

# Towards evidence based referral criteria for growth monitoring

S van Buuren, P van Dommelen, G R J Zandwijken, F K Grote, J M Wit, P H Verkerk

Arch Dis Child 2004;89:336–341. doi: 10.1136/adc.2003.027839

See end of article for authors' affiliations

Correspondence to:  
Dr S van Buuren, TNO  
Prevention and Health, PO  
Box 2215, 2301 CE  
Leiden, Netherlands;  
S.vanBuuren@pg.tno.nl

Accepted 16 July 2003

**Aims:** To evaluate the performance of growth monitoring in detecting diseases. Turner's syndrome (TS) is taken as the target disease.

**Methods:** Case-control simulation study. Three archetypal screening rules are applied to longitudinal growth data comparing a group with TS versus a reference group from birth to the age of 10 years. Main outcome measures were sensitivity, specificity, and median referral age.

**Results:** Clear differences in performance of the rules were found. The best rule takes parental height into account. Combining rules could improve diagnostic accuracy.

**Conclusion:** Growth monitoring is useful to screen for TS. A combined rule that takes absolute height SDS, parental height, and deflection in height velocity into account is the best way to do this. Similar research is needed for other diseases, populations, and ages, and the results should be synthesised into evidence based referral criteria.

Monitoring child growth and development is a routine part of child health care in many countries. In a typical scenario, the health care worker plots heights and weights on a reference diagram, and assesses whether the growth pattern of the child deviates from that of the reference population. If so, closer examination of the child might be needed. An important goal is to identify diseases and conditions that manifest themselves through abnormal growth. Examples include Turner's syndrome (TS), growth hormone deficiency, coeliac disease, malnutrition, as well as many rare diseases.

In contrast to its widespread use, current knowledge about the diagnostic performance of growth monitoring is incomplete.<sup>1,2</sup> Growth diagrams define the specificity of a single height measurement. The sensitivity of a single height measurement is unknown for even the most frequent diseases. Also, the sensitivity and specificity of measures involving two or more repeated observations, such as height gain, are unknown. The current state of affairs unfortunately precludes an informed discussion about referral criteria. Referral criteria have been evaluated,<sup>3–7</sup> but these studies have not prevented the appearance of widely different guidelines. For example, the recent UK guideline is based on just one universal height measurement at age 5.<sup>8</sup> In contrast, the Dutch consensus guidelines consist of multiple referral criteria covering infancy, childhood, and adolescence.<sup>9</sup> All in all, current practice differs among practitioners, and practices are not founded on evidence.

In order to make progress, we propose that all applications of growth monitoring should be judged along the conventional Wilson-Jungner criteria for screening tests.<sup>10</sup> Measures of diagnostic performance include sensitivity, specificity, and median referral time. The latter measure is essential to account for the temporal aspect of the problem.

Of all diseases that might be detected by monitoring growth, TS is one of the most frequent, occurring in 1:2500 female live births. Only 20–40% of the affected individuals, usually the ones with typical clinical features and somatic abnormalities, are diagnosed in the newborn period.<sup>11,12</sup> Diagnosis of the remaining patients is made during childhood (usually because of growth retardation) or later (because of lack of pubertal development<sup>12</sup>). This makes growth retardation the most important referral criterion in

the screening process of TS. The average adult height of untreated women is about 20 cm lower than the mean of the population.<sup>13</sup> Early detection of TS permits the clinician to counsel the family about the consequences of TS, such as an increased risk for cardiac, renal, thyroid, and auditory abnormalities associated with TS. Early detection also allows for the initiation of treatment with growth hormone, which increases final height substantially if started at a young age.<sup>14,15</sup> However, the diagnosis is often made too late,<sup>11,16</sup> so that the results of growth hormone treatment are less favourable. Some work has been done to identify girls with TS earlier using height velocity,<sup>17</sup> but the diagnostic value appeared limited. The goal of the present study is to gain insight into the diagnostic performance of a broader set of referral criteria for auxological screening for TS in the open population.

