Variation in Methods of Predicting Adult Height for Children With Idiopathic Short Stature
Lisa Swartz Topor, Henry A. Feldman, Howard Bauchner and Laurie E. Cohen
*Pediatrics* 2010;126;938; originally published online October 25, 2010;
DOI: 10.1542/peds.2009-3649

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://pediatrics.aappublications.org/content/126/5/938.full.html
Variation in Methods of Predicting Adult Height for Children With Idiopathic Short Stature

WHAT’S KNOWN ON THIS SUBJECT: Predicted adult height often is used in evaluations of children with short stature, because treatment may not be indicated for children whose growth is consistent with attainment of a height above the 1.2nd percentile.

WHAT THIS STUDY ADDS: We demonstrated wide variation in adult heights predicted with 3 commonly used height prediction algorithms. Because the 3 algorithms led to profoundly discrepant predicted heights, future studies are necessary to determine which is most accurate for children with short stature.

abstract

OBJECTIVE: Recombinant human growth hormone (GH) is approved for treatment of children with idiopathic short stature, and endocrinologists often depend on algorithms to predict adult height. Because algorithm performance often is included in treatment decisions, we sought to evaluate agreement among height prediction formulas.

METHODS: We identified 3 commonly used algorithms for height prediction, the Bayley-Pinneau, Roche-Wainer-Thissen, and Khamis-Roche methods. We constructed simulated samples of children with typical distributions of ages, heights, weights, bone ages, and parental heights seen in patients with idiopathic short stature, and we applied the algorithms to the simulated sample to determine whether predicted adult height was <160 cm for boys or <150 cm for girls (<1.2nd height percentiles for adults).

RESULTS: We found substantial disagreement among algorithms in the proportions of simulated cases with predicted adult heights of <1.2nd percentile, a cutoff value that may influence GH treatment decisions. With the Bayley-Pinneau formula, 43% of boys and 81% of girls had predicted adult heights below this threshold; with the Khamis-Roche method, only 3% of boys and 0.2% of girls had predicted heights of <1.2nd percentile. Roche-Wainer-Thissen predictions were between those values. Overall agreement of the methods was poor ($\kappa = 0.21$) for boys and negative for girls.

CONCLUSIONS: Wide variation exists among formulas used to predict adult heights. Because these algorithms may be used in decisions regarding whether to initiate GH treatment and assessment of the efficacy of GH in research trials, it is important for parents, pediatricians, and investigators to recognize the considerable variation involved in height predictions. Pediatrics 2010;126:938–944
Recombinant human growth hormone (GH) was approved by the US Food and Drug Administration (FDA) in 2003 for treatment of children with idiopathic short stature (ISS), defined by the FDA as height before treatment 2.25 SDs below the mean for age (1.2nd percentile) without evidence of underlying disease or GH deficiency. Although predicted adult height is not part of the FDA criteria for using GH to treat children with ISS, the FDA criteria do include a statement regarding “a growth rate that is unlikely to attain an adult height within the normal range.” In addition, the FDA identified the 1.2nd percentile cutoff values for height in adults as 63 inches (160 cm) for men and 59 inches (150 cm) for women.1

On the basis of the ISS indication for GH treatment, ~400,000 children in the United States now qualify for GH therapy.2 The approval of GH for treatment of ISS was based on 2 clinical trials, each with a small number of subjects.3,4 One study (with 33 subjects monitored to adult height) was a randomized, placebo-controlled trial, and the other (with 50 subjects monitored to adult height) was an open-label, dose-response trial that compared adult height with predicted adult height. Both trials used the Bayley-Pinneau (BP) method of height prediction. Additional studies of GH use for children with ISS have been limited by sample size. A Cochrane review included 10 randomized controlled trials of GH treatment for a total of 741 children with ISS. The average duration of treatment and height gain varied, and combined results showed height gains over 1 year of 0.7 SD.5 The review did not compare near-adult heights with predicted adult heights.

