Bone age assessment: Large-scale comparison of Greulich-Pyle method and Tanner-Whitehouse 3 method for Taiwanese children

Yeong-Seng Yuh[a,b,*], Ting Ywan Chou[c,d], Tao-Hsin Tung[e]

[a]Department of Pediatrics, Cheng-Hsin General Hospital, Taipei, Taiwan, ROC; [b]Department of Pediatrics, National Defense Medical Center, Taipei, Taiwan, ROC; [c]Department of Radiology, Cardinal Tien General Hospital, New Taipei City, Taiwan, ROC; [d]College of Medicine, Fu Jen Catholic University, New Taipei City, Taiwan, ROC; [e]Evidence-based Medicine Center, Taizhou Hospital of Zhejiang Province, Wenzhou Medical University, Linhai, Zhejiang, China

Abstract

Background: The Greulich-Pyle (GP) and Tanner-Whitehouse 3 (TW3) methods are two common methods for assessing bone age (BA). The applicability of these methods for populations other than those in the United States and Europe has been questioned. Thus, this study tested the applicability of these methods for Taiwanese children.

Methods: In total, 1476 radiographs (654 boys, 822 girls) were analyzed. A subset of 200 radiographs was evaluated to determine intrarater and interrater reliability and the time required to yield a BA assessment. BA was determined by two reviewers using the GP method and two of the TW3 methods (the Radial-Ulnar-Short bones [RUS] method and the carpals method [Carpal]). The GP and TW3 methods were directly compared using statistical techniques. A subgroup analysis by age was performed to compare BA and chronological age using a paired *t* test for each age group.

Results: The average times required to yield an assessment using the GP and TW3-RUS methods were 0.79±0.14 and 3.01±0.84 min ($p < 0.001$), respectively. Both the intrarater and interrater correlation coefficients were higher for the GP method (0.993, 0.992) than the TW3-RUS (0.985, 0.984) and TW3-Carpal (0.981, 0.973) methods. The correlation coefficient for the GP and TW3-RUS methods was highest in the pubertal stage (0.898 for boys and 0.909 for girls). The mean absolute deviations for the GP and TW3-RUS methods in the pubertal stage were 0.468 years (boys) and 0.496 years (girls). Both the GP and TW3-Carpal methods underestimated BA for boys in the prepubertal stage. Both the GP and TW3-RUS methods overestimated BA for girls in the pubertal and postpubertal stages.

Conclusion: The GP and TW3-RUS methods exhibit strong agreement in the pubertal and postpubertal stages for both sexes. With appropriate adjustments based on Taiwanese data, both methods are applicable to our children.

Keywords: Bone age assessment; Chronological age; Greulich-Pyle; Tanner-Whitehouse

1. INTRODUCTION

Bone age (BA) assessment plays an important role in clinical practice, permitting investigation of whether bone maturity is occurring at a rate consistent with the individual's chronological age (CA). In this context, BA is an effective indicator for managing children with endocrine disorders and for planning orthopedic procedures.[1,2] Numerous approaches for BA assessment have been developed.[3] Among these, two methods that are commonly used are the Greulich-Pyle (GP)[4] and the Tanner-Whitehouse

3 (TW) methods, both of which involve left hand and wrist radiographs.[5]

In the GP method, BA is evaluated by comparing the radiograph of the patient with the nearest standard radiograph in the atlas; thus, this method reflects the maturity level of all 30 bones in the hand and wrist. The GP method was developed using radiographs of individuals of European descent in North America in 1938. It has been used since 1959 and remains the most commonly used method today.[1,4] The TW method was developed in the United Kingdom in the 1950s. In the TW2-RUS method the 13 (i.e., the radius, ulnar, and short) bones are evaluated and in the TW2-Carpal method, the seven carpal bones are evaluated. The maturity level of each bone is categorized into a stage and given a score, and the sum of the scores of 20 bones (TW2-RUS + TW2-Carpal) allows for the assessment of overall skeletal maturity.[5,6] Although the TW2 method is more time-consuming than other methods, previous studies have demonstrated that it is the most accurate and reliable.[7,8] Considering the trend toward more rapid skeletal maturation in many countries, Tanner et al[5] published new reference values based on American and European data obtained between the 1960s and 1990s, and the TW3 method, an update to the TW2 method, is based on these new reference values. Priority is now

given to the RUS score (13 bones maturity) over the Carpal score (seven bones maturity) or the 20-bone maturity score; this is because "in most circumstances, the RUS score is all that is required."[5]

Following the introduction of the GP and TW3 methods in many countries, numerous studies have evaluated the applicability of these methods to various populations.[8–12] However, an increasing number of studies have found that certain methods are inappropriate for some ethnic groups due to improvements in socioeconomic status.[10,13,14] In particular, the present-day applicability of these reference standards to the Taiwanese population is unclear.

The goals of this study were (1) to test the applicability of the GP and TW3 methods for assessing BA of Taiwanese children born in the 21st century and (2) to compare the GP method with the TW3 method when applied to Taiwanese children born in the 21st century. Our null hypothesis was that both methods do not differ.

## 2. METHODS

2.1. Data sources
In this retrospective study, we collected medical records of Taiwanese children and adolescents who visited our pediatric endocrine clinic for height evaluation and final height prediction at any time from October 1, 2010, to September 30, 2020. Data on demographic characteristics, specifically sex, age, weight, and height, were collected. The protocol for the patient selection process is illustrated in Fig. 1 and was approved by the institutional review board of clinical investigation at the Chen-Hsien Hospital in Taipei. Candidates were excluded if they had a diagnosis of any genetic or endocrine diseases. Children whose height or weight was not within the 15th and 85th percentiles for the mean age-adjusted normal values for Taiwanese children were excluded from further analysis.[15] Boys aged <5 years or >17 years and girls aged <3 years or >15 years were excluded because there were too few of them. The study included 1476 children (654 boys and 822 girls).