## METHOD

### Screening rules

We investigated screening rules that are suitable for application within the setting of the child health care system. A child that is "screened in" will be referred to a physician for further investigation, eventually leading to the diagnosis of TS. We formulated three archetypal screening rules: an absolute height standard deviation score rule (HSDS), a parental height corrected rule, and a deflection rule ( $\Delta$ HSDS). Based on the *absolute HSDS* rule a child is referred if HSDS is lower than some criterion value. The *parental height corrected* rule takes genetic height potential into account by comparing the HSDS of the child to its target height SDS. The target height (TH) is the expected adult height given the heights of the biological parents and corrected for secular trend. For Dutch girls, the relevant formulas are  $TH = (\text{maternal height} + \text{paternal height} - 13)/2 + 4.5$  and  $THSDS = (TH - 170.6)/6.5$ .<sup>18</sup> The *deflection* rule signals whether an abnormal deflection in height occurs in terms of a change in HSDS per year. Table 1 gives the precise definition of each rule, the description of the free parameters, and default values of the parameters as used in the Dutch guidelines.<sup>9</sup>

**Abbreviations:** GH, growth hormone; HSDS, height standard deviation score; TH, target height; TS, Turner's syndrome

**Table 1** Three archetypal screening rules for growth monitoring with their definition, scenario parameters, interpretation, default parameter values according to the Dutch consensus guidelines,<sup>7</sup> and the parameter values used in the simulation

Screening rule	Definition	Parameter	Interpretation	Default value*	Simulation values
Absolute height SDS	For ages 0 to p years, refer if $SDS < a$ For ages p to 10 years, refer if $SDS < b$	a b	SDS referral level before age p SDS referral level after age p	-2.5 -2.5	-1.5, -2, -2.5, -3, -3.5, -4 -1, -1.5, -2, -2.5, -3
Parental height corrected	For ages q to 10 years, refer if: $SDS < c$ , AND $SDS - THSDS < d$	p c d q	Age (in years) at which the referral level changes  SDS cut off level below which SDS must lie Difference between target height SDS and SDS Age (in years) after which the rule is effective	Unspecified -1.3 -1.3 0, 3†	1, 2, 3  -1, -1.3, -1.5, -2, -2.5 -1, -1.3, -1.5, -2, -2.5 1, 2, 3
Deflection	For any pair $SDS_1$ and $SDS_2$ measured at ages $X_1$ and $X_2$ (in years), refer if: $r \leq X_1 < X_2 < 10$ , AND $X_2 - X_1 \geq e$ , AND $SDS_2 < f$ , AND $(SDS_2 - SDS_1) / (X_2 - X_1) < g$	e f g r	Minimal interval (in years) between $X_1$ and $X_2$ SDS cut off level below which $SDS_2$ must lie Height velocity change in SDS per year Age (in years) after which the rule is effective	1‡ Unspecified -0.25 0, 3†	1, 2, 3 10, 0, -1, -2, -2.5 -0.20, -0.25, -0.33, -0.50 3

\*According to Dutch consensus guidelines.

†The Dutch consensus guidelines are ambiguous.

‡The Dutch consensus guidelines require that three measurements should have been taken, each at least one half a year apart.

Each screening rule was implemented in a computer program written in S-Plus,<sup>19</sup> and each rule was applied to longitudinal height data of children with and without TS. For each screening rule, we computed the sensitivity, specificity, and median referral age for specific scenarios. A scenario is a combination of parameters. We defined scenarios by all possible parameter combinations. We first studied the properties of each screening rule separately. Given these results, we defined scenarios that combined the most promising elements of the separate rules, and computed the outcomes for combined scenarios.

**Material**

Longitudinal height curves from 777 girls with TS were collected from three sources. The National Registry of Growth Hormone Treatment in Children of the Dutch Growth Foundation contains data of all children in the Netherlands receiving growth hormone (GH) treatment. From this registry, 316 girls with TS, born between 1968 and 1996 were selected. In addition, data from 87 girls with TS, born between 1973 and 1988 from the Sophia Children’s Hospital and the data of 374 Dutch girls described by Rongen and colleagues<sup>13</sup> were used. The first two sources contain data of girls that were treated with GH and other growth promoting treatment. For this analysis we used only height measurements before treatment. Karyotype, date of diagnosis TS, the presence of congenital anomalies and/or dysmorphic features and parental height were collected when available. The average numbers of measurements per year per child during the first 10 years were 2.2, 0.7, 0.6, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, and 0.5, respectively.

A reference sample of longitudinal height data was retrospectively obtained for a cohort of all girls (n = 489) born in 1989 and 1990 in the municipality of Landgraaf, located in the south of the Netherlands. Data were collected from the records of the local child health care centre. These are routinely collected data, and they thus include all measurement errors that are being made in practice. The modal number of observations per girl was 17. Data were collected in 2001, so the oldest girls were about 11 years. The average numbers of visits per year per child during the first 10 years were 8.1, 2.2, 0.8, 0.8, 0.7, 0.7, 0.2, 0.1, 0.1, and 0.8. Table 2 contains additional information about the samples.

