- to compare attained growth between two different populations;
- to compare attained growth of the same population at different occasions;
- to detect aberrant individual growth using a single observation located in an extreme centile.

It is tempting to classify a sequence of measurements of one particular person that 'cross centiles' rapidly over time as unusual. However, the distance chart is inappropriate for assessing longitudinal patterns. The implicit assumption is that children should grow parallel along the centile curves, and that deviations from this pattern indicate unusual growth. Yet, this is not true.[23] The curve of the cross-sectional 50th centile is not the curve actually followed by any individual, even the individual who is at the 50th percentile before puberty, at the 50th percentile after puberty, and who has a spurt at the average time at the average intensity.[4] The distance chart contains no information about changes in centiles from one age to another. Several authors have emphasised this point[4,24,25], but the use of the distance diagram as a tool for monitoring longitudinal measurements is nevertheless widespread.

A number of methods for constructing reference charts (c.f. 13.3.2) allow the calculation of a Standard Deviation Score (SDS), or *z*-score, for each individual measurement. This score is normally distributed with mean 0 and variance 1, and indicates the relative position of attained size at a given age. For a normally distributed measurement SDS can be calculated as $Z = (X - M) / S$, where X is attained size, M is the mean according to the reference at age T, and where S is the corresponding standard deviation. The SDS can also be calculated for other (e.g. skewed) distributions, but the precise method depends on the fitting model that is used to construct the references.

13.3.2 Methods for creating distance references

Producing centile charts has been labeled as 'something of a black art'.[26] This section briefly reviews about 30 methods for fitting centile curves that have been proposed over the last 30 years. The material presented here draws upon more extensive reviews by Wright and Royston[27] and Borghi *et al.*[28].

The major task in centile construction is to smooth the reference distribution in two directions simultaneously, between age and within age. Borghi *et al.*[28] used and extended a classification of methods for references originally put forward by Cole. In this classification, methods are distinguished on the following characteristics:

- estimating centiles separately versus estimating them together. Methods that estimate centiles together can be further classified according to the distributional assumptions made;
- recoding data into age groups versus treating age as continuous;
- the type of age-smoothing method.

13.3.2.1 Basic method

The basic method for deriving distance references consists of three main steps:

- classify the observations into age groups;
- calculate the empirical distribution function from the ordered values of size at each age, and estimate the desired set of centile values, e.g., the 3, 10, 25, 50, 75, 90, 97 centiles, per age group;
- for each desired centile, smooth the centile values across age to form an age-dependent curve. Smoothing can be done by splines functions[29], kernel regression[30,31] or by the eye[32,33].

This method is simple to understand and provides an accurate estimate of the shape of the size distribution across age if the sample size is large. On the other hand, it is not without problems. First of all, creating age groups will bias the variance upwards if the measurements are taken from children whose exact ages differ from the midpoint age. Healy[34] proposed to reduce the variance by a factor $b^2/12$, where b is the average amount of growth occurring during the interval. Even if all children would be measured at the same exact calendar age, estimating centiles from the empirical distribution function sounds easier than it actually is. Care is needed in computing extreme centiles, as some interpolation is needed. The algorithms implemented in SPSS and SAS can give odd results, producing estimates that are too extreme, irrespective of the interpolation algorithm used. The S-Plus function *quantile*() does not have this problem.[35] Finally, because the method does not use neighboring

information, the influence of sample fluctuations can produce centile curves that could potentially cross each other.[36]

13.3.2.2 Distributional assumptions

The basic method can be improved by incorporating the assumption that size measurements within an age group follow some distribution. This assumption facilitates assessment of asymptotic behaviors and provides simple formulas for calculating SD scores. Of course, these advantages will only be realised if the assumed distribution fits the data. Methods have been developed using the normal distribution[37], possibly obtained after applying a Box-Cox transformation[26,38], a square-root transformation[39], or a variance stabilizing transformation[40].

13.3.2.3 Age as a continuous variable

Another type of improvement involves using age as a continuous variable. Koenker and Bassett[41] proposed to estimate a given regression quantile directly from the data, conditional on one or more covariates. The method is elegant because it combines both within and across age variation into one optimisation function. The method estimates quantile curves separately, which could unfortunately result in curves that touch or cross each other. The method has been adapted in several ways to prevent quantile crossing.[42,43,44] Related non-parametric methods have been developed by Wellek and Merz[45], Rossiter[46], Gasser *et al*[83] and Gannoun *et al*[47]. The HRY method[48] estimates each centile through a moving window using a high-order polynomial function. Constraints are applied to the coefficients of the polynomial so that these vary smoothly with the percentiles, thereby ensuring commonality on the centiles so that the curves do not cross. A problem with this procedure is that a fixed polynomial may not be sufficiently flexible if smoothing is applied to a wide age range. Pan and Goldstein developed extensions that improved upon the original formulation of the HRY model.[49,50]

13.3.2.4 Distributional assumptions and age as a continuous variable

Many methods have been developed that incorporate both improvements. Methods differ in the type of the assumed age-conditional distribution, and can be broadly classified into three groups:

- methods assuming a normal distribution;
- methods assuming a normal distribution after transformation;
- methods assuming a non-normal distribution.

In the first group, Aitkin[51] proposed to use maximum likelihood estimation of a linear model for the mean and a log-linear model for the variance. Altman[52] modeled the absolute residuals about the fitted mean as a function of age.

The group of methods assuming normality after a transformation is quite large. By far, the most popular method is the extended LMS method by Cole and Green[53]. The LMS method assumes that the age-conditional distribution is normal after a Box-Cox[54] type of transformation. An age-dependent L-curve models skewness, the M-curve portrays how the median attained size varies with age, and the S-curve describes in what way the coefficient of variation depends on age. The L, M and S curves are fitted by maximum penalized likelihood. Centiles are calculated in the transformed scale, and then back-transformed into the original scale. Similar methods have been proposed by Wade and Ades[55] who used a parametric smoothing method, Thompson and Theron[56] who assumed the 4-parameter family of Johnson's distribution[57] to model kurtosis, and Royston[58] who applied a shifted-logarithmic transformation. The method of Tango[59] finds transformations for both size and age such that after transformation size is a linear function of age with constant variance. If successful, the normal model can be applied on the transformed data to obtain centile curves. Royston and Wright[60,61] proposed a parametric variant of the LMS method, called the 'fractional polynomials and exponential transformation' (FPET) method. The distribution of size can be an exponential distribution with three parameters, or a modulus-exponential distribution with four parameters (which models kurtosis). The parameters of the chosen three- or four-parameter distribution are estimated by maximum likelihood. Yee[62] replaced the Box-Cox transformation by the more general Yeo-Johnson transformation[63] towards normality, which amongst others, allows for negative values of size and potentially results in better normal approximations.

Several methods based on non-normal distributions have been developed. Sorribas *et al*.[64] proposed a parametric method based on the flexible S-distribution. This method smoothes the parameters of the distribution across age. Yee[62] pioneered the use of the Box-Cox transformation towards a gamma distribution, which leads to computational benefits. Rigby and Stasinopoulos[65] generalise the LMS method so that the distribution after the Box-Cox transformation is a *t*-distribution instead of the normal. This allows to model leptokurtosis, i.e., a higher mean and thicker tails than the normal distribution. In another generalization of the LMS method, the same authors apply the Box-Cox power exponential distribution, which models skewness and both leptokurtosis and platykurtosis.[66]

13.3.2.5 Choosing a method

By now, the reader may be left somewhat bewildered by the wide array of possibilities. What method should be used in practice? The short answer is: the LMS method. The LMS method is easy to understand, fits reasonably well for many anthropometric measures, provides closed formula for calculating SD scores, comes with good model fitting tools, and has been applied by so many authors

that it has become the de facto standard. On the other hand, the LMS method is not without weaknesses. It assumes that there is no kurtosis, which may not be true. Evidence is accumulating that kurtosis can have a significant effect on the location of the extreme centiles.[67] Non-parametric methods do not make any distributional assumption, but this comes at the expense of a loss of precision and the capability to calculate SD scores. Methods using other transformations (FPET, Yeo-Johnson, Tango, Johnson family) or distributions (S-, *t*- or BCPE distributions) may give more accurate descriptions of the data while accounting for kurtosis. The LMS method is limited to single covariates, and is cumbersome to apply if the analysis sample must split into groups, e.g, in modeling the Weight-for-Height (WfH) relationship for different age groups. The methods of Royston and Wright, Yee, and Rigby and Stasinopoulos can be used with multiple covariates.

Some comparisons of methods have been published[27,68,69,70], but the performance of the more recent methods has not yet been evaluated. The WHO is currently working on a comparison involving the methods of the Box-Cox, the Box-Cox *t*, the Box-Cox-power-exponential, the FPET method, and the Johnson family of distributions. More comparative work is needed to infer whether the newer methods would actually make a difference in practice.

13.4 Velocity references

Standards based on longitudinal data are required when the same child is to be seen on more than one occasion. Conventionally, velocity is calculated as $V = (X_2\text{-}X_1) / (T_2\text{-}T_1)$, where X_1 and X_2 are the attained size at ages T_1 and T_2. Velocity can be expressed as gain in cm/years, or kg/month. Low velocity appears on the distance chart as downwards centile crossing.

An alternative is to define velocity in the rate of change in SDS per time, i.e. $W = (Z_2\text{-}\ Z_1) / (T_2\text{-}\ T_1)$, where Z_1 and Z_2 are the SDS corresponding to X_1 and X_2. This measure standardises the velocity scale to SDS per year, but does not standardise the distribution of W itself. The distribution of W depends on T_1 and T_2. This implies that a deflecting curve with a slope of -1 SDS/year can indicate dramatic growth stunting during childhood, while the same deflection could fall within the normal range during infancy. Interpretations of W that fail to take the age dependency into account might lead to misleading inferences.[71,72] Figure 13.2 provides an indication of the percentages of the children whose velocity is lower than -0.25 SDS/years, as calculated from several longitudinal studies[24,73,74,75]. Note the strong dependency on age.

-- INSERT FIGURE 13.2 ABOUT HERE –

Cole discussed a velocity measure whose distribution is independent of age: $Z_g = (Z_2-Z_1) / SD(Z_2-Z_1)$, where $SD(Z_1-Z_2) = \sqrt{(2-2r)}$, and where r is the correlation between Z_1 and Z_2 in the reference population.[76] Z_g has a standard normal distribution for all pairs of (T_1, T_2), and for that reason Z_g is preferable over W. Depending on the value of r, the conclusions emanating from both indices can be quite different.

