Methods to obtain referral criteria in growth monitoring

Paula van Dommelen and Stef van Buuren

Stat Methods Med Res published online 1 February 2013
DOI: 10.1177/0962280212473301

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–21

DOI: 10.1177/0962280212473301
smm.sagepub.com
SAGE

Methods to obtain referral criteria in growth monitoring

Paula van Dommelen[1] and Stef van Buuren[1,2]

Abstract

An important goal of growth monitoring is to identify genetic disorders, diseases or other conditions that manifest themselves through an abnormal growth. The two main conditions that can be detected by height monitoring are Turner's syndrome and growth hormone deficiency. Conditions or risk factors that can be detected by monitoring weight or body mass index include hypernatremic dehydration, celiac disease, cystic fibrosis and obesity. Monitoring infant head growth can be used to detect macrocephaly, developmental disorder and ill health in childhood. This paper describes statistical methods to obtain evidence-based referral criteria in growth monitoring. The referral criteria that we discuss are based on either anthropometric measurement(s) at a fixed age using (1) a Centile or a Standard Deviation Score, (2) a Standard Deviation corrected for parental height, (3) a Likelihood Ratio Statistic and (4) an ellipse, or on multiple measurements over time using (5) a growth rate and (6) a growth curve model. We review the potential uses of these methods, and outline their strengths and limitations.

Keywords

Growth, monitoring, screening, diagnostics, percentile, standard deviation score, likelihood ratio, ellipse

1 Introduction

Growth monitoring in infancy and childhood has been part of preventive child health programs for more than a century in both developed and underdeveloped countries. It is a popular tool for defining health and nutritional status of children.[1] At the individual level, growth monitoring includes measuring the individual's height, weight, head circumference and parental height. The individual growth curve is then plotted on a growth chart and interpreted in order to decide whether referral to a pediatrician is needed. When a child is referred, the pediatrician aims to detect the cause of the growth disorder.[2,3]

Growth monitoring is most useful in identifying growth disorders that meet two criteria. First, there should be no other clinically obvious points that might alert parents and primary care staff. Second, the growth pattern should deviate substantially from normal in most cases of the condition. The two most prevalent conditions that may meet these criteria for height are Turner's syndrome[2–6] and growth hormone deficiency.[2] Other conditions or risk factors with a slowed growth for weight

[1]Department of Statistics and Epidemiology, TNO, 2301 CE Leiden, The Netherlands
[2]Department of Methodology & Statistics, University Utrecht, Utrecht, The Netherlands

Corresponding author:
Paula van Dommelen, TNO, P.O. Box 2215, 2301 CE Leiden, The Netherlands.
Email: Paula.vanDommelen@tno.nl

or body mass index (BMI) include hypernatremic dehydration,[7] celiac disease,[8] cystic fibrosis[9] and obesity.[3,10] Conditions with an abnormal head growth are macrocephaly, developmental disorder and ill health in childhood.

There is little consensus on referral criteria for growth among industrialized countries.[11] As far as we know, three consensus-based guidelines have been published: the Finnish guideline,[12,13] the United Kingdom guideline[2] (''Coventry Consensus'') and the Dutch consensus guideline.[14] The Finnish and UK consensus guidelines were not validated. In 2004, it was shown that if the Dutch consensus guideline would be followed strictly, an unacceptably high percentage (over 80%) of all checked (healthy) children would have to be referred.[15] In order to improve the situation, several studies were initiated to obtain evidence-based referral criteria in growth monitoring.[4–10] The aim of this paper is to review the statistical methods in these studies and to obtain evidence-based referral criteria in growth monitoring. We outline the strengths and limitations of each method.

2 Quality criteria of growth monitoring as a screening program

Evidence-based referral criteria are needed if growth monitoring is used as a screening program.[10] According to the United Kingdom (UK) national screening committee (NSC), screening is defined as: ''a public health service in which members of a defined population, who do not necessarily perceive they are at risk of, or are already affected by a disease or its complications, are asked a question or offered a test, to identify those individuals who are more likely to be helped than harmed by further tests or treatment to reduce the risk of a disease or its complications.''[16] The intention of screening is to identify the disease early, thus enabling timely intervention and management in the hope to reduce mortality and suffering from the disease. Growth monitoring can be considered as a screening program. However, in general there are three important differences between growth monitoring and other screening programs.[10] First, a referral is often made on a combination of abnormal growth and other clinical symptoms, while the result of a conventional screening program usually only depends on the result of a test. Second, growth monitoring is aimed at identifying multiple diseases simultaneously, while a screening program is often aimed at identifying one disease. Third, growth monitoring is usually performed over a longer time period, while a screening program offers a test to the population at one moment in time or several tests at a given age.

The NSC developed 22 quality criteria for appraising the viability, effectiveness and appropriateness of a screening program.[16] The 22 NSC quality criteria are subdivided into four groups: (1) the epidemiology of the condition; (2) the properties of the test; (3) any treatment options and (4) the acceptability of the screening program. We will discuss several elements of these four groups in detail.[10]

2.1 Epidemiology of the condition

The first group of quality criteria deals with the epidemiology of the condition. The condition should be an important health problem. It is known that an abnormal growth can be an early sign of several conditions and that some of these conditions are important health problems. The rapid rise in childhood obesity is a good example.

2.2 Propertjes of the test

The second set of quality criteria concerns the properties of the test or referral criteria. This implies that the referral criteria should be acceptable to the population, safe, precise, simple and validated.

Growth monitoring is currently already widely implemented in practice and is generally considered as safe. Accurate measurements include a standardized measurement technique, quality equipment, which is regularly calibrated and accurate, and trained measurers.[17,18] Information on the appropriate equipment and techniques for accurate weighing and measuring of infants, children and adolescents are needed and is available from the International Organization for Standardization; a worldwide federation of national standards bodies.[19] Furthermore, the referral criteria should be simple. The simplest criterion consists of comparing a single anthropometric measurement to some norm. More advanced criteria may involve multiple measurements over time. Repeated measurements of height and weight as part of scheduled visits at child health care centers have been suggested,[20] but until now there is little information with regard to the most cost-effective number and timing of visits.

An important element of the quality criteria is that there should be a validated screening test. In growth monitoring, a screening test consists of one or several referral criteria based on growth. For example, the lowest centile of a growth chart for height can be used as a screening test. The test is then considered positive if a child's height is below this centile and negative if above. Validated or evidence-based referral criteria are able to detect, at an early stage, as many children with growth-related conditions as possible (high sensitivity) at the account of only a limited number of children with a false-positive result (high specificity).

2.3 Role of interventions

A third type of quality criteria states that there should be an effective treatment or intervention for the children with conditions identified through early detection, with evidence leading to better outcomes than late treatment. For most of the growth-related conditions, it is known that early treatment improves outcomes. Early treatment with growth hormone has proved its efficacy in the treatment of various conditions with short stature.[21] Children with Turner's syndrome have higher disease risks, especially due to the risk of dissection of the aorta and other cardiovascular diseases, as well as the risk of type 2 diabetes, osteoporosis and thyroid disease.[22] Early detection of children with Turner's syndrome enables the pediatrician and cardiologist to evaluate these children. Children with Turner's syndrome should be screened for hypertension and electrocardiographic abnormalities in addition to anatomic anomalies. Blood pressure should be monitored on a regular basis.[23] For children with celiac disease, early detection and treatment with a gluten-free diet is required to improve the immediate quality of life of the patients and to decrease the long-term risks, including a higher prevalence of malignancies, adverse pregnancy outcome, neurological problems and osteomalacia.[24] For children with cystic fibrosis, treatment may include pulmonary therapy (treatments to maintain lung function) and nutritional therapy. For infants with hypernatremic dehydration, early detection is needed to prevent serious complications, such as fits, disseminated intravascular coagulation, multiple cerebrovascular accidents and even death. Evidence on the long-term effects of treating childhood obesity is still limited.[25]

2.4 Cost-effectiveness

Finally, there are quality criteria related to the acceptability of the screening program. An important element of these criteria is that the screening program should be cost-effective. Screening should be worth the money. The direct costs of growth monitoring include equipment, staff, training of staff and the costs associated with referral for further investigation. A more demanding test is more expensive, but may also have a higher yield. Studies on economic modeling suggest that growth monitoring is associated with health improvements and is cost-effective.[3] In general, for an adequate

evaluation of the cost-effectiveness, we need insight into the diagnostic performance of various referral criteria and the costs that are associated with this. From a societal perspective, a high specificity for the referral criteria is desirable to minimize unwanted health costs, to free clinical practitioners from being overloaded by work, to evade unnecessary interventions and treatments for healthy children and to reduce parental and child's anxiety.

