

# Bone age estimation and prediction of final height in patients with $\beta$ -thalassaemia major: a comparison between the two most common methods

Athanasios Christoforidis · Maria Badouraki ·  
George Katzos · Miranda Athanassiou-Metaxa

Received: 23 July 2007 / Revised: 21 August 2007 / Accepted: 21 August 2007 / Published online: 19 October 2007  
© Springer-Verlag 2007

## Abstract

**Background** Thalassaemic patients are in need of frequent assessment of bone age because of growth failure and pubertal disorders.

**Objective** To compare the “rapid” Greulich and Pyle (G&P) method with the third edition of the Tanner and Whitehouse (TW3) method for determining skeletal maturity and predicting final height in thalassaemic patients.

**Materials and methods** A total of 191 radiographs from 58 patients (28 male, 30 female) were retrospectively evaluated by two investigators, one for each method. In 47 radiographs from 15 patients having attained their adult height, predicted final height was calculated according to each method.

**Results** The mean bone ages determined by both the G&P and TW3 methods were lower than mean chronological age, although the differences were not statistically significant ( $10.04 \pm 3.69$  years and  $9.98 \pm 3.39$  years vs.  $10.78 \pm 3.96$  years, respectively). Both methods had a tendency to over-estimate final height. Overall, the TW3 method seemed to be more accurate than the G&P method (mean absolute error  $3.21 \pm 2.51$  years vs.  $3.99 \pm 2.99$  years, respectively,  $P=0.048$ ).

**Conclusions** The same method should be used when serial assessments are performed, as both methods provide similarly reliable, although not equivalent, results. The

TW3 height prediction method seemed to be more accurate in patients with  $\beta$ -thalassaemia major than the G&P method, albeit with a large confidence interval.

**Keywords**  $\beta$ -Thalassaemia major · Bone age · Greulich and Pyle · Tanner and Whitehouse

## Introduction

Beta-thalassaemia major results from a hereditary defect in the synthesis of  $\beta$ -chains of haemoglobin leading to compensatory hyperplasia of the erythroid marrow [1]. If left untreated, it causes distinctive skeletal abnormalities, most noticeable in the short bones of the hands and feet, the ribs, the spine and the skull. Particularly in the hand, diffuse medullary expansion of the shafts of the tubular bones with thinning of the cortices with additional coarsening and thickening of the trabeculae is seen [2]. Conventional management of  $\beta$ -thalassaemia major with regular transfusions, if employed early in life, prevents the development of these skeletal deformities [3]. On the other hand, iron-induced endocrinopathies resulting in growth failure and hypogonadism are present in a significant proportion of patients with  $\beta$ -thalassaemia major [4]. Frequent assessment with hand and wrist radiographs for the determination of skeletal maturation and the prediction of final height is particularly valuable in these patients.

Currently, the two most commonly used methods for bone age assessment are the Greulich and Pyle (G&P) method [5] and the third edition of the Tanner and Whitehouse (TW3) method [6]. In both methods a radiograph of the left hand and wrist is required. Regarding the G&P method, most investigators use a modified version of the originally described technique whereby the overall appearance of the studied

A. Christoforidis (✉) · G. Katzos · M. Athanassiou-Metaxa  
1st Paediatric Department, Aristotle University of Thessaloniki,  
36 Ermou Street,  
54623 Thessaloniki, Greece  
e-mail: christoforidis@doctors.org.uk

M. Badouraki  
Paediatric Radiology Department, Ippokratio Hospital,  
Thessaloniki, Greece

radiograph is compared to the corresponding standard in the G&P atlas. Although this approach is considerably less time-consuming and complex it may lack accuracy. The TW3 method is derived from a more solid mathematical base. In this method bones of the hand and wrist are classified into one of eight or nine stages, to which scores are assigned. Summing the scores of the individually studied bones gives a total score. Skeletal age is then calculated according to scoring tables. There are two versions of this method: the first uses the radius, ulna and short bones of the thumb, middle and little fingers (TW-RUS); the second additionally uses the carpal bones (TW-20). The TW3 method is more flexible and accurate; however, it is complex in use and requires more specialist training and experience. For both the G&P and TW3 methods, equations for the prediction of adult height have been developed.