2.2. Data analysis
Radiographs of the left hands and wrists of the patients were assessed according to the GP and TW3 methods. Values obtained from the TW3 method were calculated separately based on the radius-ulna-short bone score (RUS) and the carpal bone score (Carpal). The assessments were performed by a senior pediatric endocrinologist and were reviewed by a senior pediatric radiologist. The agreement between the readings of the two reviewers was evaluated by calculating the intrarater and interrater correlation coefficients of 200 standard radiographs that were selected from each age group of both sexes (100 boys and 100 girls). The time required for each assessment was recorded.

A subgroup analysis by sex was performed. The evaluation of TW3 method was separated into TW3-RUS result and TW3-Carpal result in order to compare separately with the result of GP method. The difference between CA and BA was calculated using the GP and TW3 methods, and the results from each method were compared. A subgroup analysis by age (prepubertal, pubertal, and postpubertal) was performed. For boys, the prepubertal, pubertal, and postpubertal groups included those aged between 5 and 9, 9.1 and 14, and 14.1 and 17 years, respectively; for girls, the prepubertal, pubertal, and postpubertal groups included those aged between 3 and 8, 8.1 and 12, and 12.1 and 15 years, respectively.

BA and CA were compared using paired *t* analysis in yearly intervals to study the accuracy of the GP and TW3 methods
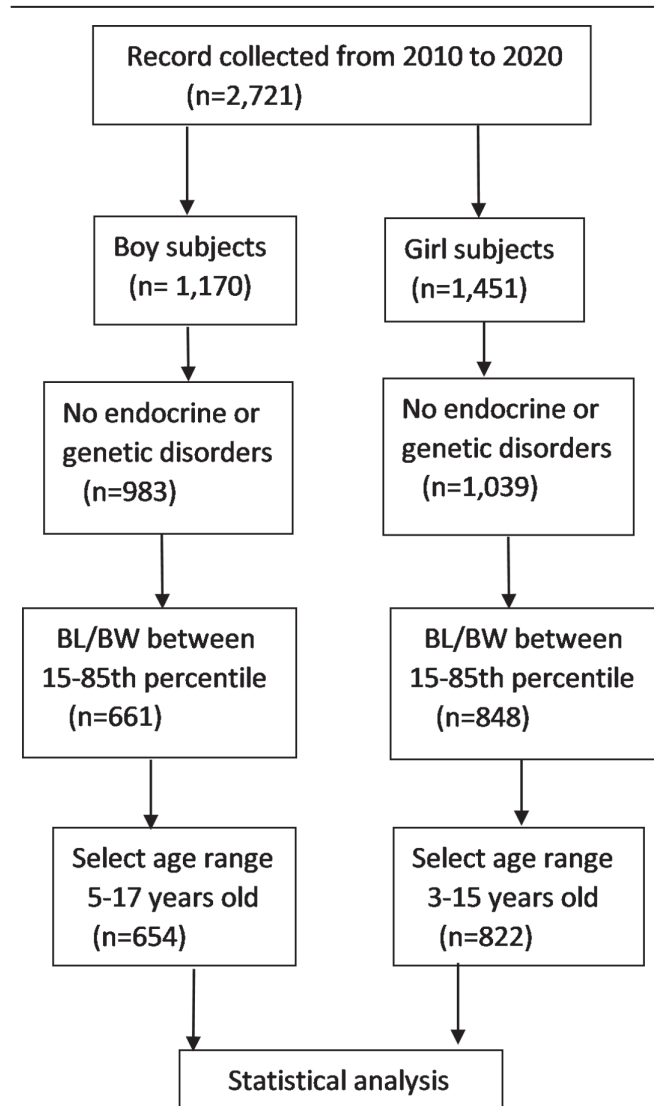
Fig. 1 Algorithm of the patient selection process. BL = body length; BW = body weight.

for assessing BA in Taiwanese children. The difference between BA and CA was calculated for each patient by subtracting the BA from the CA. The mean of the difference between BA and CA was plotted against CA in yearly intervals to compare the variation in assessment accuracy at various CAs. A positive value indicates advanced BA, whereas a negative value indicates delayed BA.

2.3. Statistical analysis
Data were analyzed using SPSS version 20 (SPSS, Chicago, IL, USA). Descriptive statistics (specifically, the mean and SD of the time required for each assessment, CA, BA, and the difference between BA and CA) were analyzed. Several statistical tests were used to compare the evaluation results. Pearson correlation coefficients were calculated to determine the strength of linear relationship. Bland–Altman plots were used to quantify agreement between methods.[16] The root mean square deviation (RMSD) and mean absolute deviation (MAD) were determined to evaluate precision. The significance of differences between BA and CA (to evaluate accuracy) and between the time required for the GP and TW3 methods to yield an assessment was calculated using a paired *t* test. Correlation coefficients of 0.7 to 1 were

defined as strong, 0.4 to 0.7 as moderate, and <0.4 as weak.[17] The method with the least MAD and RMSD was defined as the most suitable method of BA assessment. $p < 0.05$ indicated statistical significance.