In summary, many steps have to be taken into account for growth monitoring to fulfill the quality criteria of a screening program. This paper will focus on the validation of referral criteria, an element of the second group of quality criteria.

3 Statistical referral criteria in growth monitoring

Most research to date has been done on cross-sectional data, producing reference charts and growth standards. A reference chart is a graphic presentation of how children normally grow without making any claims about the health of its population, whereas a standard represents 'healthy' growth of a population and suggests a target to achieve.[26–28] In 2006, the WHO published growth standards representative of children being raised according to recommended health practices.[29] Several of these conditions are exclusive or predominant breastfeeding for four to six months and optimal health care and an optimal environment without conditions that could limit growth, such as smoking. Although the WHO outlines how children should grow, there is some controversy whether the WHO growth chart applies to all countries of the world.[30] Apart from the WHO charts, there are reference charts available for different ethnicities (also within one country), twins,[31] preterm infants[32] and syndromes.[33–43] Syndrome-specific charts are useful for comparing individual children to other children with the same diagnosis and, perhaps, to detect other pathologies within this group.

Many European countries use a local reference chart, while countries in Asia, Africa, Latin America and Caribbean, Northern America and Oceania mainly adopt the NCHS/WHO (now replaced by the WHO chart).[29] The United States is now using the growth charts of the Centers for Disease Control and Prevention (CDC).[44] There is also a wide variation in anthropometric indexes used in child growth monitoring.[45] Most countries use charts of weight for age, but some European countries and Northern America assessed length/height for age and head circumference for age.[46]

Several studies were performed to obtain evidence-based referral criteria in monitoring height, weight and/or BMI.[4–10] In the next paragraphs, we review the potential uses of methods to obtain referral criteria for growth and outline their strengths and limitations. The choice among these criteria depends on the situation in which it is to be applied. Important dimensions on which the criteria vary are:

(1) Are we screening for specific disorders? Do we incorporate disease-specific growth references into the criteria?
(2) Do we refer on a single measurement taken at a specific age, two measurements at different ages or multiple measurements at different ages?
(3) Do we refer on a single type of measurement (e.g. height) or do we combine different measurements (e.g. height and sitting height)?
(4) Are we trying to improve precision by incorporating important covariates (like height of the parents) into the criterion?

Depending on the response of these questions, some types of referral criteria will be more appropriate than others. The referral criteria that we discuss are based on either anthropometric

measurement(s) at a fixed age using (1) a Centile or a Standard Deviation Score (SDS), (2) a SDS cut-off corrected for parental height, (3) a Likelihood Ratio Statistic and (4) an ellipse, or on multiple measurements over time using (5) a growth rate and (6) a growth curve model.

3.1 Data

Receiver operating characteristic (ROC) curves are presented for the criteria based on longitudinal height from birth to 10 years of 777 girls with Turner'syndrome,[4–6,47] longitudinal height and weight from birth to 2.5 years of 102 children with celiac disease[5,8,48,49] and 216 children with cystic fibrosis,[5,9] cross-sectional height and sitting height from 10 children with Marfan and 10 children with hypochondroplasia,[50] a reference sample of longitudinal data from birth to 10 years of a cohort of 970 children (489 girls) born in Limburg in the Netherlands,[4–6] longitudinal height and weight data from birth to 2.5 years of a reference group obtained from the Social Medical Survey of Children Attending Child Health Clinics (SMOCC) cohort (n = 2151) in The Netherlands[5,8,9,51] and cross-sectional data from the Fourth nationwide growth study in the Netherlands.[52] Parental height was available for children with Turner's syndrome, celiac disease, cystic fibrosis and the references. If missing, parental height was imputed using multivariate imputation by chained equations.[53] For all details about these data, methods and analyses, we refer to van Buuren et al.,[6] van Dommelen et al.[4,8,9] and Grote et al.[5]

3.2 Referral criteria based on a fixed centile or SD curve

Some countries present their charts in centiles, while others use Standard Deviation Scores (SDS[54]). Centiles indicate a child's position within the distribution of the reference population. As an example, if a child has a height-for-age at the third centile, then 3% of the reference population who are at the same age and sex are shorter than that child. The simplest referral criteria are based on a centile (P) with a certain cut-off level (c). For an anthropometric measurement y, the referral criterion becomes as follows:

(1) if $P(y) < c$, then refer, else not (centile criterion).

Due to its simplicity, this type of referral criteria is widely being used. It is, however, unknown how many of those referred actually have a growth-related condition, or how many that have a growth-related condition are being identified by the criterion. Without any further information, the sensitivity (the true positive rate) is unknown.

In order to make progress, we need assumptions about the (height) distribution of those we wish to detect. For example, one of the conditions we wish to detect is girls with Turner's syndrome. If we have a reference for the diseased population, then we can calculate the sensitivity of a criterion. For example, we take the P_3 at 6 years as cut-off value in the centile criterion. For 6-y old girls with Turner's syndrome, the reference mean is 104.5 cm with a standard deviation of 4.2 cm. The sensitivity is then 87%, because 87% of the Turner's syndrome population has a height below the third centile of the normal reference group at 6 years of age.

There is a clear trade-off between sensitivity and specificity: higher sensitivity will usually mean lower specificity (and hence more unnecessary referrals); higher specificity will usually mean lower sensitivity (and hence more missed children with growth-related conditions). Furthermore, sensitivity and specificity depend on age and can be calculated for each cut-off level at each age or age group. For example, sensitivity is 81% in 4-year-old girls. So sensitivity at 4 years of age is

lower than at 6 years of age. Therefore, it is usually easier to find the condition if we wait longer. However, early treatment often leads to better outcomes compared to treatment at an older age. Moreover, sensitivity and specificity depend on the number of measurements of the child. For example, a child with one measurement has, by definition, a probability of 0.6% of having a height below $P_{0.6}$. However, for a child with fifteen measurements, this probability is equal to 6.2%.[15] These differences are related to the number of measurements or repeating testing. Infants are difficult to measure accurately and this causes variability in height. This variability increases the probability of having a value below a certain cut-off value. In summary, optimizing sensitivity, specificity, referral age and the number of measurements is an interrelated problem.[10]

While centiles are easy to explain, a limitation is that a measurement well below the third centile cannot accurately be defined on the growth chart. Also, the distances between the centiles are not equally distributed. For example, the difference in centimeters between the 10th and the 20th percentile is larger than between the 20th and the 30th percentile. The SDS is an alternative that expresses the measurement y relative to a reference population in units of standard deviations above or below the median. If the reference is available in the form of the LMS model and SDS is equal to z, then

$$z(y) = \frac{(y/M)^L - 1}{LS}, L \neq 0 \text{ and } z(y) = \frac{\ln(y/M)}{S}, L = 0$$

where L, M and S are three age-dependent curves of, respectively, the skewness, the median and the coefficient of variation.[47] Similar calculations are possible for other distributions like de Box-Cox Power Exponential (BCPE).[55] The BCPE distribution provides a generalization of the LMS method to data exhibiting not only skewness, which the LMS method can deal with, but also kurtosis that deviates appreciably from that of the normal distribution. For normally distributed data, such as height and head circumference, calculation of the SDS can be simplified to

$$z(y) = \frac{y - M}{MS}$$

with M the mean and MS the standard deviation at age of measurement. The 0 SDS line represents the median. At every age, 50% of children in the reference population fall above this line and 50% below it. Most children (95%) fall between the −2 and +2 SDS lines. Referral criteria for SDS and centiles are technically equivalent. For an anthropometric measurement y the referral criterion becomes as follows:

(2) if $z(y) < c$, then refer, else not (SDS criterion)

for a certain cut-off value c with a different value than in formula (1) and measurement y.

For example, the SDS of a 9-year-old boy with a height of 125 cm is equal to (125–138.52)/(138.52 * 0.0447) = –2.2 SD, because the M and S in the general population of 9-year-old boys are, respectively, 138.52 and 0.0447. This boy should be referred if a cut-off value of $c = -2$ is applied.