13.5 Conditional references

The interpretation of change over time is affected by regression to the mean.[77] An individual with an extreme size on T_1 can expect a less extreme size on T_2. The strength of the effect is determined by the correlation coefficient. Regression to the mean is crucial in the interpretation of repeated observations. A child in a low centile at T_1 will have a greater expected velocity than one of the same age starting from a higher centile. Velocity references do not take this effect into account, and are thus intrinsically flawed.[23]

Healy[78] suggested that conditioning on previous observations may overcome the problem. Cameron[74] and Berkey *et al*.[79] have put this idea to practice. A conditional reference adjusts the population reference for another variable. In longitudinal data, past observations evidently make up relevant information to condition on.[80] Cole[24,76] proposed the conditional gain score $cZ_g = (Z_2-rZ_1) / \sqrt{(1-r^2)}$, which improves upon Z_g since it properly accounts for regression to the mean.

Conditioning may also involve multiple factors. For example, Thompson and Fatti[81] proposed a weight chart for women that adjusted the overall location of the references to individual characteristics like height, age and parity. Clearly, the idea of conditioning can be exploited to develop screening and monitoring tools that might have elevated sensitivity and specificity for detecting growth-related diseases.[15]

-- INSERT TABLE 13.3 ABOUT HERE –

Table 13.3 summarises indices for age-related references, illustrating the pros and cons of each index. It will be clear that the methodology also applies to references that are conditional on other factors, such as weight or parental height.

Velocity and conditional indices rely on two measurements instead of one. A disadvantage is that both are sensitive to measurement error, especially if the period T_2 - T_1 is short. Of these two, the conditional SDS gain is less sensitive to measurement error, and should therefore be preferred. The variance of the conditional SDS gain is always lower than that of the SDS gain, especially for small r.

Methods for estimating conditional references have been described by Cameron *et al*[74], Berkey *et al*[79], Wright *et al*[82], Cole[24], Royston[80], Gasser *et al*.[83], Pan and Goldstein[84,85], Thompson and Fatti[81], Fatti *et al*[86], Wade and Ades[87], and Reinhard and Wellek[88]. It may not be so easy to get the correlation between Z_1 and Z_2 from a tabulated reference. In daily practice, the observation ages T_1 and T_2 may not be sufficiently close to the tabulated reference age intervals. It is not yet clear what the best way is to approximate r in such cases.

This section concentrated on conditional velocity reference, where conditioning is applied to one or more previous observations. Conditional distance references are also possible, and that type of references is in fact older. Examples include references that condition on parental height[89] and sibling births weight[90].

13.6 Model evaluation

It is crucial that models used to estimate reference intervals fit the data extremely well.[91] Yet, relatively little research has been done on methods for evaluating and improving the fit of the model. References can be fitted in a variety of ways, but all approaches need to specify in some way the amount of smoothness that provides a reasonable trade-off between parsimony of the curves and the fidelity to the data. This section discusses various ways to gauge this trade-off. Let us distinguish among the following approaches:[35]

Visual inspection of the shape of the reference curves. Experienced researchers may recognize the appropriateness of a given set of reference curves based on subtle features in the shape, like a 'pubertal belly' in cross-sectional data. In general, substantial exposure to reference curves is needed to develop the necessary skills.

Centiles plotted onto the individual data points. This type of plot is useful for inspecting outliers and for detecting gaps in the data and gross errors in the model, but its resolution is too limited to be helpful in choosing among different models. Attained size can be visualized in both the original and in the SD scale, but the latter possibility offers a higher resolution.

Empirical and fitted centiles plotted on top of each other. This is an old and quite accurate technique in which the observations are divided into age groups. Empirical centiles are computed for each group, and these are plotted together with the fitted curves. If everything is right, the fitted curves should be close to the point estimates (i.e. within sampling error). Various choices are possible for the vertical scale (raw, standardized for mean and/or standard deviation). A disadvantage of the raw data plot is that if the standard deviation changes with age, the same distance means different things at different ages. Van Wieringen[5] pioneered a standardized graph under the heading of 'graphical graduation', which plots empirical and fitted centiles in deviations from the median on the original scale.

Observed and expected counts. Healy *et al*.[48] suggested comparing observed and expected frequencies of observations within defined centile and age groups, a method now known as the grid test.[91] Suppose that observations are grouped into centile groups divided by the 5^{th}, 50^{th} and 95^{th} centile curves and grouped according to age. If the model fits well, we would expect to observe about 5, 45, 45 and 5 percent of the cases in each group. The difference can be summarized by a χ^2 statistic, which can be tested for statistical significance. Although intuitively appealing, the grid test lacks statistical power because of the severe data reduction steps.[91] Metcalfe[92] defended the grid test by arguing that the grid test may detect aberrant patterns near the extremes that more powerful statistical tests may fail to pick up.

Statistical tests applied to the distribution of the SD score. In all models that assume a normal distribution, the SD score should be normally distributed at all ages. A mix of clearly non-normal age-conditional SD scores can make up a marginal distribution that is close to normality. Therefore, global tests that do not account for the age-conditional nature of the SD score are uninformative and should not be used.[35] Royston and Wright[91] developed a series of age-conditional *Q*-tests, which summarise how much the SD score deviates from the normal in the mean, spread, skewness and kurtosis. Likelihood Ratio tests are also used to fit models, though it is not entirely clear what should be taken as the appropriate reference distribution.[67] Statistical tests are very useful in signaling that there is a problem, but may be less able to tell at what age the problem occurs.

Quantile-Quantile plot (Q-Q plot) of the SD scores. Q-Q plots[93] can be applied if the measurements are supposed to follow a known distribution. The display plots the quantiles of the theoretical distribution (on the horizontal axis) against those of the empirical distribution (on the vertical axis). The Q-Q plot for normal data, also known as the normal probability plot, is best known, but it can be adapted to other distributions. The plot yields insight into structural characteristics (e.g. skewness,

kurtosis) of empirical deviations from the assumed distribution. A detrended Q-Q plot is obtained if each empirical quantile is subtracted from its corresponding unit normal quantile. The detrended plot is sensitive to subtle deviations.[94] Global Q-Q plots across the entire age range are not informative, and should thus be split according to age.

Worm plot. The worm plot[35] consists of a collection of detrended Q-Q plots, each of which applies to a successive age group. The vertical axis of the worm plot portrays, for each observation, the difference between its location in the theoretical and empirical distributions. The data points in each plot form a worm-like string. The shape of the worm indicates how the data differ from the assumed underlying distribution, and when taken together, suggests useful modifications to the model. A flat worm indicates that the data follow the assumed distribution in that age group. So the aim of the model fitting process is to 'tame the worms'. Figure 13.3 is the end result of fitting height distribution of boys. All worms are reasonably flat and are within chance variation, so the model fits quite well.

-- INSERT FIGURE 13.3 ABOUT HERE --

Pan and Cole compared the behavior of the worm plot, the *Q* test and the Likelihood Ratio test by fitting LMS models to data from the Third Dutch Growth Study[33]. They concluded that all are useful tools for testing goodness of fit, and suggested in what ways they complement each other.

Outliers also influence the fit. Tango[95] proposed a method for identifying outliers. Kapitula and Bedrick[96] proposed a diagnostic for measuring the influence of individual cases on the estimated centile curves within the context of FPET model[61].

13.7 Detection of growth disorders

The growth diagram is widely used in paediatric practice, but not much is known about its performance in detecting growth disorders.[97] A growth diagram defines the specificity of a single height measurement, but its sensitivity is unknown for even the most frequent diseases. This is unfortunate because it precludes an informed discussion about referral criteria. Referral criteria have been develop and evaluated[98,99,100,101], but this has not prevented the appearance of widely different guidelines. For example, the recent UK guideline[102] is based on just one universal height measurement at age five, whereas the Dutch consensus guidelines[103] consist of multiple referral criteria covering infancy, childhood and adolescence. All in all, current practice differs among practitioners, and practices are not founded on evidence.

This issue was recently addressed by a new methodology for estimating sensitivity, specificity and median referral age for referral rules.[104] Turner Syndrome was taken as a target disease, Three basic rules were investigated: 1) height SDS below a given cut point, 2) height SDS below a cut point but corrected for parental height, and 3) height SDS velocity below a given cut point. Using two longitudinal data sets, one for Turner Syndrome and one reference, it was found that the parentally adjusted rule is superior to both others, with sensitivity near 70% and a specificity of 99.4%. Turner Syndrome is, in some sense, an 'ideal condition' to show the methodology[105], but the principle can be adapted to other disorders covered by the chapters of this book.

Implementing new screening protocols using growth references requires careful consideration of the current setting. An example is the new Dutch protocol for detecting overweight and obesity[106]. Both overweight and obesity are defined in terms of BMI, yet existing practice in The Netherlands is based on the WfH chart. It was considered undesirable to replace the WfH chart by the BMI diagram, as that would create a new obligation to calculate BMI for every child. The solution to this dilemma was to continue using WfH, and calculate BMI only for children that have a WfH SDS over +1 SD. For this reason, the 'normal range' on the Dutch WfH chart was chosen to be between -1 and +1 SD (c.f. Figure 13.1). Figure 13.4 is a flow chart of the new protocol.

-- INSERT FIGURE 13.4 ABOUT HERE --

13.8 Individualizing references

Growth references typically take the form of age-conditional reference ranges by sex. These references portray the growth expected over all members of the reference population. Increased heterogeneity among these members will lead to more variability in the references, and thus to a lower level of precision for the growth expected for an individual member.

It is often useful to eliminate unwanted variation from the population standards. Factors that can account for considerable variability between members are: individual growth history, parental height, premature birth, region, ethnicity, age of pubertal onset, being part of a twin, breast feeding, presence of a disease, and the level of education. One can condition on one or more factors. In practice, this can involve selecting subgroups from the reference population, and adapting the reference data to reflect the characteristics of this subgroup. The results may be a narrower, more precise, reference that is tailored to the characteristics of the individual.

Conditioning is not always good, and should only be done if the subgroup reference fits the purpose of the applied context. An example where this may not be true is parental weight. Creating a reference from the subgroup of children with thick parents is likely to shift the weight distribution upwards. Using this individualised reference in a normative way may miss children that actually are at risk, leading to a loss of potential health gains.