The approval of GH treatment for children with ISS has substantial potential costs for the US health care system. A 2006 article by Lee et al6 examined the 2 studies used by the FDA to support approval for ISS to estimate the cost-effectiveness of GH therapy for children with ISS. The cost-effectiveness was estimated at $52,000 per inch and, given an average incremental height gain of 1.9 inches over 5 years, the incremental cost per child was nearly $100,000.6 At this average cost per child, the potential cost of treating all eligible children would be approximately $40 billion dollars.6 These costs have been recognized by insurers, because many deny coverage for GH for treatment of ISS.7,8

Predicted adult height often is used to evaluate children with short stature, because treatment with GH may not be indicated for a child whose growth is consistent with attainment of a height above the 1.2nd percentile, according to the FDA indication. For example, for a child with constitutional delay, the current height SD score often is considerably lower than predicted near-adult height SD scores.9 Therefore, to assess the appropriateness of treatment with GH for children with ISS, endocrinologists often use an algorithm to predict whether the child’s predicted adult height will be <1.2nd percentile, corresponding to <160 cm (<63 inches) for adult men and <150 cm (<59 inches) for adult women. A 2006 survey of pediatric endocrinologists identified a predicted target height of <5th percentile as a consistent factor in the decision to treat patients with GH.10 Because the performance of such algorithms often is crucial for decision-making regarding treatment with GH, we sought to evaluate agreement among formulas used to predict adult height.

METHODS
Simulated Sample
We constructed a simulated sample of 1000 boys and 1000 girls with the typical distributions of ages, heights, weights, bone ages, and parental heights of patients seen for evaluation of short stature, including appropriate upper and lower limits and correlations among the variables. The variable distributions were determined through informal review of data from our institution’s clinicians and on the basis of a study by Grimberg et al11 that compared gender differences among referrals for evaluation of decreased growth velocity at a tertiary care center. The study by Grimberg et al11 identified referral differences according to gender, height (on average, girls had lower height SD scores than boys), and mid-parental height.

SAS software (SAS Institute, Cary, NC) was used for all analyses. Two triangular distributions of random ages were generated, spanning 6 to 16 years with a peak at 13 years for boys and 6 to 14 years with a peak at 11 years for girls. An initial sample of 2500 boys and 2500 girls was generated to allow for truncations and deletions, as detailed below, to produce the final planned sample of 1000 boys and 1000 girls. Multivariate normal z scores (mean: 0; SD: 1) were generated for each child’s height, weight, bone age and parental heights. We were unable to identify correlations for these parameters in the literature and relied on observations for our data. We programmed the random-number generator to impose correlations of 0.90 between the child’s height and weight z scores, 0.60 between the parents’ heights, 0.50 between the child’s height and each parent’s height, 0.15 between the child’s weight and bone age, 0.15 between the child’s bone age and each parent’s height, and 0 among the other z scores. In addition, to test the sensitivity of our findings to the estimated correlation values, we abolished the relatively strong assumed correlations entirely and repeated the full simulation.

The initial sample was shifted and truncated to reflect typical children...
presenting for evaluation of short stature. For girls, the height distribution was shifted by $-0.75$ and the weight distribution by $-0.25$ z scores. Upper limits of $-1$ for height, 1 for weight, and 0.25 for parental height were imposed. The $z$ score distributions were further truncated to eliminate extreme values, with deletion of cases with weight below $-3$, height below $-4$ for boys or $-4.5$ for girls, parental height below $-2.5$, or difference between chronologic age and bone age beyond $\pm 2$.

The height and weight $z$ scores were converted to centimeters and kilograms by using gender- and age-specific normative values for US children. We assumed the child’s bone age differed from chronological age by $0 \pm 1$ y mean $\pm$ SD, and thus calculated bone age in years by adding the bone age $z$ score to age in years. Parental height $z$ scores were converted to centimeters by using the Centers for Disease Control and Prevention normative values for maximal age (20 years) and were averaged to generate mid–parental height values. After the deletions described above, 2680 cases remained in the randomly simulated sample, from which the first 1000 boys and 1000 girls were selected for analysis.