Figure 1 shows the ROC curves for the SDS criterion (or similarly the centile criterion) based on children with Turner's syndrome and a reference population from Limburg stratified by the age groups 0–3 years (cut-off values −5, −4, −3.5, −3) and 3–10 years (cut-off values −3.5, −3, −2.5, −2).

For easy application, it is recommended to use cut-off levels that are similar to the centiles or standard deviation lines of most growth chart.[48,49] In this way, the referral criterion can easily be applied to each measurement in daily practice. We recommend the −2.5 SD line at ages 3–10 years as both sensitivity (74%) and specificity (99.1%) are high.[7]

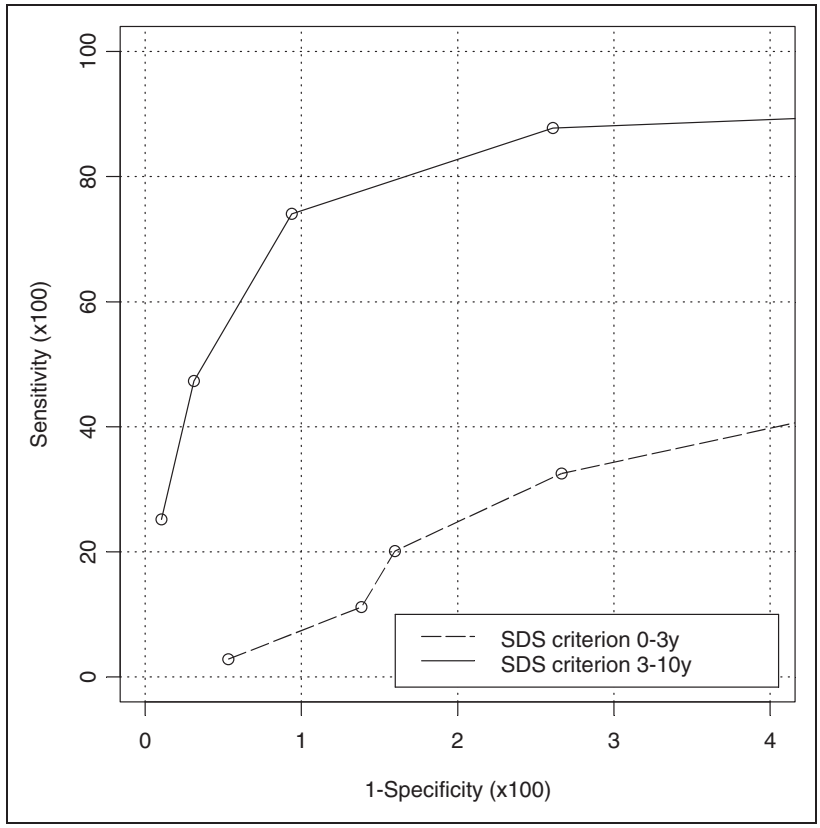

Figure 1. Receiver operating characteristic (ROC) curves of the Standard Deviation Score (SDS) criterion for ages 0–3 years and 3–10 years to detect children with Turner's syndrome.

3.3 Allowing for covariates

Criteria based on a single SDS can be extended by accounting for covariates that are known to have a significant effect on growth. An important covariate is parental height. Target height (TH) is the term used for the expected height of a child given the height of the (natural) parents. Several formulas were put forward that were all based on the mid-parental height[56] (MPH), the average height of the two parents. The MPH can be corrected for sex, secular trend and regression to the mean.[57–59] Regression to the mean is the phenomenon that very short or tall parents tend to produce less extreme offspring. The TH can be converted to the target height SDS (THSDS) by the following formula

$$z(TH) = (TH - M)/SD$$

with $z(TH)$ equal to the THSDS and M and SD, respectively, the mean and standard deviation of the final height. When using reported parental height, one should notice that a discrepancy between measured and reported heights has been noticed.[60] Men tend to overestimate and women tend to underestimate their height. The difference between reported and measured height is positively associated with age of the parent, and there is a wide individual variation between reported and

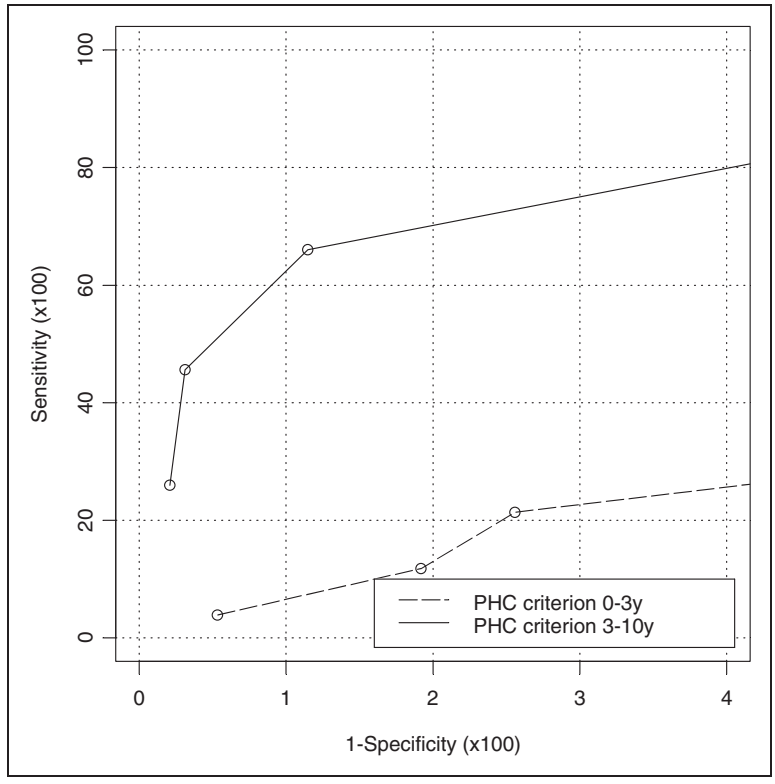

Figure 2. Receiver operating characteristic (ROC) curves of the parental height corrected (PHC) criterion for ages 0–3 years and 3–10 years to detect children with Turner's syndrome.

measured heights in both sexes.[61,62] It is recommended to obtain measured parental heights. If this is difficult in daily practice, one can use reported parental height. To correct the child's SDS for the THSDS, one can use the following referral criterion:

(3) if $z(y) - z(TH) < c$, then refer, else not (parental height corrected criterion)

with cut-off level c and measurement y. Figure 2 shows the ROC curves for the parental height corrected (PHC) criterion based on children with Turner's syndrome and a reference population from Limburg stratified by the age groups 0–3 years (cut-off values are -5, -4, -3.5) and 3–10 years (cut-off values -3.5, -3, -2.5).

To keep specificity (98.9%) and sensitivity (66%) high, a recommended cut-off level is $c = -2.5$ in the age group 3–10 years. The sensitivity is lower than according to the SDS criterion.

Sensitivity can be increased by using a combination of the SDS criterion and the PHC criterion, such that:

(4) if $z(y) - z(TH) < c_1$ and $z(y) < c_2$, then refer, else not (PHC SDS criterion).

Figure 3 shows the ROC curves for the PHC SDS criterion based on children with Turner's syndrome and a reference population from Limburg stratified by the age groups 0–3 years (optimal

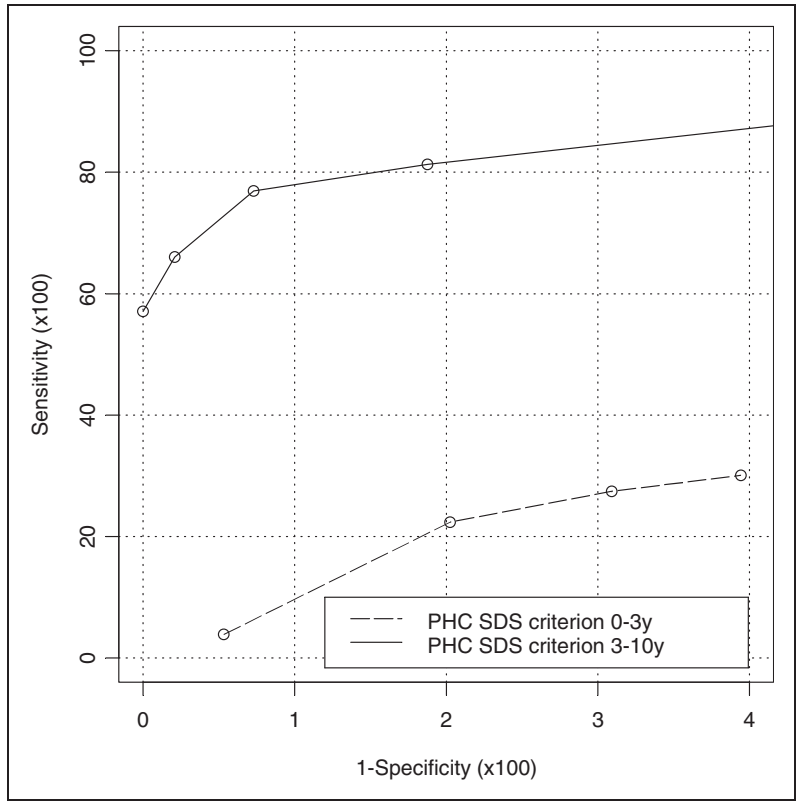

Figure 3. Receiver operating characteristic (ROC) curves of the parental height corrected (PHC) Standard Deviation Score (SDS) criterion for ages 0–3 years and 3–10 years to detect children with Turner's syndrome.