This study was conducted in order to retrospectively assess and compare the rapid G&P method with the TW3 method for the determination of skeletal maturation in radiographs of children and adolescents with  $\beta$ -thalassaemia major. Additionally, the respective adult height prediction methods were compared in patients who had attained their final height.

## Materials and methods

Skeletal maturity was evaluated in 191 radiographs of the left hand and wrist from 58 patients (28 males, 30 females) with  $\beta$ -thalassaemia major by two investigators. The first investigator was a consultant paediatric radiologist with almost 20 years of experience in this field. This reader assessed skeletal maturity using the rapid modified G&P method. An interpolated age of halfway between the standards was chosen when a patient's radiograph was thought to be intermediate between two radiographic standards in the atlas. The second estimation on the same radiographs was obtained using the TW-RUS method by a paediatrician with a special interest in paediatric endocrinology and who was specially trained in the use of this method and had 5 years of experience. Both investigators knew only the gender of the patients and were blinded to previous estimations and chronological age. In order to assess intraobserver variation for each method the same 17 radiographs were reanalyzed by the two readers.

Additionally, predicted final height was calculated according to both methods using 47 radiographs (20 from eight female patients and 27 from seven male patients; mean chronological age  $12.23 \pm 2.8$  years). All patients were aged more than 20 years when the radiographs were reassessed for this study and were considered to have attained their final height when two consecutive measurements 6 months apart gave the same height. The height of each patient at the time of initial radiographic assessment was extracted from the

medical files. Using the G&P method, based on estimating bone age, final height was calculated by recording the percentage of final height achieved from specific score tables [6]. For TW3 final height prediction, the software provided with the method was employed. This software uses TW-RUS score, height and chronological age at the time the radiograph was obtained to calculate the expected final height using specific equations described in detail in the accompanying book [7].

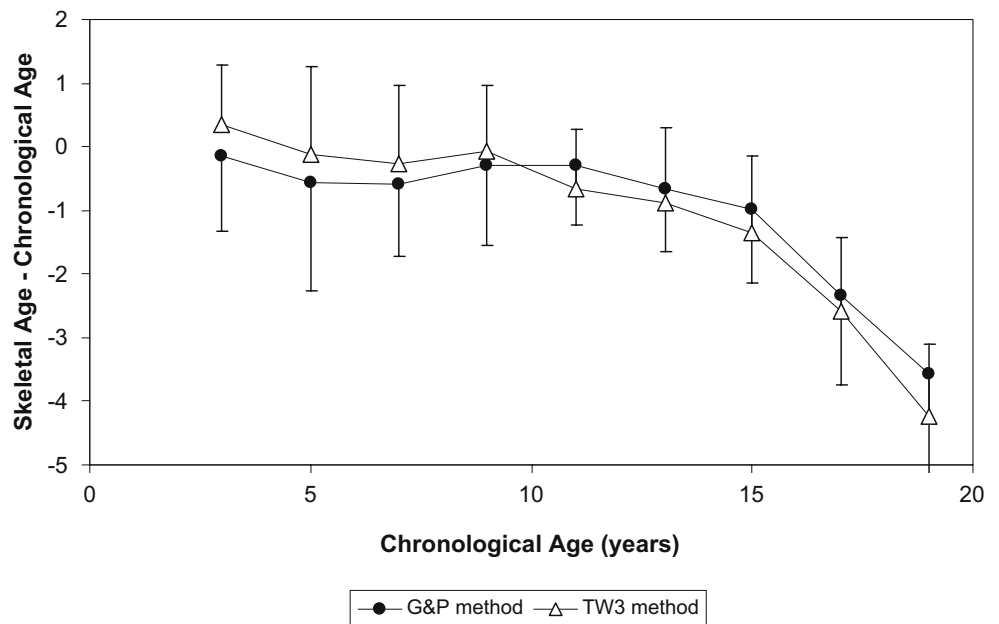
For statistical purposes and graphical demonstrations Microsoft Office Excel 2003 software was employed. Data are presented as means  $\pm$  SD. Bland-Altman scatter plots [8], being easily interpretable, were used to compare methods and intraobserver variations. The one-sample Kolmogorov-Smirnov test was used to assess the normality of the distributions of the studied parameters. As all studied parameters had a normal distribution, the paired Student's *t*-test and  $\chi^2$  test were used to assess the significance of the difference between parameters. *P* values less than 0.05 were considered statistically significant.