**Statistical analysis**

HSDS was calculated with respect to the Dutch height reference data.<sup>20</sup> Parental heights were frequently missing (55% of the Turner group, 58% in the reference group). Deleting incomplete records would not only be wasteful, but would also lead to a selective subsample. Mean HSDS of girls with TS was -3.24 for the subsample with missing parental heights, compared to -2.53 for the subsample with known parental heights, but no such differences were found in the reference group. We imputed these data under the assumption that the data are missing at random<sup>21</sup> using mice.<sup>22</sup> The method created multivariate imputations by applying sequential linear regressions, where each incomplete variable was imputed conditional on all other variables in an iterative fashion. The imputation model consisted of the last known HSDS, weight SDS, weight/height SDS, BMI SDS, age, and the height of the other parent. The number of iterations was set to 15. Predictive mean matching was used to create parental heights imputations. The imputation method possesses important properties: it includes parameter uncertainty, preserves the multivariate structure in the data, and has good coverage properties.<sup>23</sup> Figure 1 plots father’s height against mother’s height separately for the real and artificial data. It shows that the distribution is similar in both groups.

We assumed that a child is referred the first time the growth pattern meets the criteria of a given screening rule.

**Table 2** Summary statistics for the Turner and reference samples

	Turner		Reference	
	Count/mean	SD	Count/mean	SD
Total sample size (n)	777		489	
Total number of measurements	9660		7319	
Mean number of measurements per girl	12.4		15.0	
Mean height for age SDS (ages 0–10 y)	−2.44	1.13	−0.31	1.05
Mean weight for age SDS (ages 0–10 y)	−1.74	1.28	−0.12	1.05
Mean weight for height SDS (ages 0–10 y)	−0.06	1.29	0.12	1.04
Mean BMI for age SDS (ages 0–10 y)	−0.19	1.27	0.11	1.04
Height of both father and mother known	357		203	
Height of only one parent known	3		10	
Height of both parents unknown	417		276	
Mean father's height (cm)	179.4	7.50	178.4	7.57
Mean mother's height (cm)	166.4	6.29	166.7	7.38
Target height (cm)	169.7	5.89	170.6	5.70
Target height SDS	0.06	0.82	0.01	0.88
Gestational age (weeks)	38.9	2.07	39.7	1.61
Dysmorphic features (%) (n = 145)				
Cubitus valgus	31			
Large inter-nipple distance	29			
Low hair implantation	21			
Webbed neck	19			
Karyotype (%) (n = 327)				
45,X	62			
46,X,iX or 46,X,idi(X)	5			
45,X and 46,XX	5			
45,X and (46,X,iX or 46,X,idi(X))	12			
Other	16			
Median age of diagnosis of TS (years) (n = 46)				
45,X (n = 27)	6.9	4.94		
Other (n = 19)	10.4	4.81		

Multiple referrals by different rules were treated depending on the type of scenario under investigation. As long as we dealt with separate rules, the same child could be referred according to each rule—that is, as if the screening rules operated in isolation, but in any combined scenario, multiple referrals were counted as one. The screening age of children that were not referred before the age of 10 years was taken as 10 years. TS girls with a referral age of 10 under a given scenario are missed, so the proportion of such girls is the false negative rate ( $1 - \text{sensitivity}$ ). The age of 10 years was chosen because treatment of TS, if indicated, could best be started before that age.

Finally, we synthesised our results by fitting linear regression models to the main outcome variables. These models can be used to predict sensitivity, specificity, and median referral age (MRA) in intermediate cases that were not part of the simulation design.

## RESULTS

### Sensitivity and specificity

Figure 2A is the ROC plot of scenarios under the absolute HSDS rule. Only scenarios with a true positive rate (sensitivity) of at least 40%, a false positive rate ( $1 - \text{specificity}$ ) of at most 15%, and with cut off age  $p = 3$  are plotted. Under the default scenario “(−2.5, −2.5)” children are referred that have an HSDS  $< -2.5$  ( $a = -2.5$ ,  $b = -2.5$ ,  $p = 3$ ). Scenario “(−2.5, −2.5)” has a sensitivity of 70.2% and a specificity of 93.1%. Scenarios (−3, −2), (−3.5, −2), and (−4, −2) have better sensitivity and specificity for detecting TS. Specificity is, however, still on the low side for screening purposes (95–97%), thus these scenarios might lead to substantial numbers of false positives. Scenarios (−3.5, −2.5) and (−4, −2.5) cut down the number of false positives, at the expense of a loss of sensitivity. The influence of  $p$  on sensitivity and specificity was limited.

Performance of the parental height corrected rule was generally better (fig 2B). The current Dutch guideline (−1.3, −1.3) pairs a high sensitivity of 93.5% with a specificity of 95.9%. Rules using more stringent cut off points reduce the number of false positive referrals at the expense of sensitivity. Examples of interesting scenarios are (−1.5, −1.5), (−2, −2), and (−2.5, −2.5). Note that for these cases  $c = d$ . The difference with the absolute SDS rule is the extra requirement that  $\text{THSDS} > (c - d)$ —that is,  $\text{THSDS} > 0$  or taller than average parents.