(and shown) cut-off values are $(c_1, c_2) = (-3,-5)$, $(-3,-3)$, $(-2.5,-3)$, $(-2,-3)$) and 3–10 years (optimal (and shown) cut-off values $(c_1, c_2) = (-2.5,-2)$, $(-2,-2)$, $(-1.3,-2)$, $(-1.3,-1.3)$).

We suggest to apply this criterion to children aged 3 onwards as was found as the best cut-off value for age by van Buuren et al.[6] However, further analyses revealed that already from the age of 1 year, including parental height into the criterion results in higher sensitivities at fixed specificities compared to the SDS criterion. Recommended cut-off values are $c_1 = -2$ and $c_2 = -2$, such that sensitivity is 76.9% and specificity is 99.3% in 3–10-year-old children. If one accounts for regression to the mean and no secular trend, we suggest cut-off values of $c_1 = -1.6$ and $c_2 = -2$.[63]

3.4 Referral criteria based on a likelihood ratio statistic

Referral criteria based on SDS are not optimal for answering the question: "Does this child with measurement Y belong to the reference or to the diseased (or syndrome-specific) population?"[64] When determining whether the child has a specified disease, more powerful diagnostic measures exist, such as the likelihood ratio statistic (LR).[65,66] The LR can only be calculated when the reference charts for the general and the diseased population are known. Currently, reference charts are available for children with Down,[33] Turner,[34] Noonan,[35] Prader-Willi,[36] Silver-Russell,[37] Cri-du-chat,[38] Wlliams,[39] achondroplasia[40–42] and Ellis-van Crefeld syndrome.[43]

With these charts, means and SD can be used to calculate the LR. The LR measures how many times more likely patients with the disease are to have a particular result y than patients without the disease. The LR for a result y is defined by

$$LR(y) = \frac{f_1}{f_2}$$

with f_0 and f_1 the density function for, respectively, the reference and diseased population. Let z_0 and z_1 denote the SDS of the measurement y in the reference and disease populations, respectively. For example, a 6-year-old girl has a height of 109 cm.[64] This girl is short when compared with the Dutch reference[52] ($z_0 = -1.94$), but taller than girls with Turner's syndrome[34] ($z_1 = 1.07$). The density function for the reference population at 6 years of age is normally distributed with a mean of 118.7 and a standard deviation of 5.0. The density f_0 at 109 cm is then equal to

$$f_0(y = 109, \mu = 118.7, \sigma^2 = 5.02^2) = 0.0122$$

Similarly, the density for the diseased population is

$$f_1(y = 109, \mu = 104.5, \sigma^2 = 4.22^2) = 0.0535$$

The LR is thus equal to

$$LR(y) = \frac{f_1}{f_2} = \frac{0.0535}{0.0122} = 4.4$$

Therefore, observing a height y of 109 cm is 4.4 times more likely in the Turner population than in the non-diseased population. The higher the value of the LR, the stronger the evidence for the presence of the disease. Referral criteria based on LR can be defined as:[64]

(5) if $LR(y) > c$, then refer, else not (LR criterion),

with cut-off level c and measurement y.

Figure 4 shows the ROC curves for the LR criterion based on children with Turner's syndrome and a reference population from Limburg stratified by the age groups 0–3 years and 3–10 years. Only girls were selected from the Limburg data. A specific (99%) criterion for children aged 3–10 years has a cut-off value of $c = 3.5$. No good comparison can be made with the ROC of Figure 1, because only girls were selected in this example. Further analysis showed that the ROC curve of the LR criterion is almost similar to the ROC curve of the SDS criterion.

Note that if the prevalence of the disease is low, even very large LR will not give much evidence for the disease. If the disease prevalence is known, it is also possible to calculate the absolute disease risk from Y. For a low probability of the disease and when the LR is not huge, the probability of the disease given height Y of the child is approximately the LR(Y) times the prevalence of the disease, which is LR(Y) * 1/2500 for Turner's syndrome. Cut-off values for this probability can also be used as a referral criterion.

3.5 Referral criteria based on an ellipse

It is often of interest to identify people whose anthropometric measures are not in proportion to each other. Examples include: waist in proportion to hip or height in proportion to weight. Current

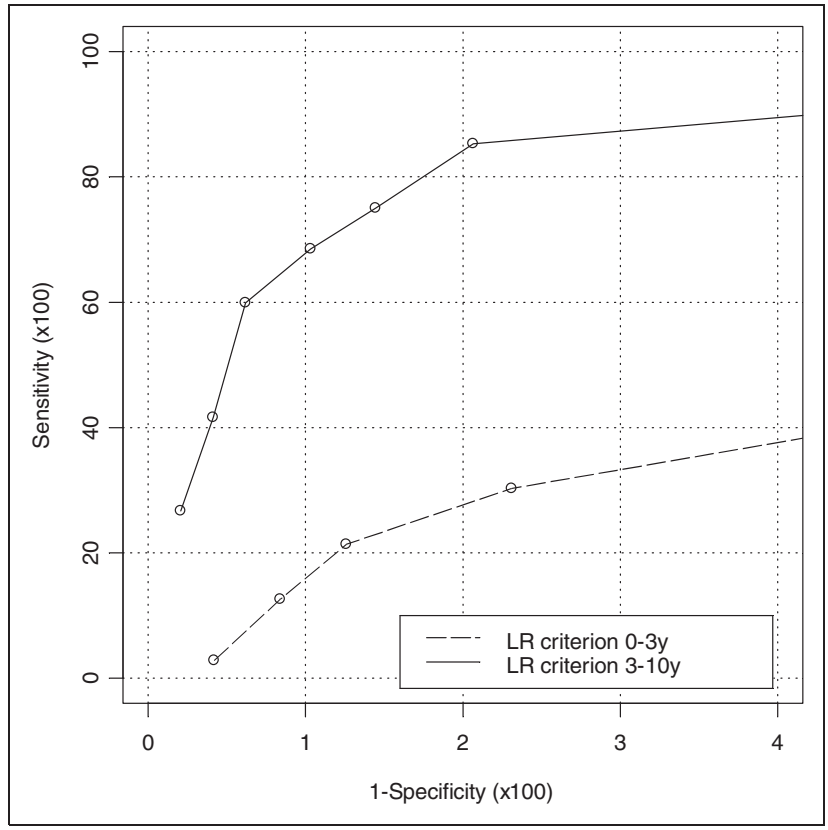

Figure 4. Receiver operating characteristic (ROC) curves of the likelihood ratio statistic (LR) criterion for ages 0–3 years and 3–10 years to detect children with Turner's syndrome.

solutions include conditional measures such as weight conditional on height, or aggregating measures like the waist-for-hip ratio or the BMI. An alternative approach is to look at the joint distribution of z (SDS of the first anthropometric measure) and z^* (SDS of the second measure). By definition, z and z^* are standard normally distributed in the reference population. If the regression of z on z^* (and vice versa) is linear then the joint distribution (z, z^*) is bivariate normally distributed.[67]

The bivariate normal distribution (z, z^*) describes the proportion of cases at each combination of z and z^*. We can find all points with the same proportion by taking slices parallel to the bottom plane. The boundaries of such a slice correspond to the equal probability contour. For example, the bivariate 90% confidence interval corresponds to the contour in which 90% of the children are located. For a bivariate normal distribution, contours correspond to ellipses. The shape of the ellipses only depends on the correlation between z and z^*.