3. RESULTS

3.1. Accuracy and time required for each individual assessment system

BA for a subset of 200 radiographs was assessed by two reviewers, and intrarater and interrater variations were analyzed. The correlation coefficients of intrarater variation for GP, TW3-RUS, and TW3-Carpal methods were 0.993, 0.985, and 0.981, respectively; and those for interrater variation were 0.992, 0.984, and 0.973, respectively. The correlation was higher for the GP method than the TW3-RUS and TW3-Carpal methods. The results for the time required for the GP, TW3-RUS, and TW3-Carpal methods to yield an assessment were 0.79 ± 0.14, 3.01 ± 0.84, and 1.85 ± 0.55 minutes, respectively. The GP method required significantly less time than the TW3-RUS and TW3-Carpal methods ($p < 0.001$).

3.2. Direct comparison between the GP method and TW3 method

The data of 1476 children were compared using the Pearson correlation comparison and method comparison technique. The results are shown in scatter plots (Fig. 2) and Bland–Altman plots (Fig. 3). The scatter plots indicated strong linear correlations of the GP with TW3-RUS and TW3-Carpal methods for both sexes. The correlation was stronger between the GP and TW3-RUS methods than for the GP and TW3-Carpal methods in females, but was weaker in males. The TW3-RUS BA ceases to apply at 16.5 years for boys and 15 years for girls, whereas the GP BA can still reflect the maturation tempo at these ages. The TW3-Carpal BA ceases to apply at 14.5 years for boys and 13 years for girls. The Bland–Altman plots indicate that the age disparity between the GP BA and TW3-Carpal BA was wider than between the GP BA and TW3-RUS BA. The variation in age was greater for boys than for girls for the GP BA and TW3-RUS BA.