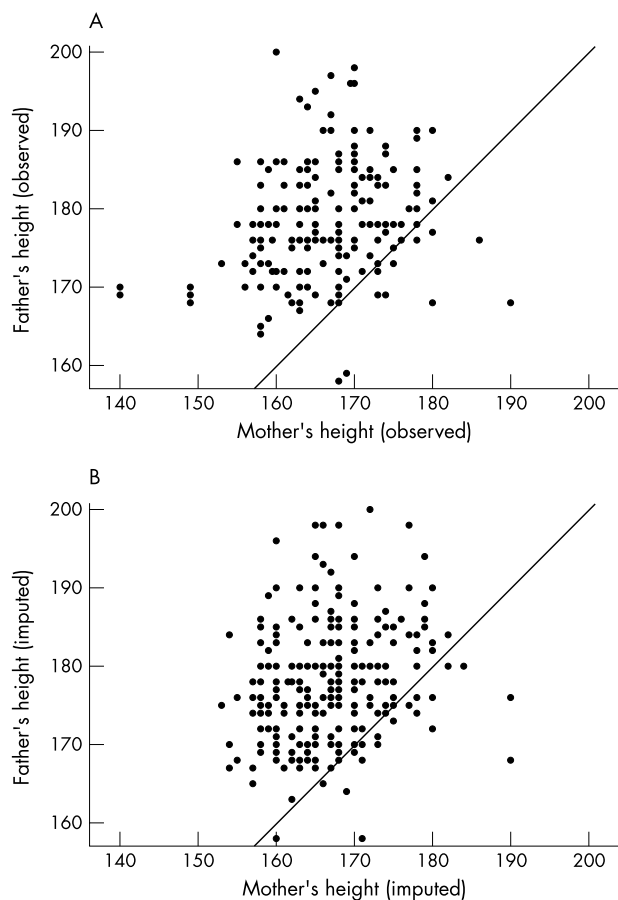
Screening based on the deflection of the growth curve has low sensitivity for rules with a specificity of at least 85% (cf fig 2C). Though not very sensitive, some deflection rules are highly specific. For example, the rule with  $e = 3$ ,  $f = -2$ , and  $g = -0.25$  (not in fig 2C) pairs a sensitivity of 23% with the maximal specificity of 100. It can be efficient to use such rules in conjunction with more sensitive rules.

### Median referral age

Median referral age in the Turner group generally did not exceed 6 years under the absolute HSDS or the parental height corrected rule (fig 3). Median referral age tends to be lower for higher sensitivity and lower specificity. Thus, more cases imply younger cases. This is especially true under scenarios that correct for parental height. The absolute HSDS rule provides the fastest detection of TS, primarily due to the fact that this is the only rule that takes measurements during infancy into account. Earlier detection of TS is possible at the expense of specificity, especially if done through the parental height rule.

### Predicting sensitivity, specificity, and median referral age

Table 3 contains a synthesis of the results. It gives estimated coefficients of the linear regression for all outcomes. As the proportion of explained variance is generally high, the



**Figure 1** Father's height plotted against mother's height in the reference sample. (A) Observed data from group in which both parental heights are known (n=203). (B) Imputed (artificial) data for cases where at least one parental height is missing (n=286). The reference line indicates the location of equal parental heights.

regression equations can be used to generate fairly accurate predictions for intermediate scenarios not listed in the table. As an example, the estimated sensitivity for scenario  $(-3, -2)$  for the absolute HSDS rule is equal to  $97.5 - 3.62a - 0.60b - 6.04ab = 73.3\%$ . The observed values vary between 74.2% (for  $p = 3$ ) and 76.5% (for  $p = 1$ ). In this case, the differences between these observed and predicted values fall within one standard deviation of the residual variation (4.97%).