The proportion of children within the ellipse is linearly related to the height of the probability contour in the bivariate normal distribution by the following formula:

$$P_e = \xi_0 + \xi_1 h \text{ with } \xi_0 = 1 \text{ and } \xi_1 = -2\pi\sqrt{1-\rho^2}$$

with $z \sim N(0,1)$, $z^* \sim N(0,1)$, $Cor(z, z^*) = \rho$, $0 \leq \rho < 1$, $0 \leq h = height < \max$.

Referral criteria can be constructed by referring all children whose measurements fall outside the ellipse:

(6) if $P_e < c$, then refer (ellipse criterion).

This identifies children for which the combination of z and z^* deviates from the reference population. Note that, when taken separately, both values can be 'normal'. For example, the combination (2,–2) lies outside the 1% ellipse, while each separate value has a probability of about 0.02. The reverse also occurs. For example, combination (–2,–2) falls within the 5% ellipse, while the separate values are more rare. The ellipse method with sitting height/height ratio (SH/H) SDS against height SDS has been tested on 10 children with Marfan syndrome and 10 children with hypochondroplasia. The ellipse was drawn around 95% of the data points in the scatter plot of SH/H SD against H SDS in children from the general population. In 4 out of 10 patients with Marfan syndrome was SH/H SDS located below the conditional −2 SD line. The ellipse criterion performed better: 6 of 10 patients with Marfan syndrome were located outside the ellipse. Furthermore, a total of 8 out of 10 patients with hypochondroplasia were located above the conditional +2 SD line, while all hypochondroplasia cases were located outside the ellipse. This suggests that the ellipse method is more powerful than criteria based on aggregated measures. The method requires that the correlation for each age and sex has to be known.

3.6 Growth rate between two time points

Several statistics can be calculated based on two measurements over time. Repeated measurements over time allow for calculation of a growth rate and can be used to define an abnormal increase or decrease in growth. An abnormal decrease of growth is often named as failure to thrive. Six screening criteria for failure to thrive were formulated.[10,11] If there are more than two measurements, criteria based on these statistics can be applied to all possible pairs of measurements. For example, if weight is measured at ages A, B and C, the growth rate can be calculated for the intervals AB, BC and AC.

1. Delta criterion

Refer a child if an absolute change (c_1) in length SDS, height SDS, weight SDS or BMI SDS occurs: $z(y(x_2)) - z(y(x_1)) < c_1$, with z the SDS and y the measurement at two different ages x_1 and x_2. For example, suppose a child has measurements at two different ages and $c_1 = -2$, then a child is referred if the second SDS is −2 below the first SDS.

2. Extended delta criterion

This criterion is equal to the first criterion with the extension that the second SDS has to be below a certain cut-off value c_2. Therefore, refer a child if $z(y(x_2)) - z(y(x_1)) < c_1$ and $z(y(x_2)) < c_2$.

3. SDS velocity

Signals whether an abnormal slowed growth for length, weight or BMI occurs in terms of change in SDS per year (c_1) in combination with a current low SDS (c_2):

$$\frac{z(y(x_2)) - z(y(x_1))}{x_2 - x_1} < c_1 \text{ and } z(y(x_2)) < c_2$$

For example, suppose a child has two measurements and $c_1 = -1$ and $c_2 = -1.5$. This child will then be referred if the difference between the second and first measurement per year exceeds 1 SDS (which corresponds to a decrease of 0.5 SDS within 6 months) and if the second measurement is less than −1.5 SDS. The term velocity commonly refers to cm or kg/year; for SDS it has the dimension 1/year.

4. Conditional weight gain criterion

Signals whether a child's conditional weight gain SDS is less than a certain value (c_1) with the restriction of having a low weight SDS (c_2):

$$\text{Conditional weight gain SDS} = \frac{z(y(x_2)) - rz(y(x_1))}{\sqrt{1-r^2}} < c_1 \text{ and } z(y(x_2)) < c_2$$

with z weight SDS and r the correlation coefficient between $z(y(x_1))$ and $z(y(x_2))$. This criterion compares an infant's current weight with that predicted from his or her previous weight, allowing for the fact that on average, light infants tend to grow faster than heavier infants[68,69] (regression to the mean).

5. Parental height deflection criterion

Signals whether a slowed growth for length SDS of the child moves away from the child's TH: $z(y(x_2)) < c_1$ and $|z(y(x_2) - z(TH)| > = |z(y(x_1) - z(TH)|$ with z length SDS. This criterion was defined because of the assumption that a correction might be needed for parental height in the first years of life: e.g. a baby that is born with a length SDS of −1 and has a target height SDS of +2 would be expected to cross the SD lines in upward direction in the first 2–3 years. A growth disorder could disturb this, and a stable length SDS of this child at −1 over the first 2 years could indicate growth pathology.

6. Combined weight and length deflection criterion

A slowed growth for length (c_2) occurs after a slowed growth for weight (c_1):

$$z(y(x_2)) - z(y(x_1)) < c_1 \text{ and } z^*(y(x_2)) - z^*(y(x_1)) < c_2$$

with z weight SDS and z^* length SDS.

The referral criteria based on failure to thrive have been tested on longitudinal weight, height and BMI between birth and 2.5 years of age of children with celiac disease, cystic fibrosis and the SMOCC reference sample. It appeared that the parental height deflection criterion and the combined weight and length deflection criterion did not have high sensitivities and specificities. The delta criterion, the SDS velocity and the conditional weight gain criterion showed sensitivities between 21 and 27% at specificities of 98.1 and 98.5%. The most optimal failure to thrive criterion was the extended delta criterion applied to BMI. Figure 5 shows the ROC curves for the extended delta criterion for BMI based on children with celiac disease (CD), cystic fibrosis (CF) and the SMOCC reference sample for children aged 0–2.5 years. Optimal cut-off values imply a very low BMI SDS (–2.5) in combination with a deflection (0.5 SDS) or a low BMI SDS (<–1.5) in combination with a large deflection (–2.5 SDS).

Traditionally, criteria based on failure to thrive are considered as a better screening tool. However, criteria based on two measurements are more sensitive to measurement error than criteria based on single growth measurements. Therefore, stricter cut-off values have to be chosen in order to have a high specificity. This has to be compensated by a high sensitivity.

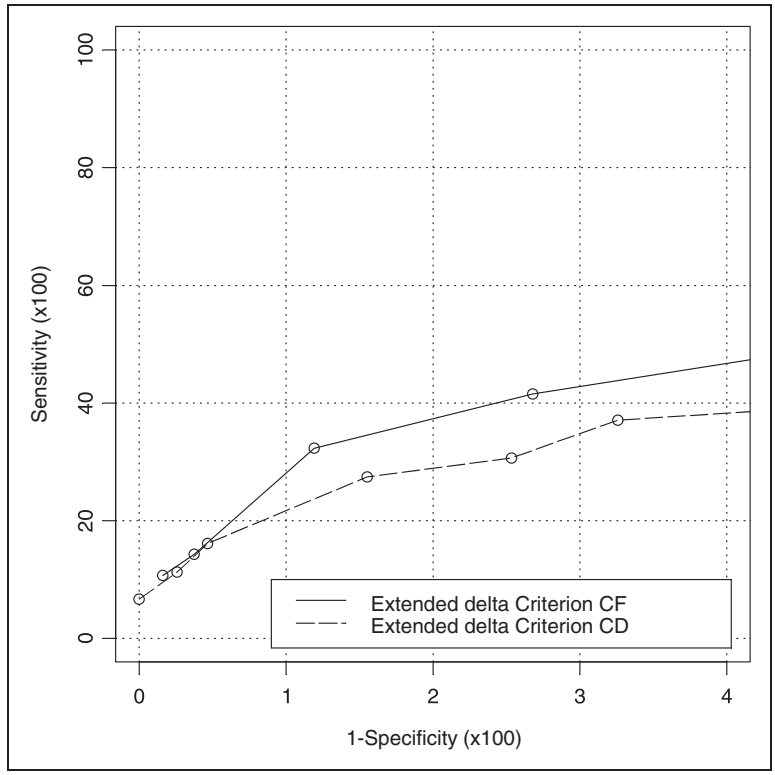

Figure 5. Receiver operating characteristic (ROC) curves of the extended delta criterion to detect children with cystic fibrosis (CF) and celiac disease (CD).