13.9 Miscellaneous topics

13.9.1 Synthetic growth charts

Growth references require large and costly studies. When resources are limited, synthetic growth reference charts[107] can help in providing quick-and-dirty growth charts. For boys, a synthetic height reference can be constructed using samples from the population at the ages of 0, 2, 6, 14 and 18 years. The remaining ages of the reference can be interpolated according to the known shape of the distance diagram. Synthetic growth charts for body weight have also been developed.[108] Little is known about the accuracy of these synthetic charts, but the idea is interesting and may stimulate the realisation of gross savings in periodical updates of reference values.

13.9.2 Chart design issues

Traditionally, the reference diagram plots a set of percentiles against age, usually the P3, P10, P25, P50, P75, P90 and P97. The extreme centile lines are sometimes combined with P5 and P95 curves.[109] Though percentiles are easy to understand, a difference on the percentile scale is meaningless. The clinical implication of 'loosing 10 percentile points' can be very different, depending on the location on the scale. If a child migrates from P50 to P40, there is probably little reason for concern. On the other hand, migrating from P11 to P1 can have profound clinical significance. The SD score does not have this problem, can be calculated outside the extreme centile curves, and be visually interpolated from the diagram. The British[110] reference uses a spacing of 2/3 SD, whereas the Dutch diagrams apply an intercurve spacing of 1 and ½ SD. A drawback is that SD-based charts are less intuitive to the uninitiated.

The traditional growth chart design fails to use 90% of the rectangular graphic area. There are few toddlers with a height of 180cm or adults with a height of 60cm. It is therefore more efficient and more informative to plot the SDS against age. Sorva *et al*.[111] developed a format in which normal

growth is shown as a horizontal line and any deviation from this indicates abnormal change in growth. Thus far, such charts seem to have been used only in Finland. Figure 13.5 plots a chart for Dutch boys 0-24 months. The curved lines can be interpreted as height contours, so the name 'meridian chart' seems appropriate for this type of display.

-- INSERT FIGURE 13.5 ABOUT HERE --

The effective display of conditional references is a major challenge. Cole[112,113] developed ways to enrich the distance diagram with velocity information by means of overlays. Given the importance of monitoring growth, more work along these lines is needed. Chart design issues are heavily tied to growth diagrams on paper. The use of computerised systems for presenting and using growth diagrams will have a major impact of the type of design decisions that need to be made.

13.9.3 Non-continuous outcomes

Relatively little work has been done on the construction of age-related references for non-continuous measures. Non-continuous measures that are of interest include: pubertal stages, indicators of motor and mental development, visual acuity, developmental milestones, clinical grades, and so on.

-- INSERT FIGURE 13.6 ABOUT HERE --

Figure 13.6 shows how the probability of attaining pubertal stages G2-G5 of Dutch boys stages varies with ages.[114] These probabilities were modeled separately for each stage as a continuous, nonlinear function of age by a generalized additive logistic model[115]. The probability of passing P10, P50 and P90 can be read off from the curve (except for P10 of G2), and used in the reference diagram as in Figure 13.1.

- - INSERT FIGURE 13.7 ABOUT HERE --

Figure 13.7 is a collection of reference diagrams of menarche probability of Dutch girls, where probability is modeled by an additive logistic model with two main factors. The first is age, the second is either weight, height or BMI, in both raw and standardized forms. A vertical course of the lines means that menarche probability is indifferent to the factor. So, the top-left plot indicates that higher weight is associated with an increase of menarche probability, but that weights beyond 60 kg do not further increase the menarche probability, irrespective of age.

Wade *et al*.[116,117] estimated references for ordinal outcomes using a fully parametric proportional odds model with asymmetric logistic models per cumulative category. More work on non-continuous outcomes would be useful.

List of key points

1. Growth references describe anthropometric variation within a group of individuals. Standards delineate variation that is considered normal, optimal or healthy.
2. Growth references are valuable for answering a wide variety of questions on both the individual and population level.
3. References can be classified as distance, velocity and conditional. Distance references are informative about attained size only, whereas velocity and conditional references describe the process of growth over time.
4. The best current methods for fitting age-conditional references treat age as a continuous and assume a distribution per age.
5. Good tools like *Q*-tests and worm plots exist to evaluate the quality of a fitted growth model.
6. Detecting growth disorders by a growth diagram requires information about its sensitivity, specificity and referral time distribution.
7. More work is needed on diagnostic performance, on methods for individualizing references, on synthetic growth charts, on computerized systems, and on non-continuous outcomes.

LEGEND

© Seminal paper

® Review paper

REFERENCES

1 World Health Organisation. *Report of WHO Expert Committee: Physical Status - The Use and Interpretation of Anthropometry*. WHO Technical report no. 854. Geneva, Switzerland: World Health Organisation, 1995.

2 Fredriks AM, van Buuren S, Burgmeijer RJF, *et al*. Continuing positive secular growth change in The Netherlans 1955-1997, *Pediatric Research* 2000; **47**: 316-323.

3 Fredriks AM, van Buuren S, Burgmeijer RJF, *et al.Groeidiagrammen (derde druk)* [Growth diagrams (3rd edition)]. Houten: Bohn, Stafleu, Van Loghum, 2004.

4 © Tanner JM. Use and abuse of growth standards. In: Falkner F, Tanner JM (eds.) *Human growth: A comprehensive treatise. Second edition, Vol. 3*. New York: Plenum Press, 1986: 95-109.

5 © Van Wieringen JC. *Seculaire groeiverschuiving [Secular growth shift]*. Leiden: Netherlands Institute for Preventive Medicine TNO, 1972.

6 World Health Organisation. An evaluation of infant growth: the use and interpretation of anthropometry in infants. *Bulletin of the World Health Organisation* 1995; **73**: 165-74.

7 World Health Organization Working Group on Infant Growth. *An evaluation of infant growth*. Report WHO/NUT/94.8. Geneva, Switzerland: World Health Organization, 1994.

8 Cole TJ, Paul AA, Whitehead RG. Weight reference charts for British long-term breastfed infants. *Acta Paediatrica* 2002; **91**: 1296-300.

9 De Onis M, Onyango AW. The Centers for Disease Control and Prevention 2000 growth charts and the growth of breastfed infants. *Acta Paediatrica* 2003; **92**: 413-9.

10 Cole TJ, Roede MJ. Centiles of body mass index for Dutch children aged 0-20 years in 1980 - A baseline to assess recent trends in obesity. *Annals of Human Biology* 1999; **26**: 303-8.

11 Fredriks AM, van Buuren S, Wit JM, Verloove-Vanhorick SP. Body index measurements in 1996-7 compared with 1980. *Archives of Diseases in Childhood* 2000; **82**: 107-12.

12 Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *British Medical Journal* 2000; **320**: 1240-3.

13 van Buuren S. Afkapwaarden van de 'body mass index' (BMI) voor ondergewicht van Nederlandse kinderen [Body-mass index cut-off values for underweight in Dutch children]. *Nederlands Tijdschrift voor de Geneeskunde* 2004; **148**: 1967-72.

14 Healy MJR. Statistics of growth standards. In: Falkner F, Tanner JM (eds.) *Human growth: A comprehensive treatise. Second edition, Vol. 3*. New York: Plenum Press, 1986: 47-58.

[15] ® Frongillo, EA. Univariate and bivariate growth references. In: Hauspie RC, Cameron N, Molinari L (eds.) *Methods in human growth research*. Cambridge: Cambridge University Press, 2004: 261-86.

[16] Haschke F, van't Hof MA, and the Euro-Growth Study Group. Euro-Growth References for length, weight, and body circumferences. *Journal of Pediatric Gastroenterology and Nutrition* 2000; **31**: S14-S38.

[17] Van't Hof MA, Haschke F, Darvay S and the Euro-Growth Study Group. Euro-Growth References on increments in length, weight, head- and arm circumference during the first three years of life. *Journal of Pediatric Gastroenterology and Nutrition* 2000; **31**: S39-S47.

[18] De Onis M, Victora CG, Garza C, *et al*. A new international growth reference for young children. In Dasgupta P, Hauspie RC (eds.) *Perspectives in human growth, development and maturation*, Dordrecht, The Netherlands: Kluwer Academic Publishers, 2001: 45-53.

[19] De Onis M, Garza C, Victora CG, *et al*. The WHO multicentre growth reference study: Planning, study design, and methodology. *Food and Nutrition Bulletin* 2004; **25**: S15-S26.

[20] © Waterlow JC, Buzina R, Keller W, *et al.* The presentation and use of height and weight data for comparing nutritional status of groups of children under the age of 10 years. *Bulletin of the World Health Organization* 1977; **55**: 489-98.

[21] Kendall M, Stuart A, Ord JK. *Kendall's advanced theory of statistics. Vol. I: Distribution theory*. London: Charles Griffin & Company Ltd, 1980.

[22] Goldstein H. Sampling for growth studies. In: Falkner F, Tanner JM (eds.) *Human growth: A comprehensive treatise. Second edition, Vol. 3*. New York: Plenum Press, 1986: 59-78.

[23] Cole TJ. Growth and development. In: Armitage P, Colton T (eds.) *Encyclopedia of Biostatistics*. New York: Wiley, 1998: 1790-7.

[24] © Cole TJ. Growth charts for both cross-sectional and longitudinal data. *Statistics in Medicine* 1994; **13**: 2477-92.

[25] Preece MA. Standardization of growth. *Acta Paediatrica Scandinavia* 1989; 349: 57-64.

[26] Cole TJ. Fitting smoothed centile curves to reference data, *Journal of the Royal Statistical Society Series A*, 1988; **151**: 385-418.

[27] ® Wright EM, Royston P. A comparison of statistical methods for age-related reference intervals, *Journal of the Royal Statistical Society Series A* 1997; **160**, 47-69.

[28] ® Borghi E, de Onis M, Garza C, *et al*. Construction of the WHO/MGRS growth references: Selection of methods for attained growth. Submitted for publication.

[29] Hamill PVV, Drizd TA, Johnson CL, *et al*. *NCHS growth curves for children birth – 18 years*. Washington DC: National Center for Health Statistics. Vital and Health Statistics Series 11, No 165, 1777.

[30] Gasser T, Köhler W, Müller H-G, *et al*. Velocity and acceleration of height growth using kernel estimation. *Annals of Human Biology* 1984; **11**: 397-411.

[31] Guo S, Roche AF, Baumgartner RN, *et al*. Kernel regression for smoothing percentile curves: reference data for calf and subscapular skinfold thicknesses in Mexican Americans. *American Journal of Clinical Nutrition* 1990; **51**: 908S-916S.