## Results

The mean chronological age of the patients at the time that all the 191 radiographs were obtained was  $10.78 \pm 3.96$  years (range 1.54–21.03 years). Both the G&P and TW3 methods gave mean bone ages lower than mean chronological age ( $10.04 \pm 3.69$  years vs.  $9.98 \pm 3.39$  years, respectively). The distribution of the mean differences between the estimated bone ages and chronological ages for the two methods in relation to chronological age groups is presented in Fig. 1. Both investigators demonstrated good intraobserver variability (95% confidence limits  $-1.06$ – $0.92$  years for the G&P method and  $-0.44$ – $0.4$  years for the TW3 method, Table 1). A Bland-Altman scatter plot is presented in Fig. 2 demonstrating the level of agreement between the two methods. The mean age disparity between the methods was  $-0.05 \pm 1.03$  years with 95% confidence limits ranging from  $-2.11$  to 2 years. Figure 3 is a plot of the differences between the estimated bone ages and chronological ages for the G&P method versus the TW3 method, and indicates that the two methods tend to provide concordant results.

Regarding final height prediction, the TW3 method seemed to be more accurate than the G&P method (mean absolute error  $3.21 \pm 2.51$  cm vs.  $3.99 \pm 2.99$  cm, respectively,  $P=0.048$ ). Both methods had a tendency to overestimate final height of patients with  $\beta$ -thalassaemia major. This was consistent in both genders with the exception of underestimation of the final height in females by the G&P prediction model (Table 2). Finally, as expected, the error in final height prediction was lower in older patients and this was more prominent in girls than in boys (Fig. 4).

**Fig. 1** The TW3 method gives more advanced bone ages than the G&P method in thalassaemic patients aged <10 years. In patients aged >10 years, however, the TW3 method gives less advanced bone ages than the G&P method. The bone ages obtained with both methods were younger than chronological ages in all age groups (error bars represent 1 SD from the mean, positive for the TW3 method and negative for the G&P method)



**Discussion**

Despite advances in conventional management, growth failure and hypogonadism are still reported among children and adolescents with  $\beta$ -thalassaemia major, especially in their second decade of life [9]. Hence, they should be routinely followed up regarding their stature and pubertal progression. Bone age estimation is a useful tool to assess skeletal maturity, to predict final height and to monitor treatment regimens. A mean of 3.3 radiographs (range 1–7) of the hand and wrist were obtained during the growth of each patient participating in this study and some of the patients were to be further investigated as they had not entered adulthood. Despite their frequent use, no study has yet evaluated and compared the most commonly used methods for estimation of bone age in this particular group of patients.