**Combining rules**

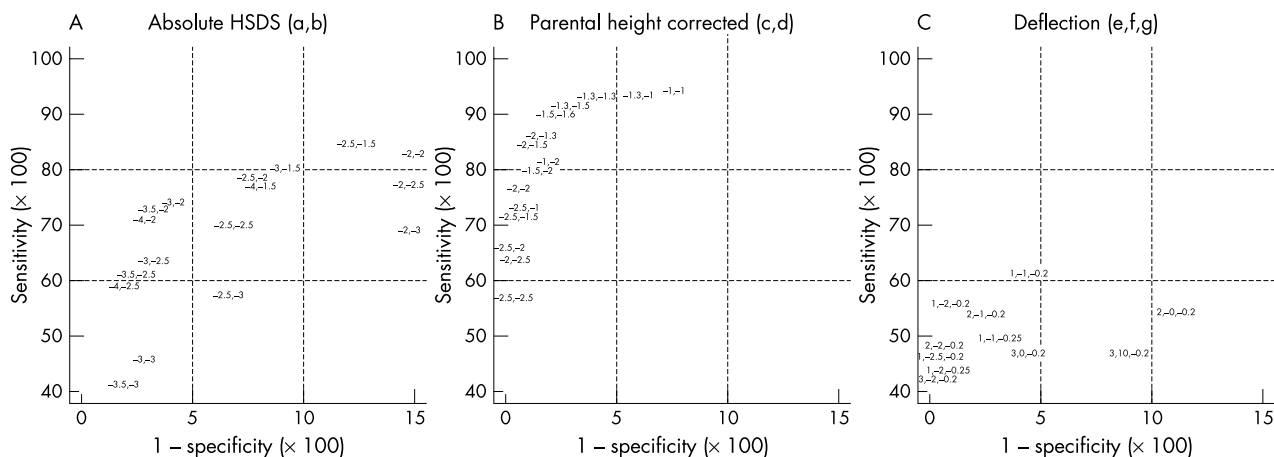
A child will be referred if he or she meets any of the rules. Sensitivity of a combined rule will be higher than that of its components, while its specificity will be lower.<sup>24</sup> Thus in order to create highly specific combinations, the component rules must have high specificity to start with.

Table 4 shows the diagnostic properties of two combinations. Combining the parental height corrected rule  $(-2, -2)$  with the absolute height corrected rule  $(-3.5, -3)$  increases sensitivity from 76.9% to 82.4%, decreases specificity from 99.4% to 97.5%, and lowers median referral age to 4.7 years. Observe that this combined rule is inferior to the parental height corrected rule  $(-2, -1.5)$  in terms of sensitivity and specificity. The story is different for the combination of the absolute rule with the deflection rule  $(3, -2, -0.25)$ , which refers children with a HSDS below  $-2$  and a deflection of at least 0.25 SDS per year during at least three years. While this rule detects only 23% of the TS group, there is not a single child in the reference group with this growth pattern. The rule picks up a few new cases. Sensitivity increases from 76.9% to 79.2%, whereas specificity remains at 99.4%. This combined rule is better than comparable parental height corrected rules.

**DISCUSSION**

Growth monitoring is important for detecting TS, but until now no evidence has been available about the diagnostic quality of possible screening procedures. We estimated sensitivity, specificity, and median referral age of TS for three screening rules, and for combinations of these rules. We found that these rules had different performance in discriminating TS. Rules that correct for parental height could identify TS better than rules using the absolute HSDS or rules based on the deflection of growth curves. Combining rules improved performance in particular cases.

The children in our control sample live in the southern part of the Netherlands, and are shorter on average ( $-0.31$  HSDS) than the Dutch reference population. This means that the specificity for the Dutch reference population might be more favourable than estimates based on the shorter population. The equations in table 3 can be used to estimate the size of the effect. For example, setting  $a = -2$ ,  $b = -2.5$ , and  $p = 1$  yields a predicted specificity of 82.5%. Had the group been  $-0.31$  shorter, then substituting  $a = -1.69$ ,  $b = -2.19$ , and  $p = 1$  predicts a specificity of 78.1% for that group. So the actual specificity for a group that is 0.31 HSDS shorter is here 4.4% lower. In order to eliminate such biases, we added 0.31 HSDS to the measurements of the reference group. The existence of regional height differences implies that the



**Figure 2** ROC plot of different scenarios under three archetypal rules. Each scenario is labeled by its parameter values according to table 1. For example, the label “ $(-3, -2)$ ” in the left plot indicates the scenario with  $a = -3$  and  $b = -2$ . Only scenarios with  $p = q = r = 3$  (cf table 1) are plotted.