[32] Van Wieringen JC, Wafelbakker F, Verbrugge HP, de Haas JH. *Growth diagrams 1965 Netherlands. Second national survey on 0-24-year-olds*. Groningen: Wolters-Noordhoff, 1971.

[33] Roede MJ, van Wieringen JC. Growth diagrams 1980: Netherlands third nation-wide survey, *Tijdschrift voor Sociale Gezondheidszorg* 1985, **63**, supplement.

[34] Healy MJR. The effect of age-grouping on the distribution of a measurement affected by growth. *American Journal of Physical Anthropology* 1962; **20**: 49-50.

[35] van Buuren S, Fredriks AM. Worm plot: A simple diagnostic device for modeling growth reference curves. *Statistics in Medicine* 2001; **20**: 1259-77.

[36] Jones MC. Discussion of "Fitting smoothed centile curves to reference data" by TJ Cole. *Journal of the Royal Statistical Society* 1988; **151**: 412-3.

[37] Tanner JM, Whitehouse RH, Takaishi M. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965; Part I. *Archives of Diseases in Childhood* 1966; **41**: 454-71.

[38] Van 't Hof MA, Wit JM, Roede MJ. A method to construct age references for skewed skinfold data, using Box-Cox transformations to normality. *Human Biology* 1985; **57**: 131-9.

[39] Niklasson A, Ericson A, Fryer JG, *et al*. An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977-1981). *Acta Paediatrica Scandinavia* 1991; **80**: 756-762.

[40] Chinn S. A new method for calculation of height centiles for preadolescent children. *Annals of Human Biology* 1992; **19**: 221-32.

[41] © Koenker RW, Bassett G Jr. Regression quantiles. *Econometrica* 1978; **46**: 33-50.

[42] He X. Quantile curves without crossing. *American Statistician* 1997; **51**: 186-92.

[43] Heagerty PJ, Pepe MS. Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Applied Statistics* 1999; **48**: 533-51.

[44] Ducharme GR, Gannoun A, Guertin MC, Jéquier JC. Reference values obtained by kernel-based estimation of quantile regressions. *Biometrics* 1995: **51**: 1105-16.

[45] Wellek S, Merz E. Age-related reference ranges for growth parameters. *Methods of Information in Medicine* 1995; **34**: 523-8.

[46] Rossiter JE. Calculating centile curves using kernel density estimation methods with application to infant kidney lengths. *Statistics in Medicine* 1991; **10**: 1693-701.

[47] Gannoun A, Girard S, Guinot C, Saracco J. Reference curves based on non-parametric quantile regression. *Statistics in Medicine* 2002; **21**: 3119-35.

[48] Healy MJR, Rasbash J, Yang M. Distribution-free estimation of age-related centiles. *Annals of Human Biology* 1988; **15**: 17-22.

[49] Pan H, Goldstein H, Yang Q. Non-parametric estimation of age-related centiles over wide age ranges. *Annals of Human Biology* 1990; **17**: 475-81.

[50] Goldstein H, Pan H. Percentile smoothing using piecewise polynomials, with covariates. *Biometrics* 1992; **48**: 1057-68.

[51] Aitkin M. Modelling variance heterogeneity in normal regression using GLIM. *Applied Statistics* 1987; **36**: 332-9.

[52] Altman DG. Construction of age-related reference centiles using absolute residuals. *Statistics in Medicine* 1993; **12**: 917-24.

[53] © Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalised likelihood. *Statistics in Medicine* 1992; **11**, 1305-19.

[54] © Box GEP, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society Series B* 1964; **26**: 211-52.

[55] Wade AM, Ades AE. Age-related reference ranges – significance tests for models and confidence intervals for centiles. *Statistics in Medicine* 1994; **13**: 2359-67.

[56] Thompson ML, Theron GB. Maximum likelihood estimation of reference centiles. *Statistics in Medicine* 1990; **9**: 539-48.

[57] Johnson NL. Systems of frequency curves generated by methods of translation. *Biometrika* 1949; **36**:149-76.

[58] Royston P. Constructing time-specific reference ranges. *Statistics in Medicine* 1991; **10**: 675-90.

[59] Tango T. Estimation of age-specific reference ranges via smoother AVAS. *Statistics in Medicine* 1998; **17**: 1231-43.

[60] Wright EM, Royston P. Simplified estimation of age-specific reference intervals for skewed data. *Statistics in Medicine* 1997; **16**: 2785-803.

[61] Royston P, Wright EM. A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. *Journal of the Royal Statistical Society Series A* 1998; **161**: 79-101.

[62] Yee TW. Quantile regression via vector generalized additive models. *Statistics in Medicine* 2004; **34**: 2295-315.

[63] Yeo I-K, Johnson RA. A new family of power transformations to improve normality and symmetry. *Biometrika* 2000; **87**: 954-9.

[64] Sorribas A, March J, Voit EO. Estimating age-related trends in cross-sectional studies using S-distributions. *Statistics in Medicine* 2000; **19**: 697-713.

[65] Rigby RA, Stasinopoulos DM. Box-Cox *t* distribution for modelling skew and leptokurtotic data with an application to centile estimation. 2004. Submitted for publication.

[66] Rigby RA, Stasinopoulos DM. Smooth centiles curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. *Statistics in Medicine* 2004, **23**: 3053-76.

[67] ® Pan H, Cole TJ. A comparison of goodness of fit tests for age-related reference ranges. *Statistics in Medicine* 2004; **23**: 1749-65.

[68] Moussa MAA. Estimation of age-specific reference intervals from skewed data. *Methods of Information in Medicine* 2002; **42**: 147-53.

[69] Bonellie SR, Raab GM. A comparison of different approaches for fitting centile curves to birthweight data. *Statistics in Medicine* 1996;**15**:2657-67.

[70] Pere A. Comparison of two methods for transforming height and weight to normality. *Annals of Human Biology* 2000; **27**: 35-45.

[71] Van Buuren S, Fredriks AM. Methoden voor het objectiveren van groeiafbuiging (in Dutch) [Objective methods for growth deflection]. In: Wit JM (ed.), *De Vierde Landelijke Groeistudie; Presentatie nieuwe groeidiagrammen*. Leiden: Boerhaave Commissie, 1998: 91-101.

[72] Van Buuren S, Fredriks AM, Verkerk PH (1999). Consensus 'Diagnostiek kleine lichaamslengte bij kinderen'. *Nederlands Tijdschrift voor de Geneeskunde*; **143**: 1585-6.

[73] Prahl-Andersen B, Kowalksi CJ, Heijendael P. *A mixed-longitudinal interdisciplinary study of growth and development: Nijmegen Growth Study*. San Francisco: Academic Press, 1979.

[74] Cameron N. Conditional standards for growth in height of British children from 5.0 to 15.99 years of age. *Annals of Human Biology* 1980; **7**: 331-337.

[75] Karlberg J. *Modelling of human growth*. Dissertation, Gotenburg, 1987.

[76] Cole TJ. Conditional reference charts to assess weight gain in British infants. *Archives of Diseases in Childhood* 1995; **73**: 8-16.

[77] © Galton F. Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute* 1886; **15**: 246-263.

[78] © Healy MJR. Notes of the statistics of growth standards. *Annals of Human Biology* 1974; **1**: 41-46.

[79] Berkey CS, Reed RB, Valadian I. Longitudinal growth standards for pre-school children. *Annals of Human Biology* 1983; **10**: 57-67.

[80] Royston P. Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal measurements. *Statistics in Medicine* 1995; **14**: 1417-36.

[81] Thompson ML, Fatti LP. Construction of multivariate centile charts for longitudinal measurements. *Statistics in Medicine* 1997; **16**: 333-45.

[82] Wright CM, Matthews JNS, Waterston A, Aynslay-Green A. What is the normal rate of weight gain in infancy? *Acta Paediatrica* 1994; **83**: 351-6.

[83] Gasser T, Molinari L, Roos M. Methodology for the establishment of growth standards. *Hormone Research* 1996; **45**(Suppl. 2), 2-7.

[84] Pan H, Goldstein H. Multi-level models for longitudinal growth norms. *Statistics in Medicine* 1997; **16**: 2665-78.

[85] Pan H, Goldstein H. Multi-level repeated measures growth modelling using extended spline functions. *Statistics in Medicine* 1998; **17**: 2755-70.

[86] Fatti LP, Senaoana EM, Thompson ML. Bayesian updating in reference centiles charts. *Journal of the Royal Statistical Society Series A* 1998; **161**: 103-15.

[87] Wade AM, Ades AE. Incorporating correlations between measurements into the estimation of age-related reference ranges. *Statistics in Medicine* 1998; **17**: 1989-2002.

[88] Reinhard I, Wellek S. Age-related reference regions for longitudinal measurements of growth characteristics. *Methods of Information in Medicine* 2001; **40**: 132-6.

[89] Tanner JM, Goldstein H, Whitehouse RH. Standards for children's height at ages 2 to 9 years, allowing for heights of parents. *Archives of Diseases in Childhood* 1970; **45**: 755-62.

[90] Tanner JM, Lejarrage H, Healy MJR. Within-family standards for birthweight, a revision. *Lancet* 1972; **ii**: 1314-5.

[91] Royston P, Wright EM. Goodness-of-fit statistics for age-specific reference intervals. *Statistics in Medicine* 2000; **19**: 2943-62.

[92] Metcalfe C. Letter to the editor. *Statistics in Medicine* 2002; **21**: 3749-50.

[93] Hoaglin DC. Using quantiles to study shape. In: Hoaglin DC, Mosteller F, Tukey JW (eds.), *Exploring data tables, trends, and shapes*. Wiley: New York, 1985: 417-59.

[94] Friendly M. *SAS System for statistical graphics, First Edition*. Cary, NC: SAS Institute, 1991.

[95] Tango T. Estimation of normal ranges in clinical laboratory data. *Statistics in Medicine* 1986; **5**: 335-346.

[96] Kapitula LR, Bedrick EJ. Diagnostics for the exponential normal growth curve model. *Statistics in Medicine* 2005, to appear.

[97] Garner P, Panpanich R, Logan S. Is routine growth monitoring effective? A systematic review of trials. *Archives of Diseases in Childhood* 2000; **82**: 197-201.

[98] Hindmarsh PC. Monitoring children's growth: Abnormal growth should also be defined by the crossing of height centiles (letter). *British Medical Journal* 1996; **312**:122.