3.7 Growth curve models

Growth curve models describe growth over time, typically using three or more time points. These models are well suited to analyze longitudinal data when the times of measurements are irregularly spaced. The models summarize individual longitudinal data into several interpretable quantities such as growth at birth, growth velocity, growth acceleration or deceleration.

A number of growth curve models have been suggested in the literature and have been shown to be representative at different periods of life.[70–73] For example, the shape invariant model (SIM) between birth and 2 years of age,[71] the Jenns-Bayley growth curve that describes height of children from birth to 8 years of age,[72] and the infancy-childhood-puberty (ICP) growth curve model that decomposes height mathematically into three additive and partly superimposed components-infancy, childhood and puberty.[73] The parameters of the model can be estimated by a nonlinear regression model via least squares or by a mixed-effects model. A mixed-effects model assumes that each parameter is the sum of a fixed and a random component. The fixed components are the same for every individual and the random components may differ between individuals according to a normal distribution. For example, according to Jenss-Bayley,[72] height can be modeled by a nonlinear mixed-effects model as:

$$y_i(t) = \alpha_1 + \alpha_2 t - \exp\{\alpha_3 + \alpha_4 t\} + \varepsilon_{it}$$

with $\alpha_k = \beta_k + \lambda_k$, β_k fixed effects and $\lambda_k$random effects, for $k = 1, \ldots, 4$ with t the age in years, $y_i(t)$ the height (in cm's) of the ith child at age t with $i = 1, \ldots, n$, n the number of children and ε_{it} is the measurement error at age t.

A nonlinear mixed-effects model is preferred, because it borrows strength across individuals in estimating individual parameters. If the longitudinal growth of children with a condition is different from the growth of healthy children, this difference will be captured by the parameters of the model.[4] Discriminant or logistic regression analyses can be used to create a model that explains the grouping of the reference children and the children with a condition. In the discriminant analyses, we can use a heteroscedastic discriminant model if the group with the condition and the reference group have different covariance matrices. This leads to a quadratic discriminant function of the form:

$$d_i(\bar{x}) = \beta_{i0} + \beta_{i1}\bar{x} + \bar{x}^T\beta_{i2}\bar{x},$$

where

$$\beta_{i0} = -\frac{1}{2}\left(p\log|\Sigma_i| + \bar{\mu}_i^T\Sigma_i^{-1}\bar{\mu}_i\right),\ \beta_{i1} = \bar{\mu}_i^T\Sigma_i^{-1} \text{ and } \beta_{i2} = -\frac{1}{2}\Sigma_i^{-1}$$

with $\sum_i$the covariance matrix of group i and p-variate normal random variables $N_p(\bar{\mu}_i, \sum_i)$ for $i = 1, 2$ (condition and reference group) and p the number of parameters of the growth model.

The discriminant score (DS) is the value resulting from applying the discriminant function $d_i(\bar{x})$ to the data for a given child. The DS can be used as referral criterion:[4]

(7) if $DS > c$, refer, else not (DS criterion).

Figure 6 shows the ROC curves of the DS criterion and the DS criterion with gestational age and parental height in the discriminant model to detect children with Turner's syndrome at 5 years of age.

In other words, if the DS is less than or equal to the cut-off (c), the child is classified as reference or, if above, it is classed as having the condition. The larger the DS, the more likely the child will have the condition.

In the logistic regression analyses, the relationship between the predictor θ and the parameters (p) of the growth model is not a linear function. Instead, the logit transformation of θ is used as follows:

$$\xi := \ln(\theta/(1-\theta)) = \beta_0 + \sum_{i=1}^{p}\beta_i x_i$$

or

$$\theta = \frac{1}{1+e^{-\xi}}$$

with β_0 a constant, β_i the parameter estimate of growth parameter i, x_i growth parameter i and p the number of growth parameters. The prognostic score (PS) is the value resulting from applying the logistic regression function θ to the data for a given child. This PS can be used as referral criterion:

(8) if $PS > c$, refer, else not (PS criterion).

In other words, if the PS is less than or equal to the cut-off (c), the child is classified as reference or, if above, it is classed as having the condition. The larger the PS, the more likely the child will have the condition. Discriminant analysis is preferred when having multivariate normally distributed parameters since it has more statistical power than logistic regression.

Note that these models are fitted from a screening perspective. The parameters of the growth curve are fitted separately for the condition and the reference group and the group allocation in the

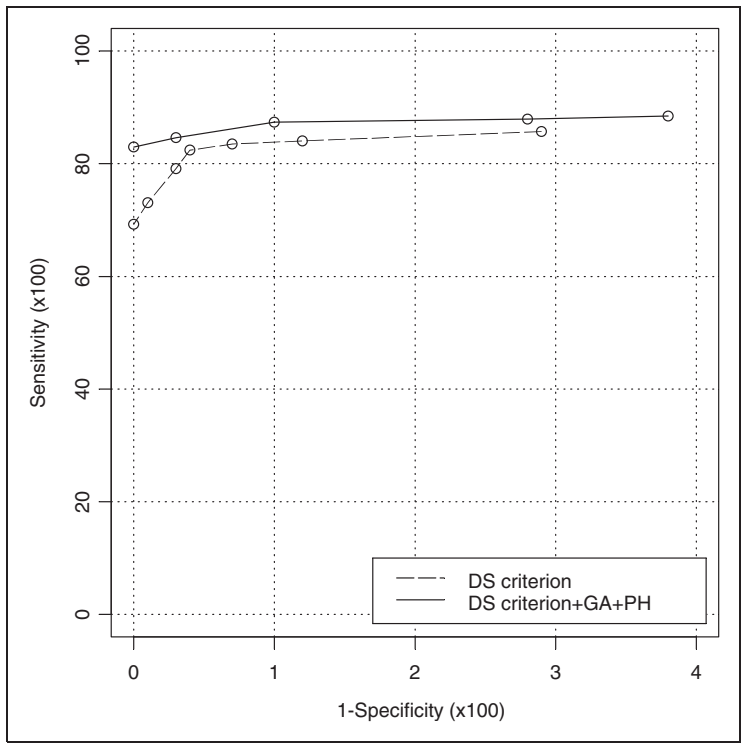

Figure 6. Receiver operating characteristic (ROC) curves of the discriminant score (DS) criterion and the DS criterion with gestational age (GA) and parental height (PH) in the model for ages 0–5 years to detect children with Turner's syndrome.

discriminant analysis was known. In an actual screening context, the information as to which group each child belongs is not present. Therefore, if we want to determine whether a new child has the condition, we have to choose in which group we estimate the growth parameters for that child. As the prevalence of most conditions is small, most children are reference children. Therefore, we most likely assume that each child is a reference child. The growth parameters of each new child should be estimated by fitting the anthropometric measures together with that of all reference children.[6] In order to obtain good estimates of the parameters of the growth curve for the children with the condition, each child have to have measurements at critical time periods: at birth and periods of rapid growth.

The individual performance of referral criteria depend on the number of measurements and the age interval. More measurements and a longer age period result in a better prediction.[4] Including parental height and gestational age in the discriminant or logistic regression model can improve the sensitivity and specificity.[6]

4 Discussion

The interest for quantitative evidence-based referral criteria is growing. Over the last 5 years several studies have been performed to obtain evidence-based referral criteria in growth monitoring.[4–10] This paper reviewed these referral criteria.

Table 1. Summary of the strengths and limitations of referral criteria in growth monitoring.