Subgroup analysis by age (prepubertal, pubertal, and postpubertal) was performed (Table 1). Each age subgroup exhibited greater within-group variation than the overall sample because the subgroups were smaller. The correlation coefficients between
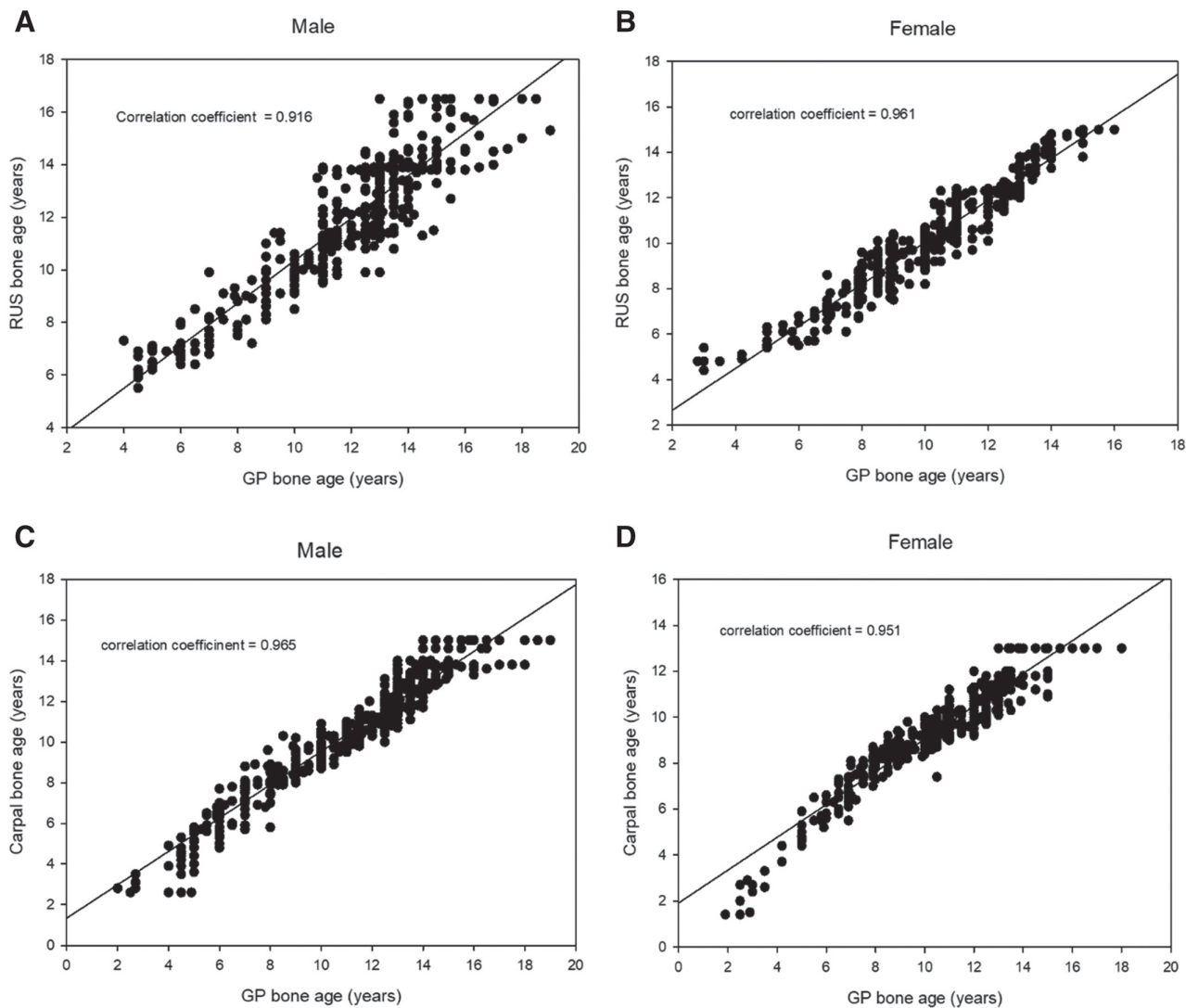
Fig. 2 Scatter plots of correlation among bone ages assessed using 2 methods. . A, GP vs TW3 RUS in male individuals. B, GP vs TW3 RUS in female individuals. C, GP vs Carpal in male individuals. D, GP vs Carpal in female individuals. Carpal, Tanner-Whitehouse 3 carpal; GP; Greulich-Pyle; TW3 RUS; Tanner-Whitehouse 3 radius, ulnar, and short bones.
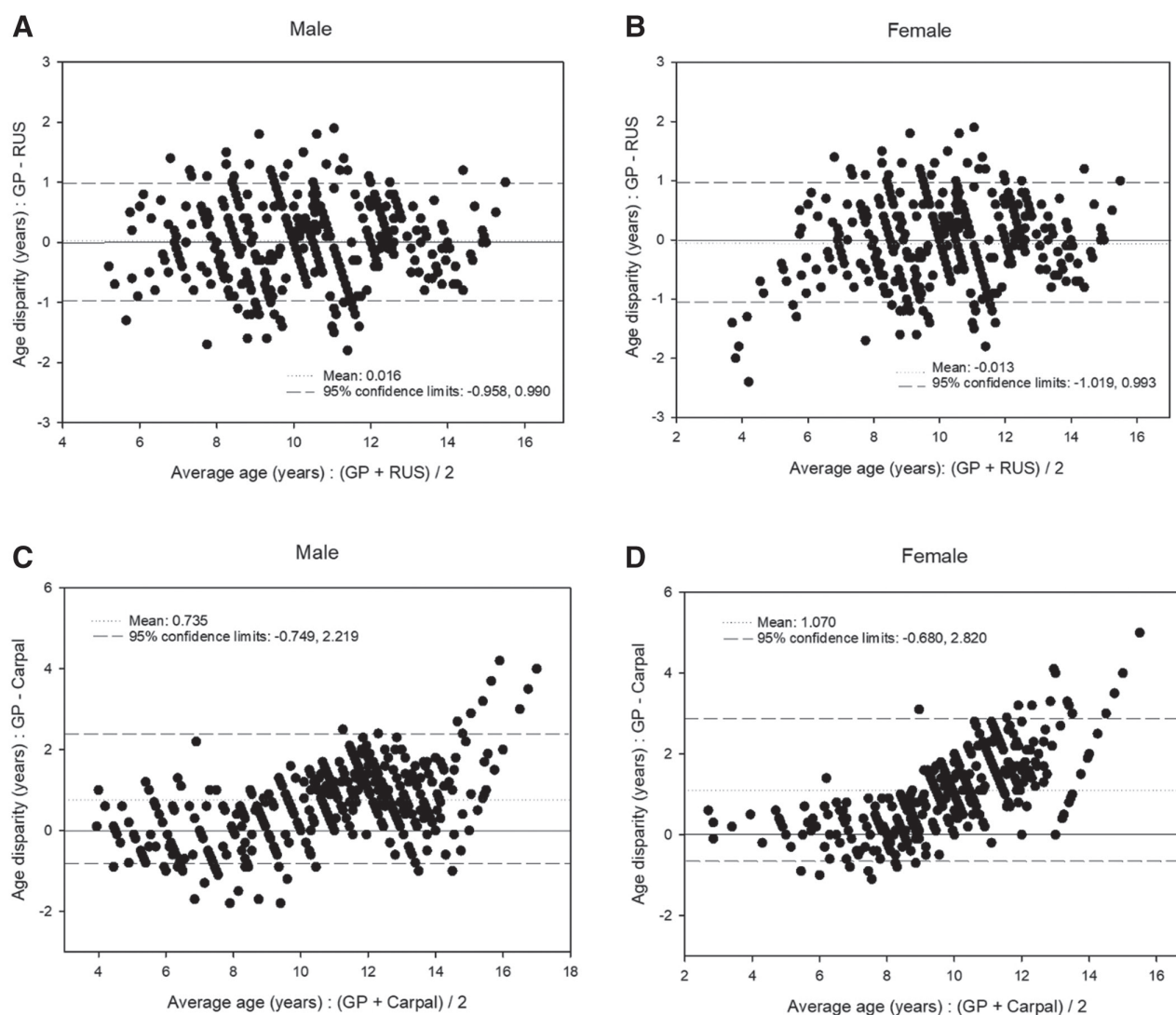
Fig. 3 Bland–Altman plots for age disparity among bone ages assessed using 2 methods. A, GP vs TW3 RUS in male individuals. B, GP vs TW3 RUS in female individuals. C, GP vs Carpal in male individuals. D, GP vs Carpal in female individuals. Carpal, Tanner-Whitehouse 3 carpal; GP, Greulich-Pyle; TW3 RUS, Tanner-Whitehouse 3 radius, ulnar, and short bones.

the GP and TW3-RUS method for the prepubertal stage were the lowest among the three pubertal stages for both sexes, and the data disparities of BAs (MAD and RMSD) for the prepubertal stage were the highest among the three pubertal stages. In contrast, the GP and TW3-RUS methods exhibited strong agreement in the pubertal and postpubertal stages for both genders. These findings suggest that the GP and TW3-RUS methods agree more for individuals at puberty and postpuberty than for individuals at prepuberty.