[99] Mulligan J, Voss LD, McCaughey ES, *et al*. Growth monitoring: testing the new guidelines. *Archives of Diseases in Childhood* 1998; **79**: 318-22.

[100] Voss LD. Changing practice in growth monitoring. *British Medical Journal* 1999; **318**:344-5.

[101] Hall DMB. Growth monitoring. *Archives of Diseases in Childhood* 2000; **82**:10-5.

[102] Hall DMD, Elliman D. *Health for all children, 4th Ed*. Oxford University Press, 2003.

[103] De Muinck Keizer-Schrama SMPF. Consensus 'Diagnostiek kleine lichaamslengte bij kinderen' [Dutch consensus guidelines for short stature]. *Nederlands Tijdschrift voor de Geneeskunde* 1998; **142**: 2519-25.

[104] van Buuren S, van Dommelen P, Zandwijken GRJ, *et al.* Towards Evidence Based Referral Criteria for Growth Monitoring. *Archives of Diseases in Childhood* 2004: **89**: 336-41.

[105] Hindmarsh PC, Cole TJ. Commentary: Height monitoring as a diagnostic test. *Archives of Diseases in Childhood*; **89**: 296-7.

[106] Bulk-Bunschoten AMW, Renders CM, van Leerdam FJM, HiraSing RA. *Signalering Overgewicht in de Jeugdgezondheidszorg*. Amsterdam: VuMC, 2004.

[107] Hermanussen M, Burmeister J. Synthetic growth reference charts. *Acta Paediatrica* 1999; **88**:809-14.

[108] Hermanussen M, Meigen C. Synthetic standards for body weight. *Homo* 2003; **54**: 142-56.

[109] Kuczmarski RJ, Ogden CK, Gummer-Strawn LM, *et al.*CDC Growth Charts: United States. Hyattsville, MD: National Center for Health Statistics. Available from hhtp://www.cdc.gov/growthcharts/.

[110] Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine* 1998; **17**: 407-29.

[111] Sorva R, Perheentupa J, Tolppanen EM. A novel format for a growth chart. Acta Paediatrica Scandinavica 1984;**73**:527-9.

[112] Cole TJ. 3-in-1 weight monitoring chart. *Lancet* 1997; **349**: 1020-30. (???)

[113] Cole TJ. Presenting information on growth distance and conditional velocity in one chart: practical issues of chart design. *Statistics in Medicine* 1998; **17**: 2697-707.

[114] Mul D, Fredriks AM, van Buuren S, *et al.* Pubertal development in The Netherlands 1965-1997. *Pediatric Research* 2000; **50**: 479-86.

[115] Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. London: Chapman and Hall, 1990.

[116] Wade AM, Ades AE, Salt AT, *et al*. Age-related standards for ordinal data: modelling the changes in visual acuity from 2 to 9 years of age. *Statistics in Medicine* 1995; **14**: 257-66

[117] Wade AM, Salt AT, Proffitt RV, *et al*. Likelihood-based modelling of age-related normal ranges for ordinal measurements: changes in visual acuity through early childhood. *Statistics in Medicine* 2004; **23**: 3623-40.

Table 13.1 Typical questions whose answers require growth references.

Individual	
Screening	Does this child have a problem with breast feeding?
	Is this child's diet inadequate?
	Does this child need supplementary food, or treatment for disease?
	Is this child overweight?
	Does this child need to be referred because of a growth problem?
Diagnosis	Does this child have organic diseases?
	Is there "failure to thrive" because of an underlying disease?
	What causes this child's overweight?
	Does an infection cause stunted growth?
Monitoring	What are the effects of increased physical exercise for this obese child?
	What are the effects of improvements in nutrition?
	What are the effects of improved health care on individual growth?
Prediction	How will this child grow in future?
	What is the risk that this child will develop a growth-related disease?
	What final height will this child attain?
Population	
Prevalence	How many people are overweight?
	What is the proportion of people in each height category?
	What is the optimal height of chairs in public transport?
Causes	What is the impact of infectious epidemic on growth?
	Can differences in growth be attributed to socio-economic differences?
Targeting	What parts of the population have the greatest need for interventions?
	How should resources be allocated?
Monitoring	At what year did the global obesity epidemic start?
	How did the height distribution evolve over the last century?
Evaluation	How many children were saved because of the nutritional aid program?
Prediction	How long will people be in the year 2020?
	What is the impact of overweight on the prevalence of future chronic diseases?

Table 13.2 95% confidence intervals for the correlation coefficient for sample sizes of $N = 100$ and $N = 200$.

Correlation	$N = 100$	$N = 200$
0.99	0.985 - 0.993	0.987 - 0.992
0.95	0.926 - 0.967	0.934 - 0.962
0.90	0.854 - 0.932	0.870 - 0.924
0.80	0.716 - 0.861	0.744 - 0.845
0.70	0.584 - 0.788	0.621 - 0.764

Table 13.3 Indices for individual attained size and growth, to be used in conjunction with age-related distance, velocity and conditional references.

Definition	Name	Advantages	Disadvantages
Distance			
X_1	Attained size at age T_1	Simple	Interpretation depends on age and sex
Z_1	Standard deviation score (SDS) at T_1	Comparable across age and sex, has standard normal distribution	Provides no information about growth over time
Velocity			
$V = (X_2 - X_1) / (T_2 - T_1)$	Velocity	Easy to calculate	Interpretation depends on age and sex
$W = (Z_2 - Z_1) / (T_2 - T_1)$	Velocity in SD scale	Adjusts velocity for age and sex	Distribution still depends on age
$Z_g = (Z_2 - Z_1) / \sqrt{(2-2r)}$	Standardised gain score (SDS gain)	Comparable across age and sex, has standard normal distribution for all T_1 and T_2	Does not account for regression to the mean, requires correlation between Z_1 and Z_2
Conditional			
$cZ_g = (Z_2 - rZ_1) / \sqrt{(1-r^2)}$	Conditional SDS gain	As Z_g, but accounts for regression to the mean	Requires correlation between Z_1 and Z_2

Figure Captions

13.1 Growth diagram of height-for-age, weight-for-height and pubertal development of Dutch boys 1-21 years.

13.2 Expected percentage in the reference population having an SDS velocity below -0.25 SDS/year. The measurement interval is one year.

13.3 Worm plot of the final LMS model (0/10/6r) fitted to height of Dutch boys 1-21 years.

13.4 Flow chart of the new Dutch protocol for identifying overweight and obesity.

13.5 Meridian chart of height of Dutch boys aged 0-24 months based on data from the Third Dutch Growth Study. Instruction: At the appropriate age, find the meridian (cm) line that equals the measurement, and mark the location. If necessary, interpolate between the two nearest meridian lines.

13.6 Empirical probabilities and fitted reference curve of genital stages G2-G5 for Dutch boys 8-21 years (Data: Fourth Dutch Growth Study).

13.7 Reference chart of menarche probability as a function of age and weight (kg and SDS), height (cm and SDS) or BMI (kg/m^2 and SDS) (Data: Fourth Dutch Growth Study).

GROEIDIAGRAM 1–21 JAAR | JONGENS

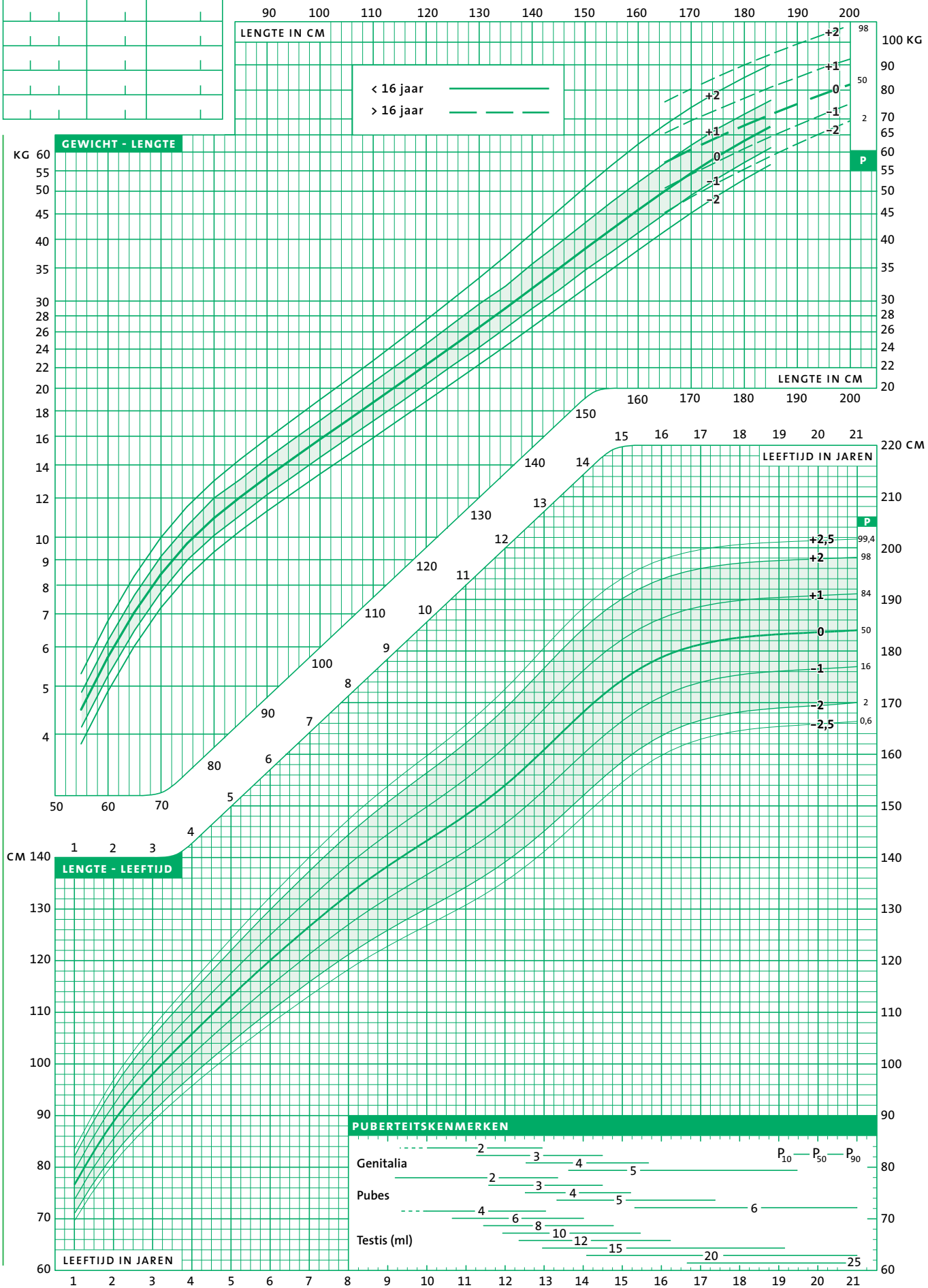
Datum	Gewicht	Lengte

Naam

Geboortedatum | Reg. nr

Vader (a/g) | cm | Moeder (a/g) | cm | TH | cm

♂



Ministerie van Volksgezondheid, Welzijn en Sport
Zorg Onderzoek Nederland ondersteunt
Pfizer
NUTRICIA