**Prediction of Adult Height**

We identified 3 commonly used algorithms for predicting adult height, from the literature and from interviews with board-certified pediatric endocrinologists. The BP method uses a series of tables that provide the child’s predicted percentage of adult height. The tables are indexed according to gender, chronological age, and skeletal age. Chronological age is trichotomized as $>1$ year behind skeletal age, within 1 year of skeletal age, or $>1$ year ahead of skeletal age. Tables are provided for ages 7 to 18 years, with omissions at each extreme depending on gender and skeletal age. By using our simulated children’s gender, age, and bone age, we determined the appropriate BP table, retrieved the percentage of adult height from the table, and calculated predicted adult height as current height divided by percentage of adult height. Omissions in the tables resulted in unobtainable values for percentage of adult height in a few cases, which yielded predictions for 975 of 1000 boys and 991 of 1000 girls.

The Roche-Wainer-Thissen (RWT) algorithm calculates predicted adult height directly from a linear combination of the child’s recumbent length, weight, and bone age, together with mid–parental height, by using gender- and age-specific coefficients. As described in the article on RWT methods, we calculated recumbent length as 1.25 cm greater than standing height. Coefficients used in the RWT method are tabulated to 14 years of age for girls and 16 years of age for boys. The Khamis-Roche (KR) algorithm calculates predicted adult height directly from a linear combination of child’s height and weight, together with mid–parental height. Gender- and age-specific coefficients are provided for ages 4 to 17.5 years.

We applied each of the 3 algorithms to calculate predicted height for the 1000 boys and 1000 girls. The predictions were dichotomized according to whether the child was in the lowest 1.2nd percentile of predicted adult height.

**Statistical Analyses**

Characteristics of the simulated sample were detailed by using standard descriptive statistics (mean, SD, range, and Pearson correlation coefficient) to confirm that they conformed to the intended profile. Predicted adult heights were calculated in centimeters and converted to $z$ scores and percentiles by using standard growth charts for the age of 20. The distribution of each measure was characterized by the mean and SD for the 1000 boys and 1000 girls, and the algorithms were compared pairwise by using the Spearman rank correlation statistic. The dichotomized height prediction was tabulated separately according to gender for each algorithm. Agreement among the 3 algorithms was assessed by using multirater $\kappa$. Pairwise agreement was assessed by using Cohen’s intrarater $\kappa$.

To test the reproducibility of the simulation, we generated 30 new sets of random heights, weights, and ages for 1000 boys and 1000 girls and examined the variability of the predictions. To test the sensitivity of the simulation to assumed parameters, we repeated it another 30 times with all correlations among height, weight, parental height, and bone age set to 0.

**RESULTS**

Table 1 shows characteristics of the simulated sample, which adhered closely to intended distributions and correlations. Age ranges within the sample had a mean of 11.5 years for boys and 10.3 years for girls and, by design, the sample included equal numbers of each gender. The mean $\pm$ SD height $z$ score for boys was $-2.03 \pm 0.66$ and that for girls was $-2.50 \pm 0.76$, reflecting typical patients evaluated for short stature. Age, height, and weight were strongly intercorrelated, according to design, with pairwise Pearson coefficients for these variables ranging from 88% to 97% for boys and from 77% to 96% for girls. The other simulated variables did not correlate with each other or with the cluster of height, weight, and age, with a maximal Pearson correlation of 19%.
Figure 1 shows the distribution of adult heights predicted by each algorithm in a single replicate of the simulation (1000 boys and 1000 girls). In each graph, the red line is drawn at the 1.2nd height percentile. Although this threshold is not a FDA criterion for GH use to treat children with ISS, we were interested in the distribution of and agreements in predicted adult heights around the 1.2nd percentile.