The accuracy of bone age estimation depends upon the investigator’s experience. In this study, two experienced readers were used. Hence, 95% confidence limits for both methods were satisfactory (–1.06–0.92 for the G&P method and –0.44–0.4 for the TW3 method). The absolute error

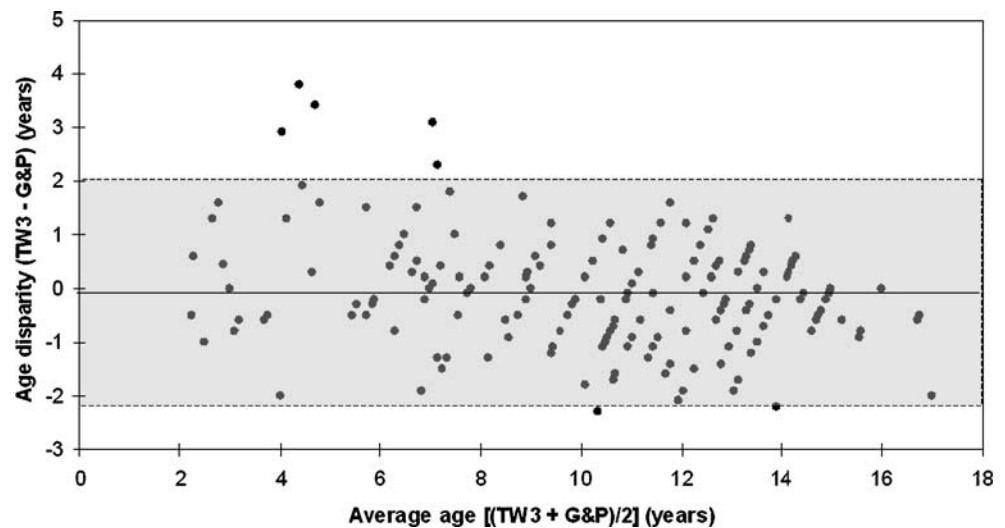
between readings was greater for the G&P method than for the TW3 method, although the difference was not statistically significant. This greater error in reproducibility with the G&P method was probably due to the use of the rapid modified version. According to this, the bone age recorded is the age given for the closest reference match, which may result in discrepancies when an individual bone is more or less advanced than its closest match. Although the originally described method included a provision for assigning bone age individually to each bone [1], this technique has seldom been used [10, 11]. Indeed, the majority of investigators have used the rapid G&P method, which is less complex and time-consuming. In a study performed by King et al. [12] the average time taken for age estimation with the G&P method was only 1.4 min, compared to 7.9 min for the TW2 method.

Previous comparative studies in normal populations have shown that bone is estimated as younger with the G&P method than with the TW method [13–17]. This consistent finding has mainly been attributed to racial and socioeconomic differences between the reference populations used for the two methods. The G&P method was based on study of American children of high socioeconomic status in the 1940s whereas the TW method was based on British children of low socioeconomic status in the 1950s. Previous comparative studies have compared the G&P method with either the TW1 or the TW2 methods. In the 2001 third edition of the TW method (now termed TW3) there are considerable changes in the reference population, which now includes population data from North America and Europe. Thus, bone ages estimated with the TW3 method are 1 year younger than those estimated with the TW2

**Table 1** Intraobserver variation (mean age disparity is the first minus the second reading; mean absolute error is the absolute value of the difference between the two readings)

Method	Mean age disparity (years)	SD of disparity (years)	95% confidence limits	Mean absolute error (years)
G&P	–0.07	0.49	–1.06–0.92	0.27±0.39
TW3	–0.02	0.21	–0.44–0.4	0.12±0.17

**Fig. 2** Bland-Altman scatter plot of the variation between the TW3 and G&P methods (mean $\pm$ SD age disparity  $-0.05\pm 1.03$  years; 95% CL  $-2.11$ – $2$ )