**Table 3** Regression equations for predicting sensitivity, specificity, and median referral age of the absolute HSDS (sds), parental height corrected (phc), and deflection (def) screening rules, the residual standard error, and the proportion of explained variance ( $r^2$ )

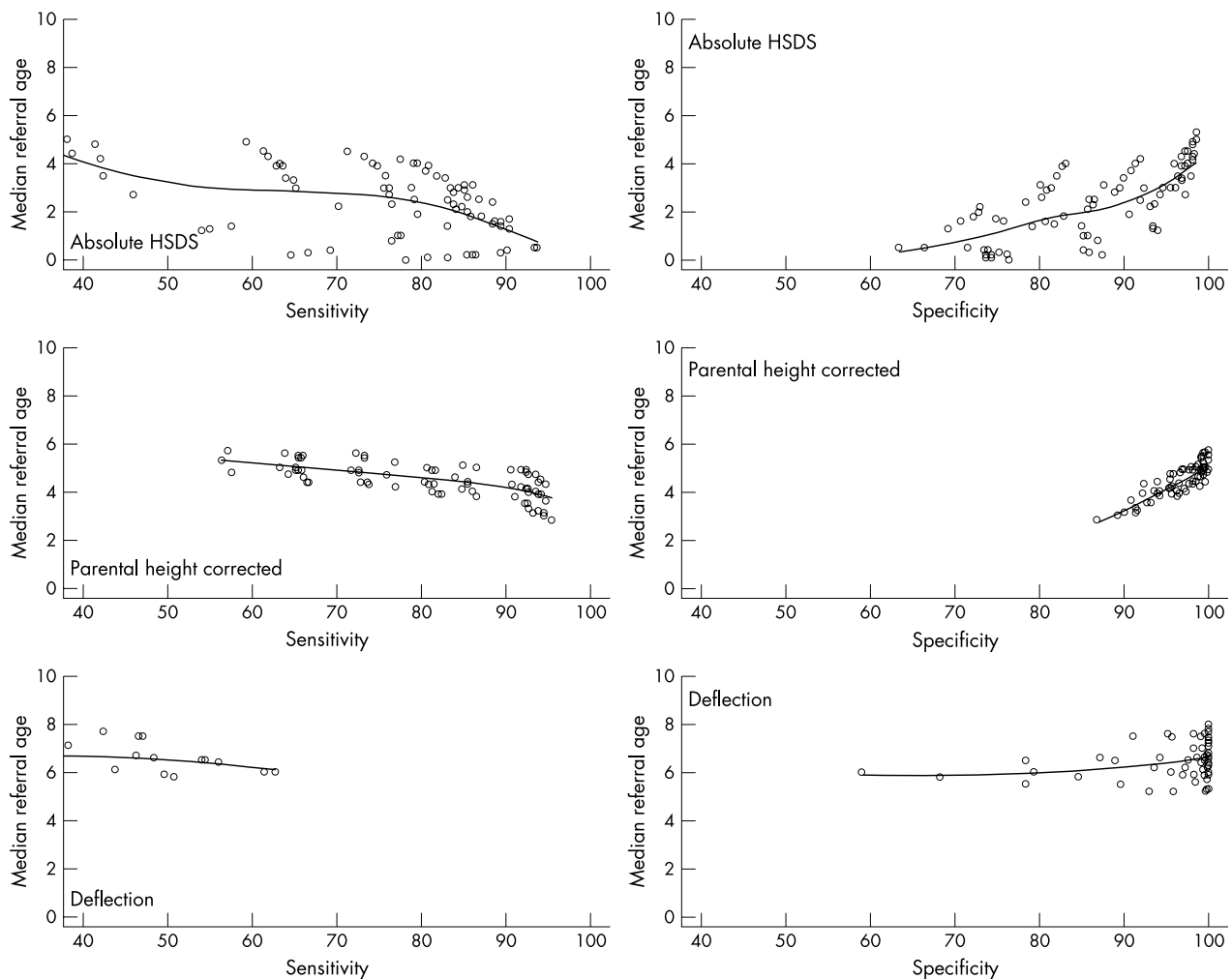
Rule	Outcome	Predictive equation	Resid. SE	$r^2$
sds	Sensitivity	$97.5-3.62a-0.60b-6.04ab$	4.97	0.89
	Specificity	$58.4-3.55a-2.32b+1.20p+2.00ab$	4.56	0.79
	Median referral age	$1.45+0.33a+1.13b-0.04p+0.59ab-0.32ap+0.20bp$	0.38	0.94
phc	Sensitivity	$145+21.6c+26.8d+5.95cd$	3.66	0.91
	Specificity	$67.0-11.0c-11.3d+4.82q-3.41cd+1.06cq-1.15dq$	0.68	0.95
	Median referral age	$-1.09-1.76c-1.92d+1.26q-0.62cd+0.26cq+0.32dq+0.11cdq$	0.10	0.98
def	Sensitivity	$89.9-8.59e+0.40f+145g$	7.46	0.86
	Specificity	$82.9+3.48e-1.05f-21.1g$	5.47	0.57
	Median referral age	$5.63+0.82e-0.02f+2.28g$	0.28	0.87

actual false-positive rates can vary across the country. The parental height corrected rule and the deflection rule are less sensitive to such differences. Using the equations in table 3, it is straightforward to compute the effect of regional differences on sensitivity and specificity. Region specific screening rules can be created if the effect is substantial. Similar considerations apply to ethnic minority groups.