Table 2 shows summary statistics for the distribution of heights predicted by the 3 algorithms and for agreement among algorithms, averaged over 30 random replicates. The BP method predicted lowest adult stature for both genders, with mean height 7 cm below the KR prediction for boys and nearly 10 cm lower for girls. Adult height predictions obtained with the RWT method were between those obtained the other 2 methods. We found substantial disagreement among the 3 methods in the proportions of simulated patients with predicted adult heights of <1.2nd percentile (Fig 1 and Table 2). The BP method predicted adult heights of <1.2nd percentile for 43% of boys and 81% of girls, whereas the KR method predicted adult heights of <1.2nd percentile for only 3% of boys and 0.2% of girls. Both the BP and RWT methods predicted short stature for a greater proportion of girls, whereas the KR method predicted short stature for a greater proportion of boys.

Rank correlation among methods for predicting a given child’s adult height was moderately strong (0.66–0.89) (Table 2); this statistic does not take into account the systematic differences noted above for mean heights predicted by the 3 algorithms. Agreement among the methods in predicting height of <1.2nd percentile produced \( \kappa \) values ranging from 0.42 (BP versus RWT methods for boys) to virtually 0 (KR versus BP or RWT methods for girls). Overall agreement among the methods had a \( \kappa \) of 0.21 for boys and was negative (worse than chance) for girls (Table 2).

Variation among the 30 replicates was negligible, in keeping with the large sample size for each. Mean heights fluctuated by 0.1 to 0.2 cm around the average presented in Table 2, mean \( z \) scores by 0.02 to 0.04, and pairwise \( \kappa \) values by 0.00 to 0.05, with similar findings for other parameters. Abolishment of the assumed correlations among height, weight, parental height, and bone age had a negligible effect, with the only noticeable impact on Table 2 findings being attenuation of the small \( \kappa \) values.

**DISCUSSION**

Our study assessed the variation of 3 height prediction algorithms used for children with short stature. We found wide differences in height predictions among commonly used formulas. Our results showed that, on average, the BP method predicted lower adult heights than did other methods.

Our study is the first to compare height prediction algorithms by using large-scale models to represent typical short children presenting for endocrinologic evaluation. The variation in height predictions among methods is important for clinical decision-making and for interpretation of research results. Many re-
search protocols and published studies of GH treatment for children with ISS use an end point that compares near-adult height and predicted adult height.\(^4\),\(^{18}\),\(^{19}\) In addition, predicted adult height has been used by investigators as a marker of growth response to other treatments for children with ISS, including recombinant human insulin-like growth factor \(^{20}\) and combination therapy with GH and a gonadotropin hormone-releasing agonist.\(^{21}\) Use of predicted adult height as an outcome measure for children with ISS should be interpreted cautiously, because we have shown that different height prediction algorithms provide different results. In addition, although the recent consensus statement on the use of GH as treatment for ISS acknowledges the inaccuracy of predictions of adult heights, it concludes that predicted adult height may be helpful in combination with other criteria for determination of GH use.\(^{22}\) Therefore, practitioners and investigators may continue to rely on predicted adult height in decision-making.

As demonstrated by studies with small sample sizes, height prediction methods for different patient populations lead to both underestimates and overestimates of predicted adult height, with wide variations in accuracy. Maes et al\(^{23}\) compared the accuracy of 3 height prediction methods for 62 boys and 28 girls with short stature and found that the BP method was most accurate for short boys, whereas the Tanner-Whitehouse II (TWII) algorithm was most accurate for short girls. We did not use the TWII algorithm in our study because it requires a different technique for assessment of bone age than the BP and RWT algorithms. Roemmich et al\(^{24}\) compared 3 bone age methods and 3 height prediction models for 23 boys without growth disorders. Overall, the TWII method showed underprediction of near-adult height, whereas the other methods (BP method, RWT method, and RWT method with Fels bone age method) led to overpredictions.\(^{24}\) Sperlich et al\(^{25}\) explored the accuracy of height predictions for 49 boys with constitutional delay, comparing the results of the BP, TWII, and RWT formulas with adult heights. Although average predicted heights all correlated significantly with average adult heights, individual adult heights showed impressive deviations from predicted adult heights; \(>30\%\) of the

![Figure 1](https://example.com/figure1.png)