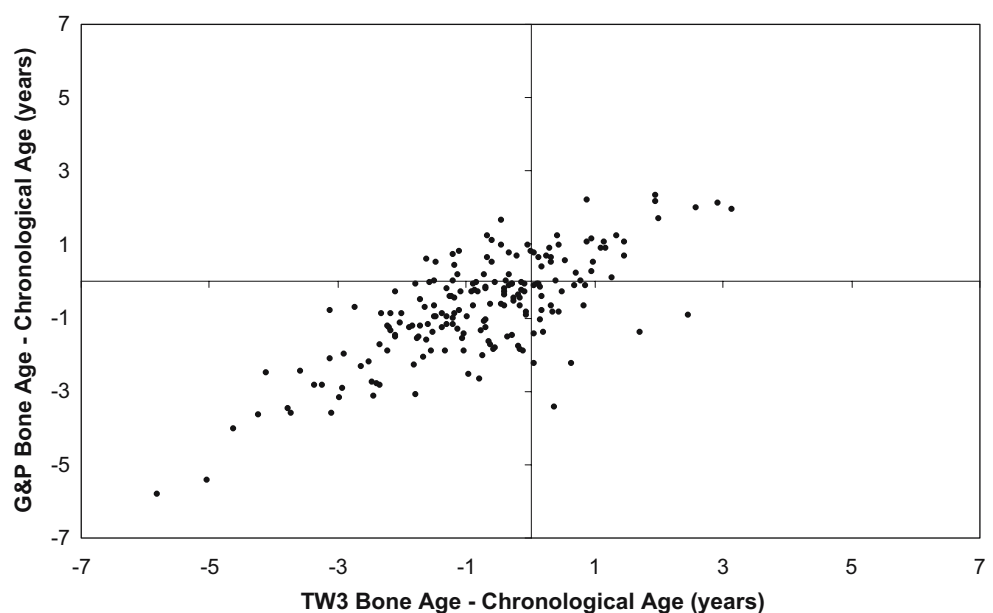


method for children aged from 10 years upwards, but show smaller differences at younger ages [7]. A recent study comparing the TW2 and TW3 methods has indeed shown that TW3 estimates of bone age are younger than TW2 estimates [18]. Our study, the first to compare the newly developed TW3 method with the G&P method, supports this observation. Furthermore, our results show that the TW3 method estimates bone ages as more advanced than the G&P method in thalassaemic patients aged <10 years, but this was reversed in older patients (Fig. 1). Bone ages obtained with both methods were younger than chronological ages in all age groups except in thalassaemic patients aged <5 years assessed with the TW3 method. Bone ages estimated with both methods were closer to chronological age in younger patients, whereas in older patients estimated

bone ages were significantly younger than chronological ages. This can be mainly attributed to the multifactorial growth and pubertal retardation observed in thalassaemic children and adolescents [19].

Prediction of final height based on bone age estimations is both challenging and clinically important as it influences to a large extent decisions on therapeutic interventions. Our results indicate that both methods generally overestimated final height in our patients, but mean errors were within acceptable limits although with wide ranges (Table 2). Indeed, if we set the confidence limits of attained final height to be predicted final height  $\pm 5.0$  cm then only 64% of the predicted final heights obtained with the G&P method and 79% of those predicted with the TW3 method were within the limits, with a statistically significant difference between

**Fig. 3** Plot of the differences between estimated bone ages and chronological ages for the TW3 method versus the G&P method. The age differences had the same sign (either positive or negative) in 151/191 radiographs (79%) and different signs in 40/191 radiographs (21%)



**Table 2** Error in final height prediction (difference in centimetres between predicted and actual final height) and absolute error (in centimetres) for the two studied methods in boys, in girls and in the total population (values are means±SD)