3.3. Comparison between assessment methods using the mean difference in BA and CA for each age group
To obtain more comprehensive findings, we compared our findings for both methods against the mean difference in BA and CA at yearly intervals of CA (Fig. 4). The relationship between the difference between the BA-CA difference and CA, for both the GP and TW3 methods, was determined. In prepubertal boys, the mean age difference using the GP method was close to the mean age difference using the TW3-Carpal method, although both methods underestimated BA. The pattern of underestimation was similar for both methods. The TW3-RUS method

overestimated BA for both boys and girls in the prepubertal stage. The mean age difference using the GP method was close to that of the TW3-RUS method in the pubertal and postpubertal stages, with a tendency toward overestimating BA. The mean age difference using the TW3-Carpal method was severely underestimated in the pubertal and postpubertal stages.

Statistical data on 654 males and 822 females, including CA, BA assessed by different methods, the difference between BA and CA, and the results of paired t tests, are illustrated in Table 2. The age range of statistically agreement between BA and CA was broader in the GP data than that of those from TW3-RUS and TW3-Carpal methods.

4. DISCUSSION
This study confirms that "the TW3 norms are, incidentally, very close to the level of maturity represented by the old Gruelich–Pyle atlas" in BA assessment for Taiwanese children.[5] The TW3-RUS methods and GP method exhibit a reasonable level of agreement for individuals in the pubertal and postpubertal stages. Our results show that the MADs are 0.34–0.47 years in

Table 1

Agreement between bone ages assessed by the GP and the Tanner-Whitehouse 3 RUS methods and between the GP and the Tanner-Whitehouse 3 carpal (Carpal) methods, overall and for three pubertal stages

Methods	Age	n	Correlation coefficient	MAD (y)	RMSD (y)
GP vs RUS	Boys overall (5–17 y)	654	0.958	0.487	0.698
	Girls overall (3–14 y)	822	0.961	0.503	0.709
	Prepubertal boys (5–9 y)	148	0.860	0.575	0.758
	Prepubertal girls (3–8 y)	215	0.882	0.653	0.808
	Pubertal boys (9–14 y)	413	0.898	0.468	0.684
	Pubertal girls (8–12 y)	500	0.909	0.496	0.704
	Postpubertal boys (14–17 y)	93	0.861	0.335	0.579
	Postpubertal girls (12–15 y)	107	0.899	0.354	0.595
GP vs Carpal	Boys overall (5–17 y)	654	0.962	0.955	0.977
	Girls overall (3–14 y)	822	0.949	1.070	1.035
	Prepubertal boys (5–9 y)	148	0.871	0.589	0.772
	Prepubertal girls (3–8 y)	215	0.930	0.217	0.413
	Pubertal boys (9–14 y)	413	0.915	0.907	0.953
	Pubertal girls (8–12 y)	500	0.888	1.122	1.059
	Postpubertal boys (14–17 y)	93	0.661	1.537	1.219
	Postpubertal girls (12–15 y)	107	0.671	1.995	1.194

GP = Greulich-Pyle; MAD = mean absolute deviation; RMSD = root mean square deviation; RUS = radius, ulnar, and short bones.
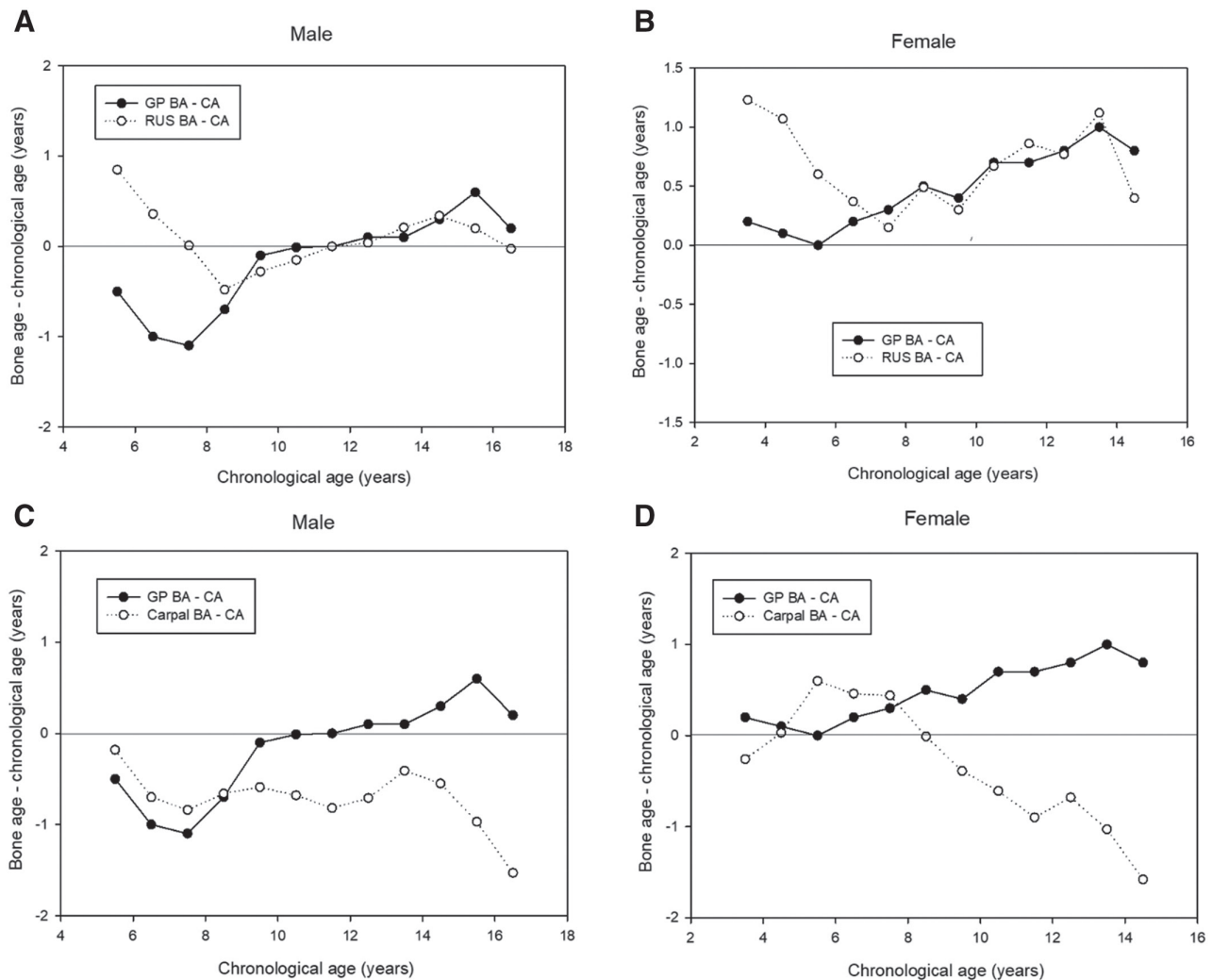
Fig. 4 Mean age difference between CA and BA assessed using 2 methods. Zero baseline represents no difference between assessed BA and CA. A, Comparison between GP BA-CA and TW3 RUS BA-CA in male individuals. B, Comparison between GP BA-CA and TW3 RUS BA-CA in female individuals. C, Comparison between GP BA-CA and Carpal BA-CA in male individuals. D, Comparison between GP BA-CA and Carpal BA-CA in female individuals. BA = bone age; CA = chronological age; GP = Greulich-Pyle; TW3 RUS = Tanner-Whitehouse 3 radius, ulnar, and short bones.