Diagnosis of TS is often unnecessarily delayed. Excluding the 20–40% of the patients identified in infancy, the median age of diagnosis is somewhere between 10 and 12 years.<sup>11 16</sup> Including 30% of the early cases into the calculation would lower the median age of diagnoses to the range of 7–8.4

years. By the time of diagnosis, patients were extremely short (mean  $-3.0$  HSDS). We found that the median referral age of most screening rules studied here is between 4 and 6 years. Some rules even identify 50% or more of the cases within the first year. The current policy in the Netherlands is that GH treatment in girls with TS is applied if HSDS  $<-1.5$  and if the child is older than 6 years, but preferably younger than 9. Before the age of 6 years treatment is only started if HSDS  $<-2.5$ . Our results suggest that systematic growth monitoring is able to find the large majority of cases in time.

The occurrence of missing parental heights complicated the analysis. It is inappropriate to simply ignore the records with



**Figure 3** Median referral age of girls with TS as a function of sensitivity and specificity under each rule. Every dot corresponds to a scenario.

**Table 4** Combining rules using a high specificity strategy

Row	Rule	Scenario parameters						Sensitivity (*100)	Specificity (*100)	MRA	
		a	b	c	d	e	f				g
1	phc			-2.0	-2.0			76.9	99.4	5.2	
2	sds	-3.5	-3.0					41.4	98.1	4.8	
3	phc-sds	-3.5	-3.0	-2.0	-2.0			82.4	97.5	4.7	
4	phc			-2.0	-2.0			76.9	99.4	5.2	
5	def					3	-2.0	-0.25	100.0	7.7	
6	phc-def			-2.0	-2.0	3	-2.0	-0.25	79.2	99.4	5.3
7	phc			-2.0	-1.5			84.9	98.8	5.1	

Rows 1–3 list a parental height corrected (phc), an absolute sds (sds), and their combined (phc-sds) rule. Rows 4–6 list a parental height corrected (phc), a deflection (def), and their combined (phc-def) rule. Row 7 is a single parental height corrected rule that is better than row 3 but not preferable to row 6. MRA, median referral age.

incomplete parental heights because the shorter TS girls drop out more frequently. This leads to sensitivity estimates that are too low. The effect is substantial. For example, using just the complete cases in scenario  $c = d = -1.3$  and  $q = 3$  results in a sensitivity estimate of 88.7%, compared to 93.5% based on the imputed sample. As it would be unfair to exclude the incomplete cases only for the parental height rule, sensitivity estimates for other rules would also be affected. Imputation yields unbiased estimates for the TS group as a whole. The precision of these estimates is lower than found in the hypothetical case in which we would have had complete data, but it is higher than obtained in the inappropriate complete-case analysis just discussed.

Our results enable informed decisions about specific choices in screening rules for identifying TS. Although growth charts are also used to detect other anomalies, like growth hormone deficiency or coeliac disease, growth monitoring should at least be able to detect TS. If monitoring cannot pick up TS, then it almost certainly will fail in more complicated cases where the effects on growth are less pronounced. It is likely that repeating our study for other diseases will lead to different estimates for sensitivity and specificity. Additional complexities will surface, for example, the lack of a gold standard for diagnosis of growth hormone deficiency. However, such studies would probably not lead to a different ranking among the three rules. We expect that rules that take parental height into account are generally preferable to rules that do not.

The findings appear to be only partially in harmony with published guidelines and proposals.<sup>7–9</sup> As anticipated,<sup>17–25</sup> we found that centile crossing has low sensitivity and specificity, and in this sense, the Dutch guidelines may need re-evaluation. Marked differences occur with respect to the correction for parental height. Hall and Elliman<sup>8</sup> dismissed a correction for parental target height on practical grounds, whereas we found that it represents a substantial improvement, in line with earlier observations by Massa and Vanderschueren-Lodeweyckx.<sup>16</sup>

We conclude that growth monitoring is useful to screen for TS. The parental height corrected rule will refer 60–77% of the girls with TS before the age of 10 at tolerable levels of false positives—that is, at a maximum of 1%. We recommend the use of the combined rule “phc-def” listed in table 4. This rule refers children older than age three if HSDS is below -2 and if either HSDS is more than 2 SD below the target HSDS, or HSDS shows a deflection of 0.25 SDS per year or more during a period of at least three years. This rule picks up almost 80% of the girls with TS, while it refers only 0.6% of the non-TS population. We also recommend that similar research should be done for other diseases, populations, and ages. The results should be synthesised into general evidence based referral criteria.