Referral criteria	Formula	Strengths	Limitations
Centile	$P(y) < c$	Easy application.	Distances between the percentiles are not equally distributed.
SDS	$z(y) < c$	Easy application. Distances between the SDS are equally distributed.	SDS less easy to interpret than centiles.
PHC	$z(y) - z(TH) < c$	Target height can also be plotted in the same chart.	Mother's height and father's height needed.
PHC SDS	$z(y) - z(TH) < c_1$ and $z(y) < c_2$,	Target height can also be plotted in the same chart. Higher validity than the PHC criterion.	See PHC
LR	$LR(y) > c$	Most powerful test for two groups.	Computer needed. Density of the diseased group must be known.
Ellipse	$P_e < c$	Powerful test for proportions.	Assumption of normality for each measure. The regression of z on z^* (and vice versa) should be linear. Correlation between z and z^* must be known.
Delta	$z(y(x_2)) - z(y(x_1)) < c_1$	Growth rate can be used to define an abnormal increase or decrease in growth.	Sensitive to measurement error. Pairwise application is cumbersome.
Extended delta	$z(y(x_2)) - z(y(x_1)) < c_1$ and $z(y(x_2)) < c_2$	Growth rate can be used to define an abnormal increase or decrease in growth. Higher validity than the delta criterion.	See Delta.
SDS velocity	$\frac{z(y(x_2)) - z(y(x_1))}{x_2 - x_1} < c_1$ and $z(y(x_2)) < c_2$	Growth rate that corrects for the time interval.	See Delta.
Conditional weight gain	$\frac{z(y(x_2)) - rz(y(x_1))}{\sqrt{1-r^2}} < c_1$ and $z(y(x_2)) < c_2$ with z weight SDS	Growth rate per year that corrects for regression to the mean.	See Delta. Correlation is needed for each age and time interval.
Parental height deflection	$z(y(x_2)) < c_1$ and $\lvert z(y(x_2) - z(TH)\rvert \geq \lvert z(y(x_1) - z(TH)\rvert$ with z length SDS	Growth rate that corrects for parental height.	See Delta and PHC.

(continued)

Table 1. Continued

Referral criteria	Formula	Strengths	Limitations
Combined weight and length deflection	$z(y(x_2)) - z(y(x_1)) < c_1$ and $z^*(y(x_2)) - z^*(y(x_1)) < c_2$ with z weight SDS and z^* length SDS	Growth rates that depend on weight and length.	See Delta.
Discriminant score	$DS > c$	Includes all available information.	Needs flexible growth models with measurements at birth and periods of rapid growth. Assumption of multivariate normally distributed growth parameters.
Prognostic score	$PS > c$	Includes all available information. No need for multivariate normally distributed growth parameters.	Needs flexible growth models with measurements at birth and periods of rapid growth.

y: measurement; x_1: age at first measurement; x_2: age at second measurement; P: centile; z or z^*: standard deviation score; PHC: parental height corrected; TH: target height; LR: likelihood ratio statistic; DS: discriminant score, PS: prognostic score; r: correlation between $z(y(x_1))$ and $z(y(x_2))$.

Table 1 shows a summary of the strengths and limitations of the criteria discussed. In general, the more advanced methods can incorporate more information and can thus provide a better trade-off between sensitivity and specificity. The downside is that application of these criteria in practice requires appropriate software to be available for those that make the referral decision.

More work needs to be done to increase the validity of detecting conditions by using referral criteria in computer-based systems. If all children have a medical file in an automated system in child health care, more information can be used to decide whether the child should be referred. Granted that appropriate software is available, computer-assisted screening saves time, eliminates error and continuously updates each record so that it is accessible for colleagues.[10]

The methods listed in Table 1 can be extended in several ways. For example, the ellipse criterion works only for one population. It is straightforward to generalize it to the two populations setting using the LR method. The densities for any possible pair of measurements can be calculated under both populations, and their ratio defines the LR. It would also be useful to formulate a general framework of which all criteria are special cases.

Of course, whatever referral criteria are used, they have to fit practice. All criteria can be implemented if a computer system is used, but some are difficult to apply by paper-and-pencil.

The DS criterion allows to incorporate most information, and thus might be a good candidate for digital systems. For paper-and-pencil settings, the SDS criterion and the PHC criterion are more convenient to use, though slightly less efficient.

The UK growth monitoring system is based on taking one measurement at the age of 5 years. Since referral rules based on parental heights and multiple measurements lead to better decisions, Fry pleaded for measuring both parents and for more than one measurement.[74] Hall et al. agreed that the issue of adjusting the school entry height measurement for parental height deserved to be revisited, but also mentioned the practical difficulties in obtaining measured heights from the parents. Hall et al. also mentioned the low coverage and inaccurate measurements in the UK and pleaded for a better implementation of the "Coventry Consensus".[75] Our hope is that the work collected in the present paper will stimulate an informed discussion of such issues.

References

1. de Onis M and Habicht JP. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. *Am J Clin Nutr* 1996; **64**: 650–658.
2. Hall DM. Growth monitoring. *Arch Dis Child* 2000; **82**: 10–15.
3. Fayter D, Nixon J, Hartley S, et al. Effectiveness and cost-effectiveness of height screening programmes during the primary school years: a systematic review. *Arch Dis Child* 2008; **93**: 278–284.
4. van Dommelen P, van Buuren S, Zandwijken GR, et al. Individual growth curve models for assessing evidence-based referral criteria in growth monitoring. *Stat Med* 2005; **24**: 3663–3674.
5. Grote FK, van Dommelen P, Oostdijk W, et al. Developing evidence-based guidelines for referral for short stature. *Arch Dis Child* 2008; **93**: 212–217.
6. van Buuren S, van Dommelen P, Zandwijken GR, et al. Towards evidence based referral criteria for growth monitoring. *Arch Dis Child* 2004; **89**: 336–341.
7. van Dommelen P, van Wouwe JP, Breuning-Boers JM, et al. Reference chart for relative weight change to detect hypernatraemic dehydration. *Arch Dis Child* 2007; **92**: 490–494.
8. van Dommelen P, Grote FK, Oostdijk W, et al. Screening rules for growth to detect celiac disease: a case-control simulation study. *BMC Pediatrics* 2008; **8**: 35.
9. van Dommelen P, Grote FK, Oostdijk W, et al. Growth monitoring to detect children with cystic fibrosis. *Horm Res* 2009; **72**: 218–224.
10. van Dommelen P. *Evidence-based referral criteria in growth monitoring*. Academic Thesis, Utrecht University, Utrecht, The Netherlands, 2008.
11. Grote FK, Oostdijk W, de Muinck Keizer-Schrama SM, et al. Growth monitoring and diagnostic work-up of short stature: an international inventorization. *J Pediatr Endocrinol Metab* 2005; **18**: 1031–1038.
12. Sorva R, Tolppanen EM and Perheentupa J. Variation of growth in length and weight of children. I. Years 1 and 2. *Acta Paediatr Scand* 1990; **79**: 490–497.
13. Sorva R, Lankinen S, Tolppanen EM, et al. Variation of growth in height and weight of children. II. After infancy. *J Acta Paediatr Scand* 1990; **79**: 498–506.
14. van Buuren S, Fredriks AM and Verkerk PH. Comment on "Consensus 'Diagnostiek kleine lichaamslengte bij kinderen" [Comment on Dutch consensus guidelines for short stature]. *Ned Tijdschr Geneeskd* 1999; **143**: 1585–1586.
15. van Buuren S, Bonnemaijer-Kerckhoffs DJ, Grote FK, et al. Many referrals under Dutch short stature guidelines. *Arch Dis Child* 2004; **89**: 351–352.