**FIGURE 1**
Distributions of adult heights predicted by using 3 height prediction algorithms. Bars to the left of the red line represent predicted adult heights that are below the 1.2nd percentile for adult height (160 cm or 63 inches for men and 150 cm or 59 inches for women). The blue and pink bars (at extreme right) represent simulated samples that did not allow for height prediction with the BP algorithm because bone age was less than the value required for the algorithm.
### TABLE 2

**Performance and Comparison of 3 Algorithms for Predictions of Adult Height for Simulated Samples of Children with ISS**

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BP</td>
<td>RWT</td>
</tr>
<tr>
<td>Adult height, mean ± SD, cm</td>
<td>161.3 ± 7.0</td>
<td>164.4 ± 4.3</td>
</tr>
<tr>
<td>Adult height percentile, mean ± SD</td>
<td>59.9 ± 9.9</td>
<td>67.2 ± 6.9</td>
</tr>
<tr>
<td>Adult height z score, mean ± SD</td>
<td>−2.18 ± 0.96</td>
<td>−1.74 ± 0.80</td>
</tr>
<tr>
<td>PAH of &lt;12nd percentile, %</td>
<td>42.6</td>
<td>17.1</td>
</tr>
</tbody>
</table>

PAH indicates predicted adult height. All parameters were averaged over 30 random samples, with each sample including 1000 boys and 1000 girls.

*a Predicted adult height of <12nd percentile is defined as a predicted adult height of <160 cm (<5 ft 3 in) for boys and <150 cm (<4 ft 11 in) for girls.

*b Spearman correlation coefficient, identical for height in centimeters, percentile, and z score.

*c κ coefficients measure agreement among algorithms, ranging from 1 (perfect agreement) to 0 (chance agreement only) to −1 (systematic disagreement).

**CONCLUSIONS**

We demonstrated wide variations in predicted adult heights by using 3 different common height prediction algorithms. The Greulich and Pyle, rather than through cultural inter-rater variability, which reduces height prediction errors. Although this method was approved in Europe, it requires extensive training of bone age readers. Another method for determining bone age is radiography of the hand and wrist. However, the use of this method was restricted to children with short stature and was not used for children with ISS.

The Greulich and Pyle method was found to be more precise than the other methods used. Given the substantial medical, psychological, and financial implications of using GH for children with ISS, it is important that parents, clinicians, and investigators understand the consequences of using GH for children with ISS.

A limitation of this study is the use of simulated patients. However, the sample used in this study was generated to reflect a realistic patient population evaluated for short stature and was used to compare 3 height prediction methods in a large cohort of children with short stature. The sample included children with ISS, and they found no significant differences among bone age methods based on bone age. The Greulich and Pyle method was found to be more precise than the other methods used. Given the substantial medical, psychological, and financial implications of using GH for children with ISS, it is important that parents, clinicians, and investigators understand the consequences of using GH for children with ISS.

The Greulich and Pyle method was found to be more precise than the other methods used. Given the substantial medical, psychological, and financial implications of using GH for children with ISS, it is important that parents, clinicians, and investigators understand the consequences of using GH for children with ISS.
considerable uncertainty of adult height predictions.

ACKNOWLEDGMENTS
Dr Swartz Topor was supported by the Clinical Investigator Training Program, Harvard University/Massachusetts Institute of Technology Health Sciences and Technology-Beth Israel Deaconess Medical Center, in collaboration with Pfizer Inc and Merck & Co. This work also was conducted with support from Harvard Catalyst, Harvard Clinical and Translational Science Center, National Institutes of Health grant UL1 RR 025758, and financial contributions from Harvard University and its affiliated academic health care centers (Dr Feldman).

We thank Joseph Wolfsdorf, MB, BCh, for his careful reading of the manuscript.

REFERENCES

944 SWARTZ TOPOR et al Downloaded from pediatrics.aappublications.org at UNIV OF CHICAGO on May 27, 2013
Variation in Methods of Predicting Adult Height for Children With Idiopathic Short Stature
Lisa Swartz Topor, Henry A. Feldman, Howard Bauchner and Laurie E. Cohen

Pediatrics 2010;126:938; originally published online October 25, 2010; DOI: 10.1542/peds.2009-3649