Table 2

CA and BA assessed using the GP method, Tanner-Whitehouse 3 RUS method and the Tanner-Whitehouse 3 carpal (Carpal) method, and the difference (BA–CA) between these measures for different age groups

Sex	Age group (y)	n	CA (y)	GP BA (y)	RUS BA (y)	Carpal BA (y)	GP BA-CA (y)	RUS BA-CA (y)	Carpal BA-CA (y)	p value (GP vs CA)	p value (RUS vs CA)	p value (Carpal vs CA)
M	5–5.9	27	5.6 ± 0.2	5.1 ± 0.8	6.4 ± 0.6	5.5 ± 0.9	−0.5 ± 0.9	0.8 ± 0.6	−0.2 ± 0.8	0.182	0.006	0.464
	6–6.9	41	6.5 ± 0.3	5.5 ± 0.7	6.8 ± 0.5	5.8 ± 1.0	−1.0 ± 0.7	0.4 ± 0.6	−0.7 ± 1.0	<0.001	0.025	0.004
	7–7.9	39	7.4 ± 0.3	6.4 ± 1.3	7.4 ± 0.9	6.7 ± 1.4	−1.1 ± 1.1	0.0 ± 0.8	−0.8 ± 1.2	0.001	0.976	0.002
	8–8.9	41	8.4 ± 0.3	7.7 ± 1.3	7.9 ± 0.9	7.8 ± 1.0	−0.7 ± 1.5	−0.5 ± 0.9	−0.7 ± 1.1	0.034	0.033	0.002
	9–9.9	48	9.5 ± 0.3	9.3 ± 1.4	9.2 ± 0.8	9.0 ± 1.1	−0.1 ± 1.4	−0.3 ± 0.8	−0.6 ± 1.0	0.650	0.131	0.002
	10–10.9	78	10.5 ± 0.3	10.5 ± 1.1	10.1 ± 0.9	9.8 ± 0.9	0.2 ± 1.0	−0.2 ± 0.9	−0.7 ± 0.9	0.912	0.005	<0.001
	11–11.9	125	11.4 ± 0.3	11.5 ± 1.2	11.4 ± 0.9	10.7 ± 0.9	0.0 ± 1.2	0.0 ± 0.9	−0.8 ± 0.9	0.902	0.969	<0.001
	12–12.9	92	12.5 ± 0.3	12.6 ± 1.0	12.6 ± 1.0	11.8 ± 1.0	0.1 ± 1.0	0.0 ± 1.0	−0.7 ± 1.0	0.230	0.677	<0.001
	13–13.9	70	13.5 ± 0.3	13.6 ± 0.8	13.7 ± 0.7	13.1 ± 0.8	0.1 ± 0.8	0.2 ± 0.7	−0.4 ± 0.8	0.195	0.043	<0.001
	14–14.9	51	14.5 ± 0.3	14.8 ± 1.3	14.8 ± 1.0	13.9 ± 0.7	0.3 ± 1.2	0.3 ± 0.9	−0.6 ± 0.8	0.232	0.074	<0.001
	15–15.9	27	15.6 ± 0.3	16.2 ± 1.6	15.8 ± 0.8	14.5 ± 0.6	0.6 ± 1.6	0.2 ± 0.8	−1.0 ± 0.6	0.211	0.426	<0.001
	16–16.9	15	16.4 ± 0.3	16.5 ± 1.3	16.1 ± 0.7	14.8 ± 0.3	0.1 ± 1.4	0.0 ± 0.7	−1.5 ± 0.4	0.704	0.184	<0.001
F	3–3.9	13	3.5 ± 0.3	3.4 ± 0.6	4.7 ± 0.3	3.3 ± 0.8	0.1 ± 0.6	1.2 ± 0.4	−0.3 ± 0.8	0.541	<0.001	0.523
	4–4.9	18	4.5 ± 0.3	4.6 ± 1.2	5.6 ± 0.4	4.5 ± 1.2	0.1 ± 1.2	1.1 ± 0.4	0.0 ± 1.1	0.805	0.001	0.949
	5–5.9	33	5.5 ± 0.3	5.5 ± 0.5	6.1 ± 0.5	6.1 ± 0.8	0.0 ± 0.6	0.6 ± 0.7	0.6 ± 0.8	0.905	0.039	0.003
	6–6.9	52	6.5 ± 0.2	6.7 ± 0.4	6.9 ± 0.9	7.0 ± 1.0	0.2 ± 0.5	0.4 ± 1.1	0.5 ± 1.1	0.273	0.209	0.047
	7–7.9	99	7.6 ± 0.3	7.9 ± 0.8	7.7 ± 0.3	8.0 ± 0.7	0.3 ± 0.7	0.2 ± 0.9	0.4 ± 0.7	0.010	0.320	<0.001
	8–8.9	136	8.6 ± 0.3	9.0 ± 0.8	9.1 ± 0.9	8.5 ± 0.7	0.5 ± 0.8	0.5 ± 0.9	−0.3 ± 0.6	<0.001	<0.001	0.718
	9–9.9	151	9.5 ± 0.3	10.0 ± 1.1	9.8 ± 1.0	9.2 ± 1.0	0.4 ± 1.1	0.3 ± 1.0	−0.4 ± 0.7	<0.001	0.002	<0.001
	10–10.9	126	10.5 ± 0.3	11.2 ± 0.9	11.2 ± 0.8	9.9 ± 0.6	0.7 ± 0.9	0.7 ± 0.9	−0.6 ± 0.7	<0.001	<0.001	<0.001
	11–11.9	87	11.5 ± 0.3	12.2 ± 0.9	12.4 ± 0.7	10.6 ± 0.6	0.7 ± 0.8	0.9 ± 0.7	−0.9 ± 0.6	<0.001	<0.001	<0.001
	12–12.9	49	12.4 ± 0.3	13.3 ± 0.9	13.2 ± 1.0	11.8 ± 0.9	0.8 ± 0.8	0.8 ± 0.9	−0.7 ± 0.9	<0.001	<0.001	<0.001
	13–13.9	38	13.4 ± 0.3	14.4 ± 0.8	14.5 ± 0.5	12.4 ± 0.7	1.0 ± 0.7	1.1 ± 0.4	−1.0 ± 0.7	<0.001	<0.001	<0.001
	14–14.9	20	14.4 ± 0.3	15.1 ± 0.4	14.8 ± 0.4	12.9 ± 0.4	0.8 ± 0.3	0.4 ± 0.6	−1.6 ± 0.6	<0.001	0.008	<0.001