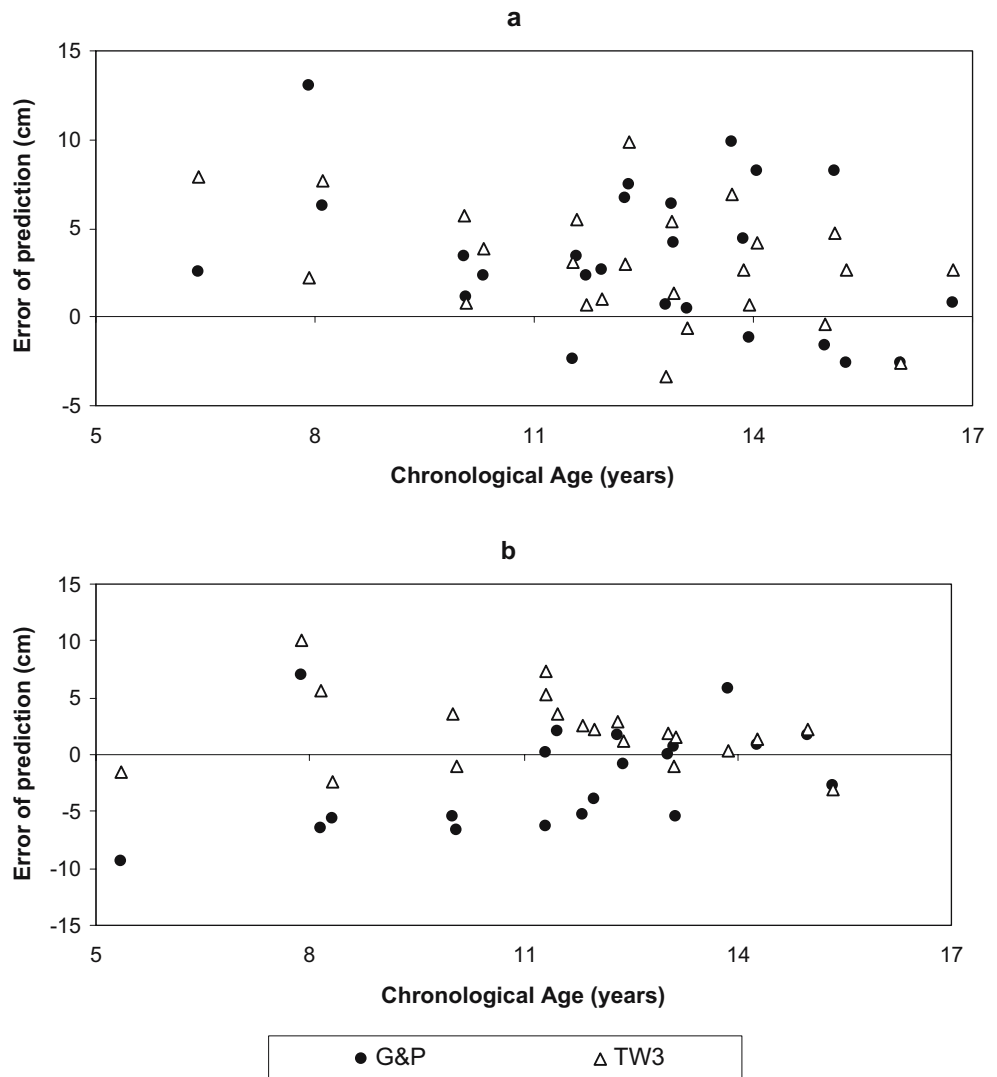
Method	Boys (n=27)			Girls (n=20)			Total (n=47)		
	Error in prediction	Absolute error	Range	Error in prediction	Absolute error	Range	Error in prediction	Absolute error	Range
G&P	3.28±3.99	4.04±3.18	-2.62–13.07	-1.91±4.47	3.91±2.78	-9.36–6.98	1.07±4.89	3.99±2.99	-9.36–13.07
TW3	2.77±3.25	3.34±2.64	-3.4–9.9	2.13±3.25	3.03±2.38	-3–10	2.5±3.29	3.21±2.51	-3.4–10

methods ( $P=0.029$ ). The prediction of final height with the G&P method in girls seems to be particularly inaccurate as only 50% of girls attained a final height within the limits of predicted final height  $\pm 5.0$  cm. Hence, these large confidence intervals must be taken into account and thalassaemic patients should be informed about the full range of predicted final height and not just the mean value.