16. Gray JA. New concepts in screening. *Br J Gen Pract* 2004; **54**: 292–298.
17. World Health Organization. Physical Status: The Use and Interpretation of Anthropometry. Report of a WHO Expert Committee. Geneva: World Health Organization Tech Rep Ser 854; 1995.
18. Henry JJ. Routine growth monitoring and assessment of growth disorders. *J Pediatr Health Care* 1992; **6**(5 Pt 2): 291–301.
19. Basic human body measurements for technological design (ISO 7250:1996).
20. Saunders NR and Shouldice M. Health maintenance visits: A critical review. In: Feldman W (ed.) *Evidence-based pediatrics*. Hamilton: B.C. Decker Inc, 2000, pp.16–37.
21. Krysiak R, Gdula-Dymek A, Bednarska-Czerwińska A, et al. Growth hormone therapy in children and adults. *Pharmacol Rep* 2007; **59**: 500–516.
22. Hjerrild BE, Mortensen KH and Gravholt CH. Turner syndrome and clinical treatment. *Br Med Bull* 2008; **86**: 77–93.
23. Bondy CA. Congenital cardiovascular disease in Turner syndrome. *Congenit Heart Dis* 2008; **3**: 2–15.
24. Goddard CJ and Gillett HR. Complications of coeliac disease: are all patients at risk? *Postgrad Med J* 2006; **82**: 705–712.
25. Oude Luttikhuis H, Baur L, Jansen H, Shrewsbury VA, O'Malley C, Stolk RP and Summerbell CD. Interventions for treating obesity in children. *Cochrane Database of Systematic Reviews* 2009; (1): CD001872. DOI: 10.1002/14651858.CD001872.pub2.
26. World Health Organization. Working Group. Use and interpretation of anthropometric indicators of nutritional status. *Bull WHO* 1983; **64**: 929–941.
27. Cole TJ. The use of growth references. In: Ulijaszek SJJF and Preece MA (eds) *The Cambridge Encyclopedia of Human Growth and Development*. Cambridge, UK: Cambridge University Press, 1998, pp.80–89.
28. Garza C, Frongillo E and Dewey KG. Implications of growth patterns of breast-fed infants for growth references. *Acta Paediatr Suppl* 1994; **402**: 4–10.
29. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatr* 2006; **450**: 76–85.
30. van Buuren S and van Wouwe JP. WHO Child Growth Standards in action. *Arch Dis Child* 2008; **93**: 549–551.
31. van Dommelen P, de Gunst M, van der Vaart A, et al. Growth references for height, weight and body mass index of twins aged 0–2.5 years. *Acta Paediatr* 2008; **97**: 1099–1104.
32. Bocca-Tjeertes IF, van Buuren S, Bos AF, Kerstjens JM, Ten Vergert EM and Reijneveld SA. Growth of preterm and full-term children aged 0–4 years: integrating median growth and variability in growth charts. *J Pediatr* 2012; **161**: 460–465.
33. Van Gameren-Oosterom HB, Van Dommelen P, Oudesluys-Murphy AM, Buitendijk SE, Van Buuren S and Van Wouwe JP. Healthy growth in children with Down syndrome. *PLoS One* 2012; **7**: e31079.
34. Rongen-Westerlaken C, Corel L, van den Broeck J, et al. Reference values for height, height velocity and weight in Turner's syndrome. Swedish Study Group for GH treatment. *Acta Paediatr* 1997; **86**: 937–942.
35. Witt DR, Keena BA, Hall JG, et al. Growth curves for height in Noonan syndrome. *Clin Genet* 1986; **30**: 150–153.
36. Grauer ML, Wollmann HA, Schulz V, et al. Reference values for height and weight in Prader–Willisyndrome based on 315 patients. *Hormone Res* 1997; **48**(Suppl. 2): 54.
37. Wollman HA, Kirchner T, Enders H, et al. Growth and symptoms in Silver–Russell syndrome: review on the basis of 386 patients. *Eur J Pediatr* 1995; **154**: 958–968.
38. Marinescu RC, Mainardi PC, Collins MR, et al. Growth charts for cri-du-chat syndrome: an international collaborative study. *Am J Med Genet* 2000; **94**: 153–162.
39. Martin ND, Smith WR, Cole TJ, et al. New height, weight and head circumference charts for British children with Williams syndrome. *Arch Dis Child* 2007; **92**: 598–601.
40. Hoover-Fong JE, McGready J, Schulze KJ, et al. Weight for age charts for children with achondroplasia. *Am J Med Genet* 2007; **143**: 2227–2235.
41. Hunter AG, Hecht JT and Scott CI, Jr. Standard weight for height curves in achondroplasia. *Am J Med Genet* 1996; **62**: 255–261.
42. Horton WA, Rotter JI, Rimoin DL, et al. Standard growth curves for achondroplasia. *J Pediatr* 1978; **93**: 435–438.
43. Verbeek S, Eilers PH, Lawrence K, et al. Growth charts for children with Ellis-van Creveld syndrome. *Eur J Pediatr* 2011; **170**: 207–211.
44. CDC Growth Charts for the United States: Methods and Development. Series Report 11, No. 246, 2000, pp.201.
45. de Onis M, Wijnhoven TM and Onyango AW. Worldwide practices in child growth monitoring. *J Pediatr* 2004; **144**: 461–465.
46. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatr* 2006; **450**: 76–85.
47. Rongen-Westerlaken C, Corel L, van den Broeck J, et al. Reference values for height, height velocity and weight in Turner's syndrome. Swedish Study Group for GH treatment. *Acta Paediatr* 1997; **86**: 937–942.
48. Damen GM, Boersma B, Wit JM, et al. Catch-up growth in 60 children with celiac disease. *J Pediatr Gastroenterol Nutr* 1994; **19**: 394–400.
49. Boersma B, Houwen RH, Blum WF, et al. Catch-up growth and endocrine changes in childhood celiac disease. Endocrine changes during catch-up growth. *Horm Res* 2002; **58**(Suppl 1): 57–65.
50. Fredriks AM, van Buuren S, van Heel WJ, et al. Nationwide age references for sitting height, leg length, and sitting height/height ratio, and their diagnostic value for disproportionate growth disorders. *Arch Dis Child* 2005; **90**: 807–812.
51. Herngreen WP, Reerink JD, van Noord-Zaadstra BM, et al. The SMOCC-study: Design of a representative cohort of live-born infants in the Netherlands. *Eur J Public Health* 1992; **2**: 117–122.
52. Fredriks AM, van Buuren S, Burgmeijer RJ, et al. Continuing positive secular growth change in The Netherlands 1955–1997. *Pediatr Res* 2000; **47**: 316–323.
53. van Buuren S. *Flexible Imputation of Missing Data*. Boca Raton, FL: Chapman & Hall/CRC Press, 2012, 342 pages.
54. Cole TJ and Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; **11**: 1305–1319.
55. Rigby RA and Stasinopoulos DM. Generalized additive models for location, scale and shape, (with discussion). *Appl Statist* 2005; **54**: 507–554.
56. Galton F. Regression towards mediocrity in hereditary stature. *J Anthropological Inst* 1886; **15**: 246–263.
57. Tanner JM, Goldstein H and Whitehouse RH. Standards for children's height at age 2–9 years allowing for height of parents. *Arch Dis Child* 1970; **45**: 755–762.
58. van Dommelen P, Schönbeck Y and van Buuren S. A simple calculation of the target height. *Arch Dis Child* 2012; **97**: 182.
59. Hermanussen M and Cole TJ. The calculation of target height reconsidered. *Horm Res* 2003; **59**: 180–183.
60. Braziuniene I, Wilson TA and Lane AH. Accuracy of self-reported height measurements in parents and its effect on mid-parental target height calculation. *BMC Endocr Disord* 2007; **7**: 2.

61. Cizmecioglu F, Doherty A, Paterson WF, et al. Measured versus reported parental height. *Arch Dis Child* 2005; **90**: 941–942.
62. Engstrom JL, Paterson SA, Doherty A, et al. Accuracy of self-reported height and weight in women: an integrative review of the literature. *J Midwifery Womens Health* 2003; **48**: 338–345.
63. Kamphuis M, Obenhuijsen NH, van Dommelen P, et al. Jeugdgezondheidszorg. Guideline for preventive child health care: 'Detection and referral criteria in short stature'. *Ned Tijdschr Geneeskd* 2010; **154**: A2366.
64. van Buuren S. Improved accuracy when screening for human growth disorders by likelihood ratios. *Stat Med* 2008; **27**: 1527–1538.
65. Neyman J and Pearson ED. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika* 1928; **20A** Part I: 175–240.
66. Neyman J and Pearson ED. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika* 1928; **20A** Part II: 263–294.
67. Arnold BC, Castillo E and Sarabia JM. *Conditional specification of statistical models*. Springer: Berlin, 1999.
68. Cole TJ. Conditional reference charts to assess weight gain in British infants. *Arch Dis Child* 1995; **73**: 8–16.
69. Cole TJ. Presenting information on growth distance and conditional velocity in one chart: practical issues of chart design. *Stat med* 1998; **17**: 2697–2707.
70. Thissen D and Bock RD. Linear and nonlinear curve fitting. In: A von Eye (ed.) *Statistical methods in longitudinal research*, Volume II: Time series and categorical longitudinal data. Orlando: Academic press, 1990, pp.289–318.
71. Beath KJ. Infant growth modelling using a shape invariant model with random effects. *Stat Med* 2007; **26**: 2547–2564.
72. Jenss RM and Bayley N. A mathematical method for studying the growth of a child. *Hum Biol* 1937; **9**: 556–563.
73. Karlberg J, Engstrom I, Karlberg P, et al. Analysis of linear growth using a mathematical model. I. From birth to three years. *Acta Paediatr Scand* 1987; **76**: 478–488.
74. Fry T. If it's worth doing, let's do it! *Arch Dis Child* 2008; **93**: 267–268.
75. Hall D, Cole T, Elliman D, et al. Growth monitoring. *Arch Dis Child* 2008; **93**: 717–718.