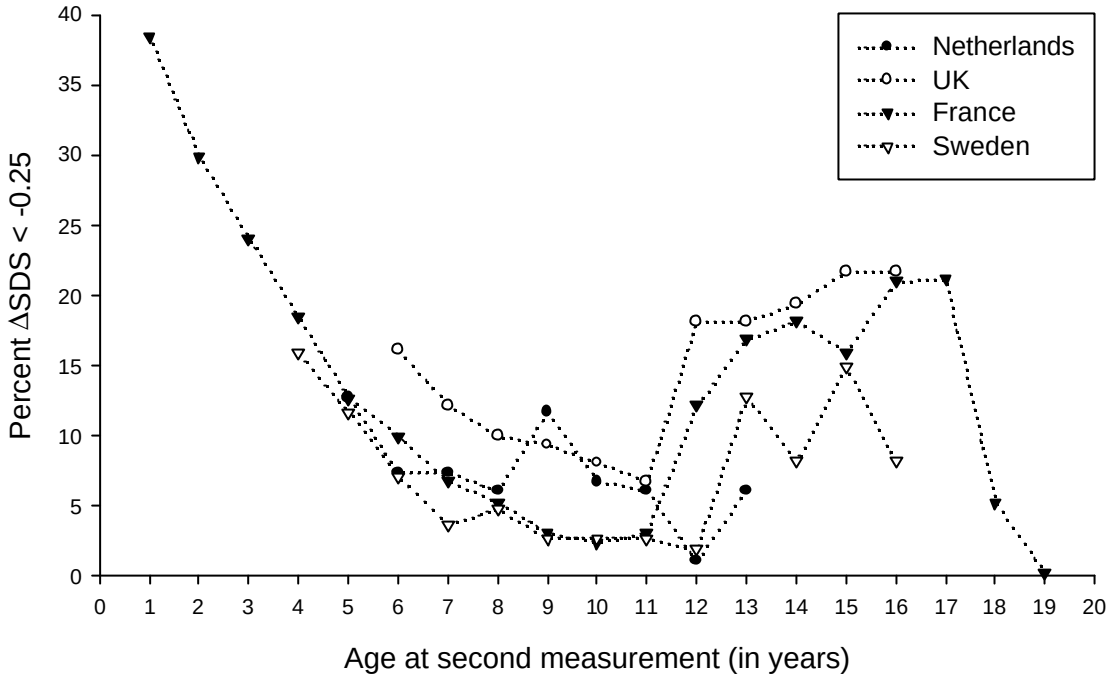

Netherlands
UK
France
Sweden
Percent ΔSDS < -0.25
Age at second measurement (in years)
0
5
10
15
20
25
30
35
40
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Figure 2

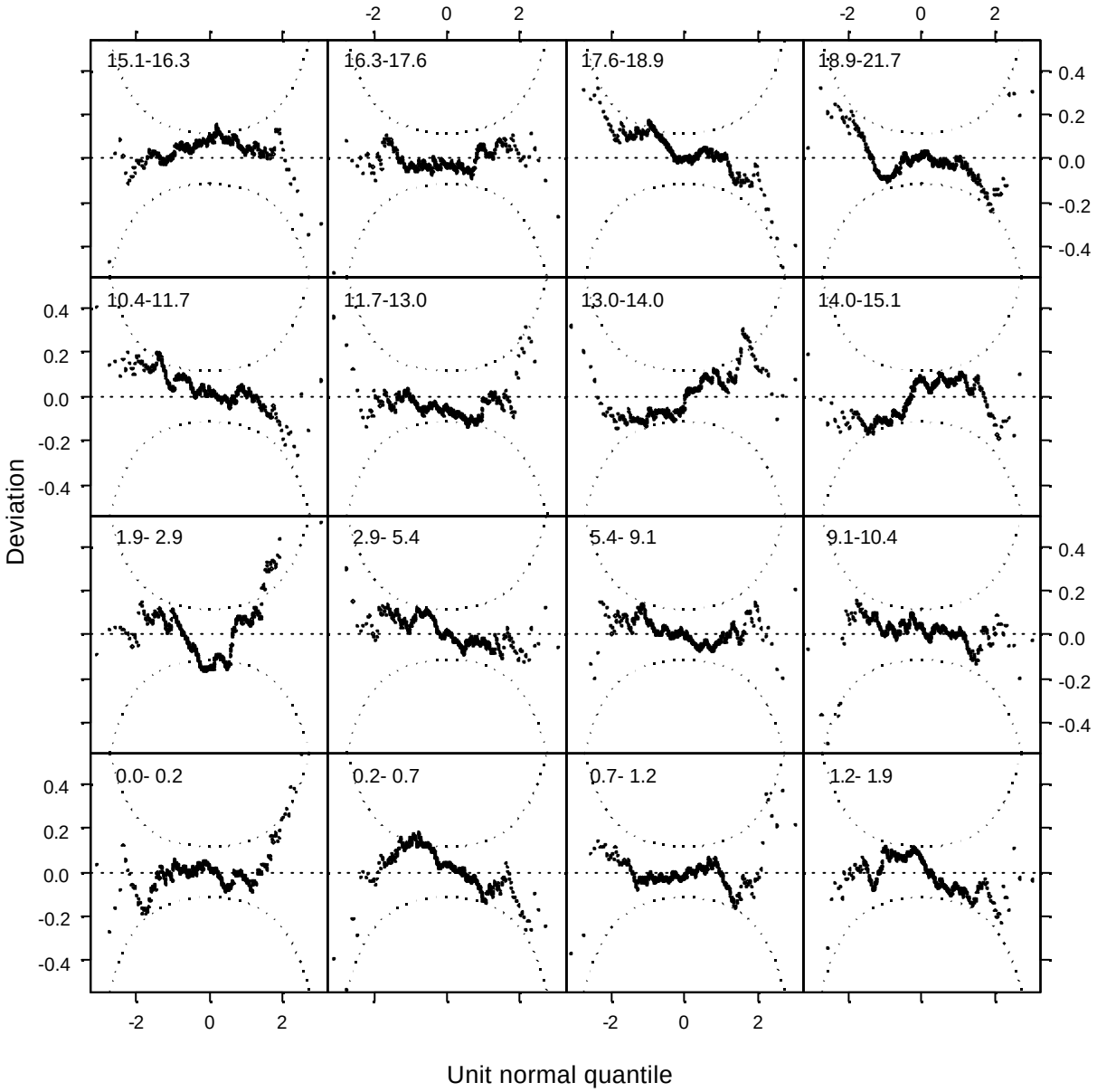
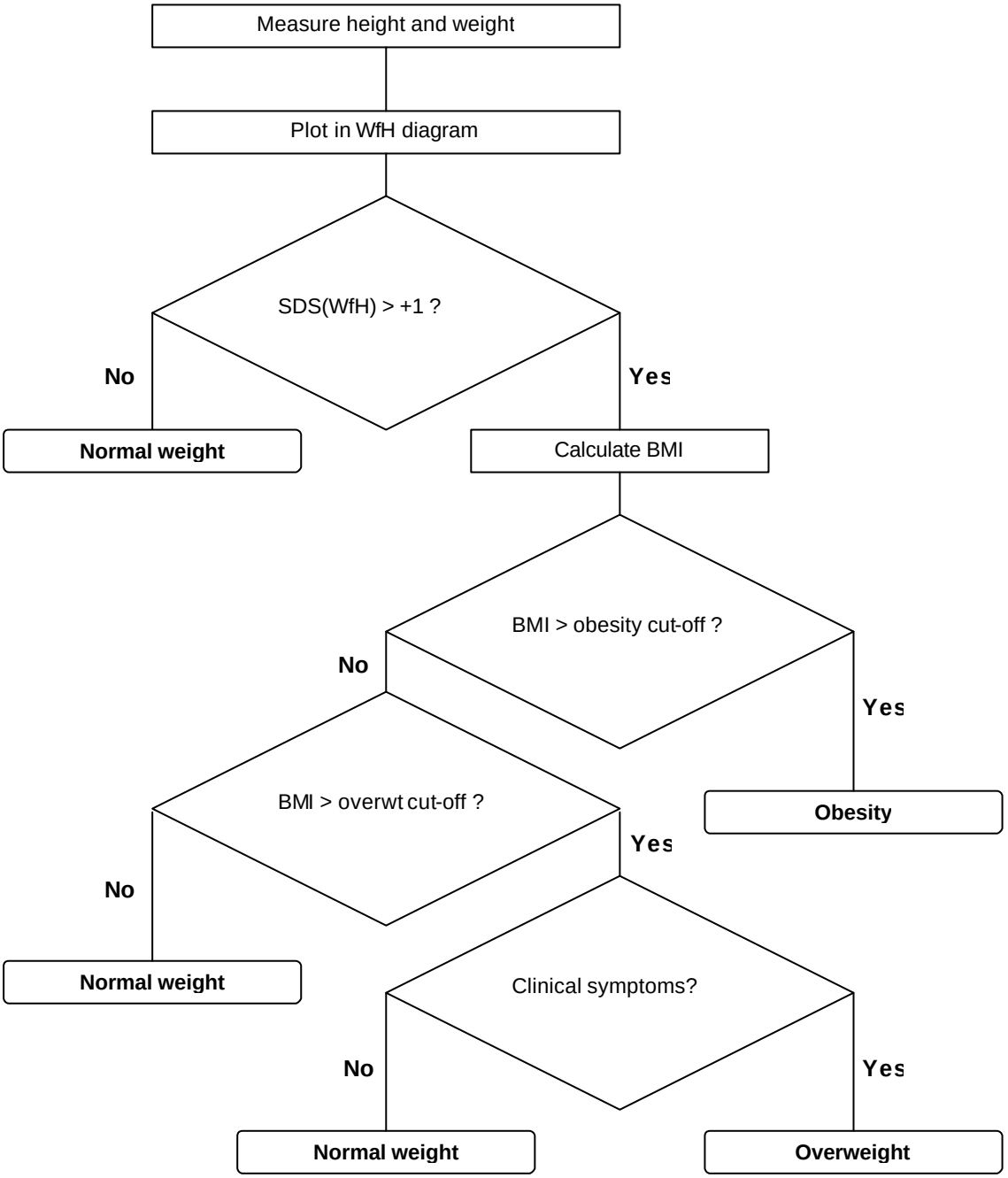

