Bone age assessment: a large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods

R K Bull, P D Edwards, P M Kemp, S Fry, I A Hughes

Abstract

Purpose—Comparison of bone age assessed using either the “atlas matching” method of Greulich and Pyle or the “point scoring system” of Tanner and Whitehouse (TW2).

Materials and methods—362 consecutive “bone age” radiographs of the left hand and distal radius performed in a large provincial teaching hospital. Data were analysed using the “method comparison” statistical technique. Ten per cent of the radiographs were re-analysed to assess intra-observer variation.

Results—The 95% confidence interval for the difference between the two methods was 2.28 to −1.52 years. Intra-observer variation was greater for the Greulich and Pyle method than for the TW2 method (95% confidence limit, −2.46 to 2.18 v −1.41 to 1.43).

Conclusion—The two methods of bone age assessment as used in clinical practice do not give equivalent estimates of bone age and we suggest that one method only (preferably the TW2) should be used when performing serial measurements on an individual patient.

Keywords: bone age; Greulich and Pyle; Tanner and Whitehouse

Bone age is commonly assessed by one of two main methods: either the Greulich and Pyle method or the Tanner and Whitehouse (TW2) method. The Greulich and Pyle method as originally described involves a complex comparison of all of the bones in the hand and wrist against reference “normal” radiographs of different ages. In most institutions a “rapid” modified version of this technique is used, whereby the overall appearance of a given radiograph is compared with the reference radiographs and the nearest match is selected. Although this modified approach is considerably faster than the original it may be less accurate. The TW2 method relies on the systematic evaluation of the maturity of all the bones in the hand and wrist. Several small studies have compared the two methods, and have suggested that there is close agreement between them. However, the data from these studies were analysed by regression analysis, which is inappropriate for this type of comparison.

Our study compares the rapid Greulich and Pyle method, as used commonly in clinical practice, with the TW2 method in a large group of subjects. Data are analysed using the more appropriate “method comparison” technique.

Materials and methods

All bone age radiographs of the left hand, including the wrist and distal radius, performed in our institution between 1992 and 1996 for assessment of bone age were analysed (362 radiographs). The children were aged between 2 and 18 years and came from the general population of a large provincial teaching hospital (children aged < 2 years were excluded because bone age assessment from radiographs of the wrist in this age group is unreliable).

Over the four year period of study, the radiographs were assessed by a succession of radiology trainees (total 12) according to the method of Greulich and Pyle. The same radiographs were also assessed by the TW2 method by one of two nurse auxologists: specialist nurses who had received specific training in the use of the TW2 method.

Thirty nine of the radiographs (~ 10%) were then reassessed by both methods by the same observers to assess intra-observer variation for each method.

Results

Statistical analysis involved comparison of bone age assessed by these two methods. Results are shown on a scatter graph (fig 1)
Bone age assessment

As the years (age 2 to 18) increase, the greater the correlation. Furthermore, the wider the range of values for the measurements being compared (in this case from 2 to 18), the greater their strength of association between two variables, not the agreement between them; therefore, the range of values being compared (in this case from 2 to 18 years), the greater the correlation. V

<table>
<thead>
<tr>
<th>Methods</th>
<th>Mean age disparity (years)</th>
<th>SD of disparity (years)</th>
<th>95% CL.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-observer variation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greulich and Pyle method</td>
<td>0.14</td>
<td>1.16</td>
<td>−2.46 to 2.18</td>
</tr>
<tr>
<td>Tanner and Whitehouse method</td>
<td>0.01</td>
<td>0.71</td>
<td>−1.41 to 1.43</td>
</tr>
<tr>
<td>Variation between methods</td>
<td>0.38</td>
<td>0.95</td>
<td>2.28 to −1.52</td>
</tr>
</tbody>
</table>

For intra-observer variation, mean age disparity is reading 2 − reading 1. For variation between methods, mean age disparity is TW2 age − G & P age.

We gratefully acknowledge the work undertaken by Ms C Hawley and Ms P Moreland. The study was supported by the Child Growth Foundation and Pharmacia and Upjohn.

References