BA = bone age; CA = chronological age; GP = Greulich-Pyle; MAD = mean absolute deviation; RMSD = root mean square deviation; RUS = radius, ulnar, and short bones.

boys and 0.35–0.50 years in girls within these stages. In clinical practice, when a single measurement of BA is required for diagnosis, a tolerance of ±0.5 years was suggested to be acceptable.[18,19] Our study also indicates that BA assessed using the TW3-Carpal method severely deviates from CA in the pubertal and postpubertal stages for both sexes. The TW3-Carpal method is no longer used in most circumstances due to its low accuracy. Previous studies have suggested that intrarater correlation is greater for the TW2 method than the GP method.[7,8] Our results showed that both intrarater and interrater correlations for the GP method were higher than the TW3 method. No study has evaluated the time required for the TW3-RUS assessment alone to yield a result. Previous studies have only shown that the average time required for BA assessment was 7.9 minutes for the TW2-20-bone method and 1.4 minutes for the GP method.[1,2] In our findings, the average time required for BA assessment was 3.01 minutes for the TW3-RUS method, 1.85 minutes for the TW3-Carpal method, and 0.79 minutes for the GP method.

We found that the skeletal maturation tempo in Taiwanese children differs from those in the reference populations for both the GP and TW3 methods. The most notable deviation was that of the GP method, which tended to underestimate BA (by about 1 year) for boys aged 6 to 8 years. A similar pattern was noted with the TW3-Carpal method. The TW3-RUS method tended to overestimate BA for boys aged 6 to 8 years. Similar findings were reported by Griffith in the Hong Kong population.[10] For girls in the prepubertal stage, the GP method accurately estimated CA, and the TW3-RUS method overestimated CA, after which there was a rapid advancement of BA. Taiwanese children reach the end of maturity prior to the age observed through the TW3 method. One of the advantages of the GP method over the TW3-RUS method is that the GP method yields better results for postpubertal adolescents; thus, the GP method can be applied to a sample with a wider age range. For both sexes, BA assessed using the GP method could continuously reflect the maturation tempo of CA in the pubertal and postpubertal stages, whereas BA assessed using the TW3-RUS and TW3-Carpal methods were limited by a saturated skeletal maturity score of 1000. According to our results, the TW3 method is inappropriate for late teenagers.