15.1-16.3
16.3-17.6
17.6-18.9
18.9-21.7
10.4-11.7
11.7-13.0
13.0-14.0
14.0-15.1
1.9- 2.9
2.9- 5.4
5.4- 9.1
9.1-10.4
0.0- 0.2
0.2- 0.7
0.7- 1.2
1.2- 1.9
Deviation
Unit normal quantile
-2
0
2
0.4
0.2
0.0
-0.2
-0.4

figure 4


Measure height and weight
Plot in WfH diagram
SDS(WfH) > +1 ?
No
Yes
Normal weight
Calculate BMI
BMI > obesity cut-off ?
No
Yes
Obesity
BMI > overwt cut-off ?
Yes
No
Normal weight
Clinical symptoms?
No
Yes
Normal weight
Overweight

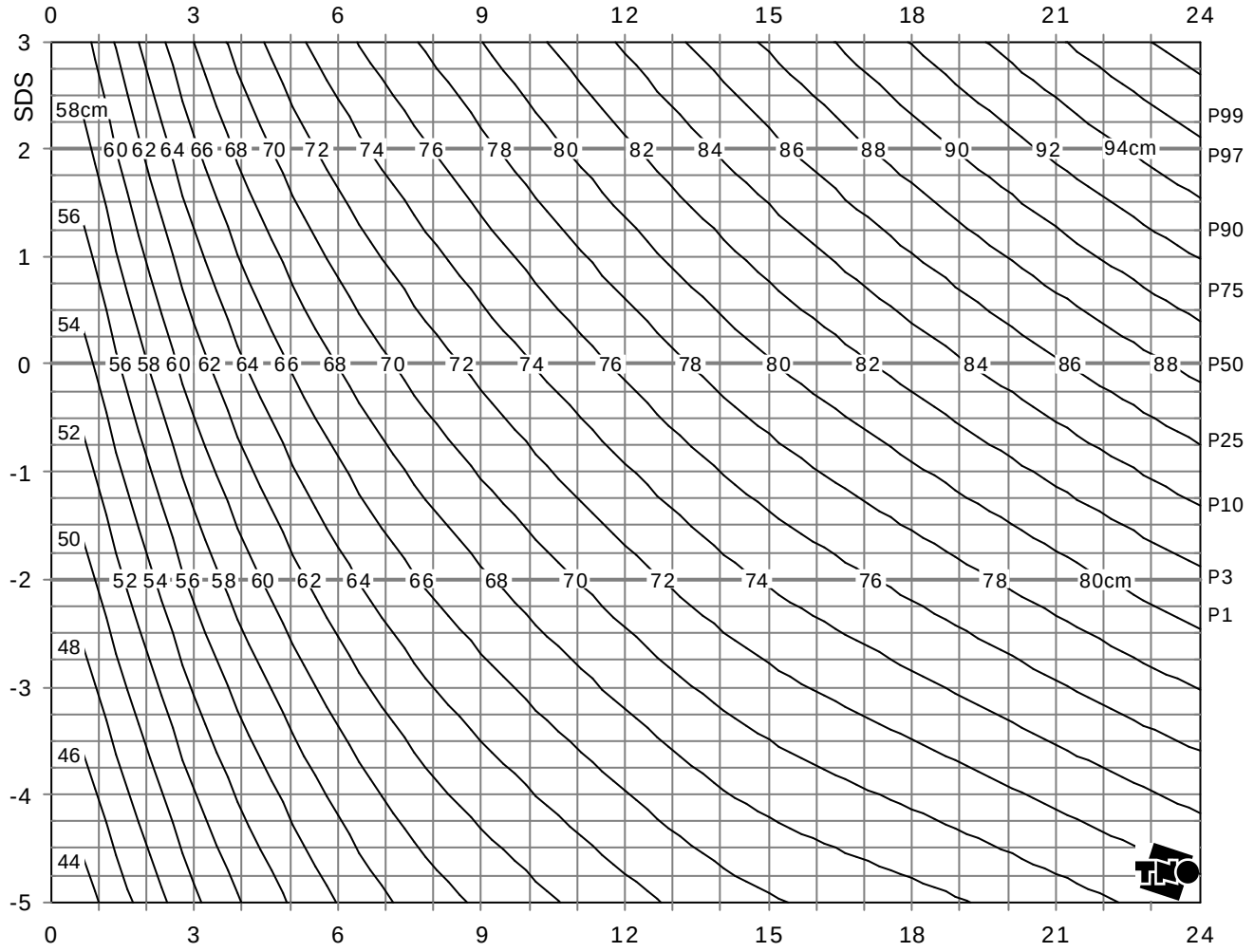
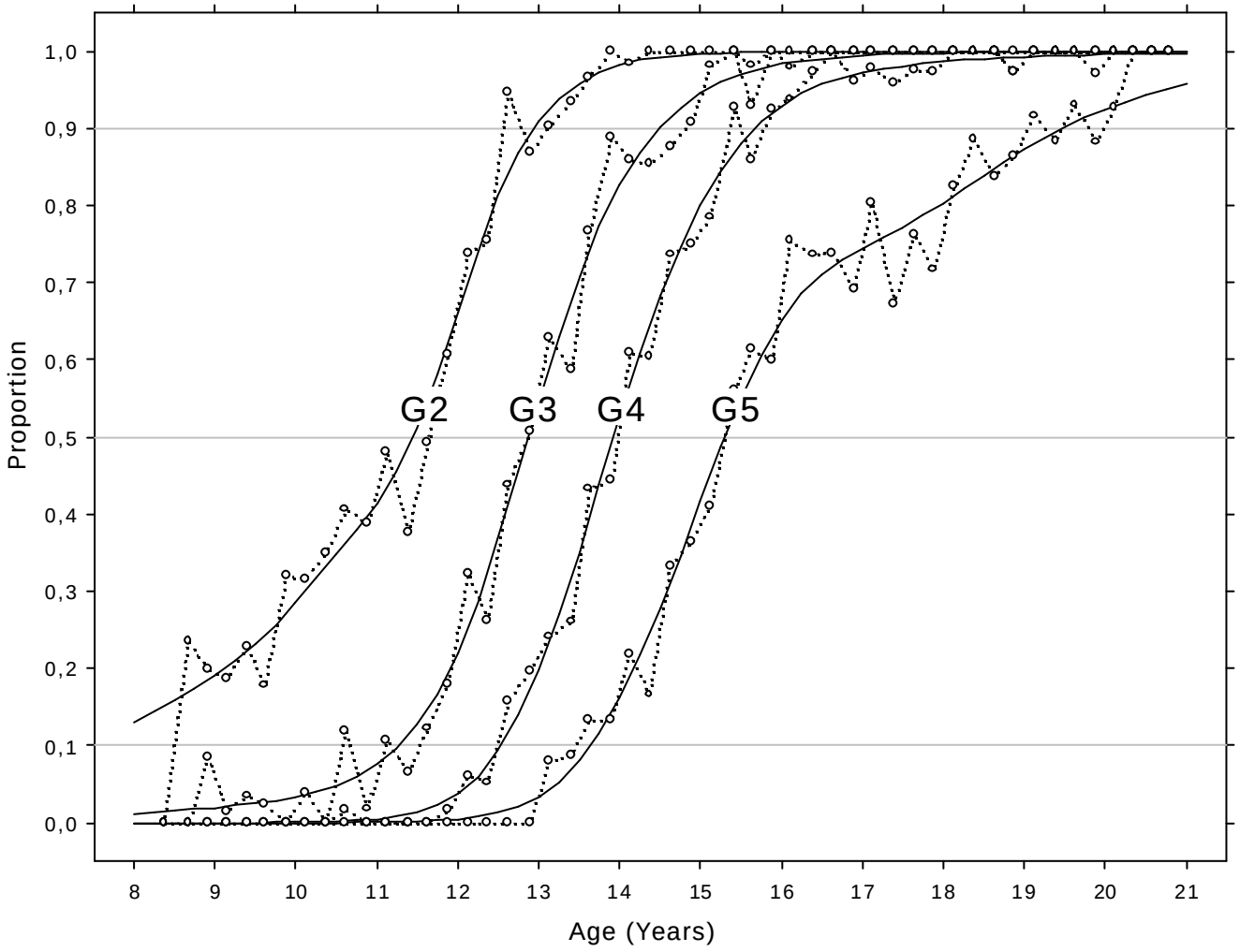
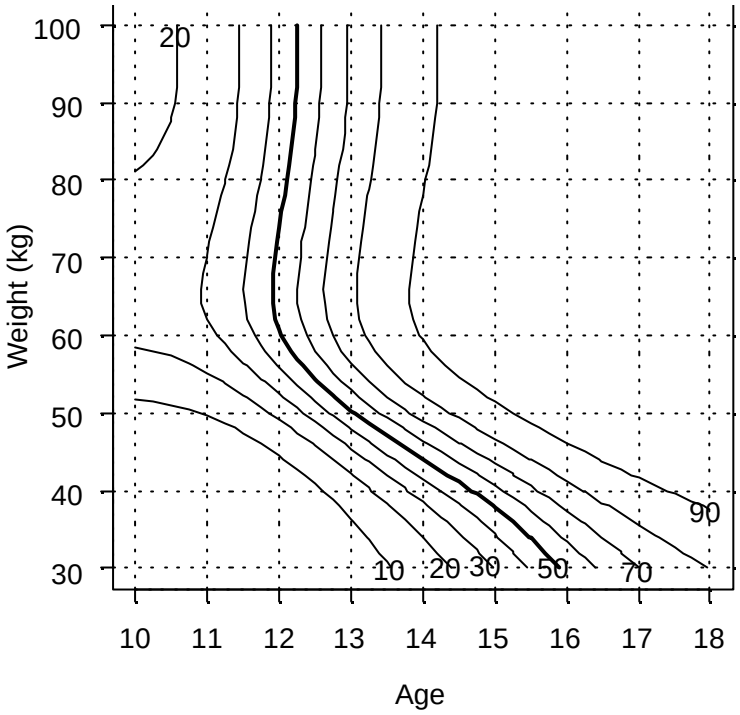
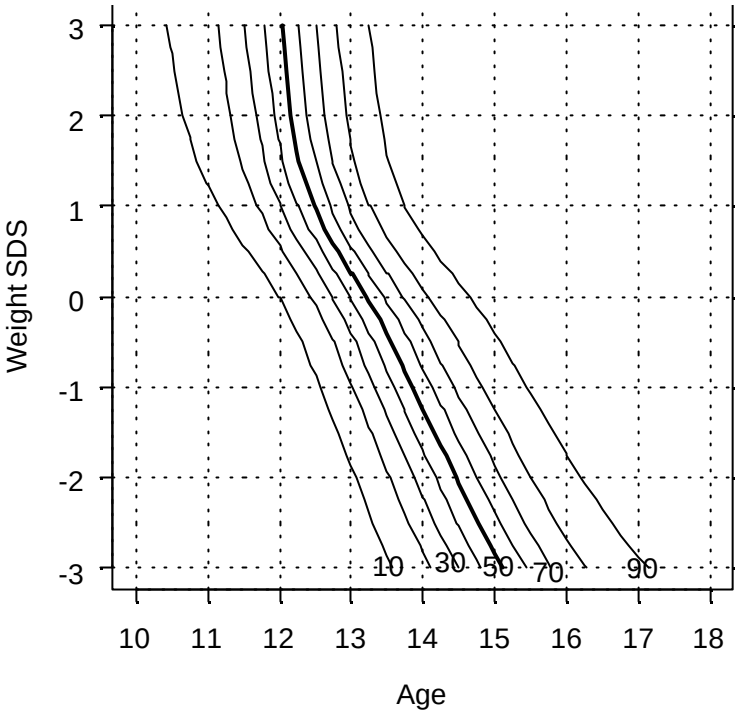