**Conclusion**

Both studied methods (G&P and TW3) gave similar results regarding reliability, although the estimates were not equivalent. Thus, for serial assessments in an individual patient, one method should be used exclusively. The TW3 height prediction method seems to be more accurate in

**Fig. 4** Age distribution of error in prediction (predicted minus actual final height) with the G&P method (black circles) and the TW3 method (white triangles) in boys (a) and in girls (b). The error in final height prediction is lower in older patients and this is more prominent in girls than in boys



patients with  $\beta$ -thalassaemia major than the G&P method; however, large confidence intervals must be taken into account.

## References

- Weatherall DJ (1998) Pathophysiology of thalassaemia. *Baillieres Clin Haematol* 11:127–146
- Tyler PA, Madani G, Chaudhuri R et al (2006) The radiological appearances of thalassaemia. *Clin Radiol* 61:40–52
- Piomelli S, Danoff SJ, Becker MH et al (1969) Prevention of bone malformations and cardiomegaly in Cooley's anemia by early hypertransfusion regimen. *Ann N Y Acad Sci* 165:427–436
- De Sanctis V, Eleftheriou A, Malaventura C; Thalassaemia International Federation Study Group on Growth and Endocrine Complications in Thalassaemia (2004) Prevalence of endocrine complications and short stature in patients with thalassaemia major: a multicenter study by the Thalassaemia International Federation (TIF). *Pediatr Endocrinol Rev* 2 [Suppl 2]:249–255
- Greulich WW, Pyle SI (1959) Radiographic atlas of skeletal development of the hand and wrist, 2nd edn. Stanford University Press, Stanford, CA
- Bayley N, Pinneau SR (1952) Tables for predicting adult height from skeletal age: revised for use with the Greulich-Pyle hand standards. *J Pediatr* 50:423–441
- Tanner JM, Healey MJ, Goldstein H et al (2001) Assessment of skeletal maturity and prediction of adult height (TW3 method), 3rd edn. Harcourt, London
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1(8476):307–310
- De Sanctis V (2002) Growth and puberty and its management in thalassaemia. *Horm Res* 58 [Suppl 1]:72–79
- Roche AF, Davila GH, Pasternack BA et al (1970) Some factors influencing the replicability of assessments of skeletal maturity (Greulich-Pyle). *Am J Roentgenol Radium Ther Nucl Med* 109:299–306
- Wingerd J, Peritz E, Sproul A (1974) Race and stature differences in the skeletal maturation of the hand and wrist. *Ann Hum Biol* 1:201–209
- King DG, Steventon DM, O'Sullivan MP et al (1994) Reproducibility of bone ages when performed by radiology registrars: an audit of Tanner and Whitehouse II versus Greulich and Pyle methods. *Br J Radiol* 67:848–851
- Acheson RM, Vicinus JH, Fowler GB (1966) Studies in the reliability of assessing skeletal maturity from X-rays. Part 3. Greulich-Pyle atlas and Tanner-Whitehouse method contrasted. *Hum Biol* 38:204–218
- Buckler JM (1977) Comparison of systems of estimating skeletal age. *Arch Dis Child* 52:667–668
- Milner GR, Levick RK, Kay R (1986) Assessment of bone age: a comparison of the Greulich and Pyle, and the Tanner and Whitehouse methods. *Clin Radiol* 37:119–121
- Castriota-Scanderbeg A, Sacco MC, Emberti-Gialloreti L et al (1998) Skeletal age assessment in children and young adults: comparison between a newly developed sonographic method and conventional methods. *Skeletal Radiol* 27:271–277
- Bull RK, Edwards PD, Kemp PM et al (1999) Bone age assessment: a large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods. *Arch Dis Child* 81:172–173
- Ahmed ML, Warner JT (2007) TW2 and TW3 bone ages: time to change? *Arch Dis Child* 92:371–375
- Low LC (1997) Growth, puberty and endocrine function in beta-thalassaemia major. *J Pediatr Endocrinol Metab* 10:175–184