#### Authors' affiliations

**S van Buuren, P van Dommelen**, Dept of Statistics, TNO Prevention and Health, Leiden, Netherlands  
**G R J Zandwijken**, Dutch Growth Foundation, Rotterdam, Netherlands  
**F K Grote, J M Wit**, Dept of Paediatrics, Leiden University Medical Center, Leiden, Netherlands  
**P H Verkerk**, Dept of Child Health, TNO Prevention and Health, Leiden, Netherlands

#### REFERENCES

- Panpanich R, Garner P. Growth monitoring in children. *Cochrane Database Syst Rev* 2000;**2**:CD001443.
- Garner P, Panpanich R, Logan S. Is routine growth monitoring effective? A systematic review of trials. *Arch Dis Child* 2000;**82**:197–201.
- Hindmarsh PC. Monitoring children's growth: abnormal growth should also be defined by the crossing of height centiles [letter]. *BMJ* 1996;**312**:122.
- Paterson BA. A review of growth monitoring in Aboriginal school children in Australia. *BMJ com*, electronic letter, 5 March 1999.
- Mulligan J, Voss LD, McCaughey ES, et al. Growth monitoring: testing the new guidelines. *Arch Dis Child* 1998;**79**:318–22.
- Voss LD. Changing practice in growth monitoring. *BMJ* 1999;**318**:344–5.
- Hall DMB. Growth monitoring. *Arch Dis Child* 2000;**82**:10–15.
- Hall DMD, Elliman D. *Health for all children*, 4th edn. Oxford: Oxford University Press, 2003.
- De Muinck Keizer-Schrama SMPF. Consensus 'Diagnostiek kleine lichaamslengte bij kinderen' [Dutch consensus guidelines for short stature]. *Ned Tijdschr Geneesk* 1998;**142**:2519–25.
- Wilson JM, Jungner G. *Principles and practice of screening for disease*. Public Health Paper Number 34. Geneva: World Health Organisation, 1968.
- Sävendahl L, Davenport ML. Delayed diagnoses of Turner's syndrome: proposed guidelines for change. *J Pediatr* 2000;**137**:455–9.
- Gicquel C, Gaston V, Cabrol S, et al. Assessment of Turner's syndrome by molecular analysis of the X chromosome in growth retarded girls. *J Clin Endocrinol Metab* 1998;**83**:1472–6.
- Rongen-Westelaken 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**:937–42.
- Sas TC, de Muinck Keizer-Schrama SM, Stijnen T, et al. Normalization of height in girls with Turner syndrome after long-term growth hormone treatment: results of a randomized dose-response trial. *J Clin Endocrinol Metab* 1999;**84**:4607–12.
- Reiter EO, Bletthen SL, Baptista J, et al. Early initiation of growth hormone treatment allows age-appropriate estrogen use in Turner's syndrome. *J Clin Endocrinol Metab* 2001;**86**:1936–41.
- Massa G, Vanderschueren-Lodeweyckx M. Age and height at diagnosis in Turner syndrome: influence of parental height. *Pediatrics* 1991;**88**:1148–52.
- Van den Broeck J, Hokken-Koelega A, Wit JM. Validity of height velocity as a diagnostic criterion for idiopathic growth hormone deficiency and Turner syndrome. *Horm Res* 1999;**51**:68–73.
- Fredriks AM, van Buuren S, Burgmeijer RJF, et al. *Groeiendiagrammen (2<sup>e</sup>, herziene en vermeerderde druk)* [Growth diagrams, second edition]. Houten: Bohn Stafleu van Loghum, 2002.
- Insightful Corp. *S-Plus 2000 user's guide*. Seattle, WA: Insightful, 2000.
- Fredriks AM, van Buuren S, Burgmeijer RJ, et al. Continuing positive secular growth change in The Netherlands 1955–1997. *Pediatr Res* 2000;**47**:316–23.
- Little RJA, Rubin DB. *Statistical analysis with missing data*, 2nd edn. Wiley: New York, 2002.
- Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;**18**:681–94.
- Brand JPL, van Buuren S, Groothuis-Oudshoorn CGM, et al. A toolkit in SAS for the evaluation of multiple imputation methods. *Stat Neerl* 2003;**57**:36–45.
- Macaskill P, Walter SD, Irwig L, et al. Assessing the gain in diagnostic performance when combining two diagnostic tests. *Stat Med* 2002;**21**:2527–46.
- Van Buuren S, Fredriks AM, Verkerk PH. Comment on "Consensus 'Diagnostiek kleine lichaamslengte bij kinderen' [Comment on Dutch consensus guidelines for short stature]. *Ned Tijdschr Geneesk* 1999;**143**:1585–6.