0
3
6
9
12
15
18
21
24
SDS
3
2
1
0
-1
-2
-3
-4
-5
58cm
60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94cm
56
56 58 60 62 64 66 68 70 72 74 76 78 80 82 84 86 88
54
52
52 54 56 58 60 62 64 66 68 70 72 74 76 78 80cm
50
48
46
44
P99
P97
P90
P75
P50
P25
P10
P3
P1

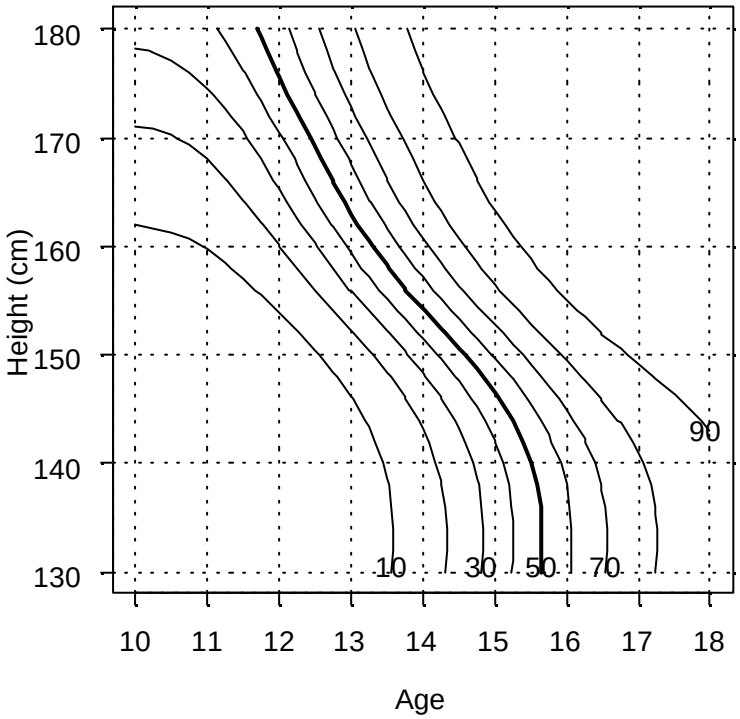
figure 5

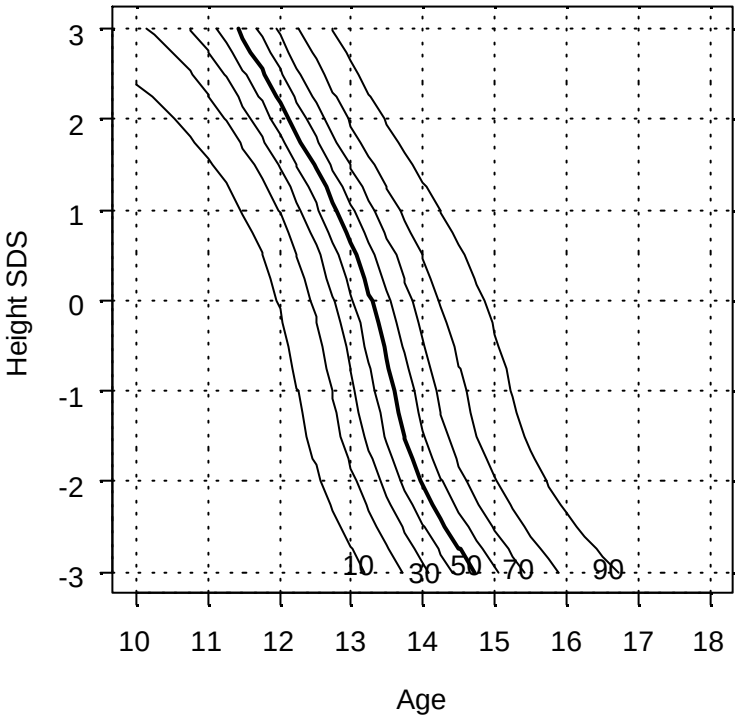
figure 6

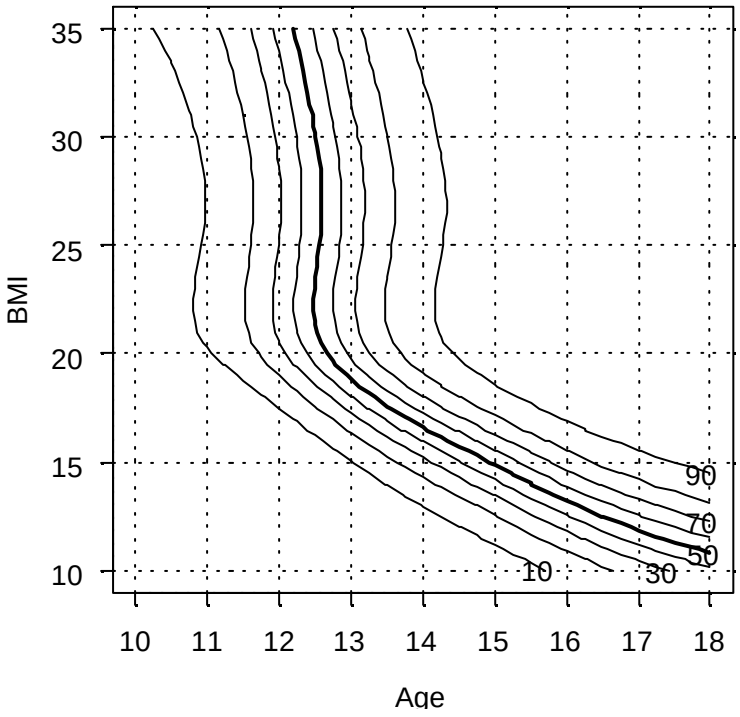

Proportion
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0,0
G2
G3
G4
G5
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Age (Years)

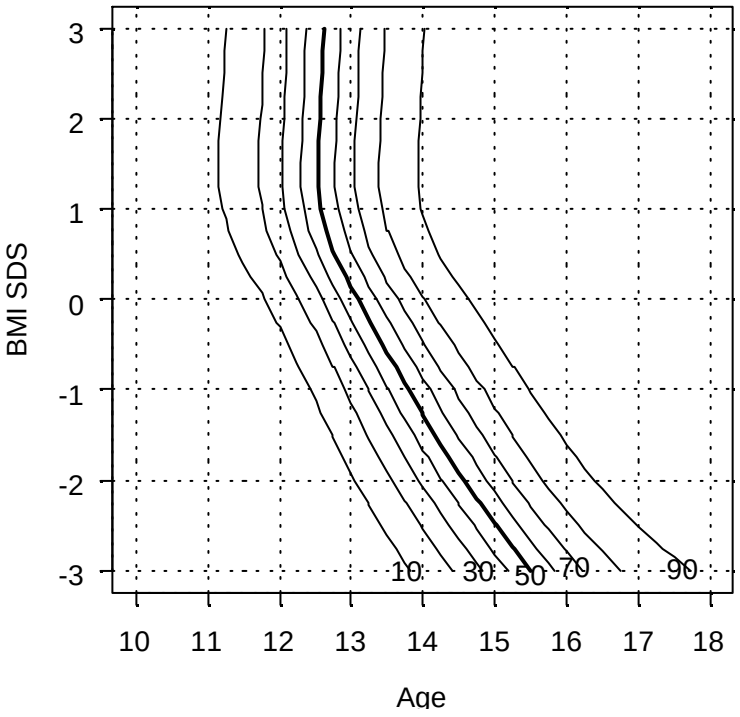

Weight (kg)
100
90
80
70
60
50
40
30
20
10
20
30
50
70
90
10
11
12
13
14
15
16
17
18
Age


Weight SDS
3
2
1
0
-1
-2
-3
10
30
50
70
90
10
11
12
13
14
15
16
17
18
Age


Height (cm)
180
170
160
150
140
130
90
10
30
50
70
10
11
12
13
14
15
16
17
18
Age


Height SDS
3
2
1
0
-1
-2
-3
10
30
50
70
90
10
11
12
13
14
15
16
17
18
Age


BMI
35
30
25
20
15
10
90
70
50
10
30
10
11
12
13
14
15
16
17
18
Age


BMI SDS
3
2
1
0
-1
-2
-3
10
30
50
70
90
10
11
12
13
14
15
16
17
18
Age