A secular trend toward early onset of puberty in girls has been developing in many countries, including Taiwan, since the late 1800s.[20–22] A similar trend in skeletal maturity has been developing over the past 80 years, but this trend has differed between the sexes.[23] Our previous study showed that mean BA assessed using the GP method in girls was generally advanced by 0.3 to 1 year between 7 and 15 years of age.[11] In this study, results obtained using the TW3-RUS method confirmed that trend. The rate of skeletal maturity differs between boys and girls. BA for people of Chinese descent, especially those in China and Taiwan, was delayed in early childhood and advanced in adolescence.[10,11,24,25] Our previous study showed that ulnar bone maturity was delayed in young Taiwanese boys.[11] Our data in this study suggest that carpal bone maturity is also delayed in the prepubertal stage. A delay in ulnar and carpal bone maturity may represent a normal tempo for Taiwanese boys in the prepubertal stage. Due to the potential effects of ethnicity and secular trends, the adaptation of BA assessment methods for the contemporary Taiwanese population should be considered. Per our findings, we recommend the normalization of BA assessment methods among the Taiwanese population. For example, for a Taiwanese boy with a CA of 7.5 years, 1.1 years should be added to the BA indicated by the GP method. Similarly, 1 year should be subtracted from the BA of a Taiwanese girl with a CA of 13 years. The tempo of skeletal maturation differs between populations. Thus, reference to local standards may be valid.

Given the differences noted in the BA assessments in our study, a population-specific standard may be more useful in assessing BA. Revised TW3 reference values (China-05) for BA assessment were recently developed in mainland China.[26,27] Taiwanese people are ethnically similar to Southern Chinese people, but further study is required to determine whether the China-05 reference values are applicable to Taiwanese children.

This study used a variety of statistical techniques to compare BA assessment methods when applied to Taiwanese children. Most previous studies comparing different methods of BA estimation have used correlation analysis.[21,28–30] The correlation coefficient alone measures the strength of linear agreement between two variables, but it does not measure the accuracy and precision. The wider the range of values being compared (in this case, the overall age compared with a specific pubertal stage), the greater the correlation.[17] Our approach accurately measured the agreement of accuracy and precision between the GP and TW3 methods.

Our study has the following limitations. First, children involved in this study were from a single clinical department in Taipei city; thus, our findings do not account for potential country-wide variation. Second, selection bias may have been present because the health status of our patients potentially deviated from that in the average population. To minimize this bias, we selected children without endocrine disorders and with an age-adjusted height and weight that was within the 15th and 85th percentiles of the national standard. Third, our sample was limited by the small sample size of patients at the extreme ends of the age range. With an increase in the number of patients, the absolute difference between BA and CA may vary. Fourth, the reviewers had more experience using the GP method than the TW3 method. To become familiar with the TW3 method, the reviewers were trained over a 3-month period. The TW3 method is generally considered to be more objective than the GP method; thus, we consider a 3-month training period to be sufficient.

In conclusion, Taiwanese children exhibit a different pattern of skeletal maturation than the children on whom the GP and TW3 methods were originally based. With the incorporation of reference values based on Taiwanese data, both the GP and TW3-RUS methods can be made applicable to Taiwanese children. The GP norms were compatible with the TW3-RUS norms for boys aged between 9 and 15 years and for girls aged between 7 and 13 years. Because the GP method requires less time to yield a result and covers a wider age range, it has greater clinical utility.

ACKNOWLEDGMENTS

REFERENCES

1. Cavallo F, Mohn A, Chiarelli F, Giannini C. Evaluation of bone age in children: a mini-review. *Front Pediatr* 2021;**9**:580314.
2. Creo AL, Schwenk WF 2nd. Bone age: a handy tool for pediatric providers. *Pediatr* 2017;**140**:1–11.
3. Prokop-Piotrkowska M, Marszałek-Dziuba K, Moszczyńska E, Szalecki M, Jurkiewicz E. Traditional and new methods of bone age assessment—an overview. *J Clin Res Ped Endocrinol* 2021;**13**:251–62.
4. Greulich WW, Pyle SI. *Radiographic atlas of skeletal development of the hand and wrist*. 2nd ed. Standford, CA: Standford University Press; 1959. Available at https://www.sup.org/books/.
5. Tanner JM, Healy M, Goldstein H, Cameron N. *Assessment of skeletal maturity and prediction of adult height (TW3 method)*. London: WB Saunders; 2001. Available at http://www.elsevierhealth.com/.

6. Tanner JM. *Growth at adolescent*. 2nd ed. Springfield, IL: Blackwell Scientific Publications; 1962. Available at http://www.blackwellpublishing.com/.

7. Bull RK, Edwards PD, Kemp PM, Fry S, Hughes IA. Bone age assessment: a large scale comparison of the Greulich and Pyle, and Tanner-Whitehouse (TW2) methods. *Arch Dis Child* 1999;**81**:172–3.

8. King DG, Steventon DM, O'Sullivan MP, Cook AM, Hornsby VP, Jefferson IG, et al. Reproducibility of bone ages when performed by radiology registrars: an audit of Tanner and Whitehouse II versus Greulich and Pyle methods. *Br J Radiol* 1994;**67**:848–51.

9. Alshamrani K, Messina F, Offiah AC. Is the Greulich and Pyle atlas applicable to all ethnicities? A systemic review and meta-analysis. *Eur Radiol* 2019;**29**:2910–23.

10. Griffith JF, Cheng JCY, Wong E. Are western skeletal age standards applicable to the Hong Kong Chinese population? A comparison of the Greulich and Pyle method and the Tanner and Whitehouse method. *Hong Kong Med J* 2007;**13**:28–32.

11. Yuh YS, Chou TW, Chow JC. Applicability of the Greulich and Pyle bone age standards to Taiwanese children: a Taipei experience. *J Chin Med Assoc* 2022;**85**:767–73.

12. Alshamrani K, Offiah AC. Applicability of 2 commonly used bone age assessment methods to twenty-first century UK children. *Eur Radiol* 2020;**30**:504–13.

13. Benjavongkulchai S, Pittayapat P. Age estimation methods using hand and wrist radiographs in a group of contemporary Thais. *Forensic Sci Int* 2018;**287**:218.e1–8.

14. Oh MS, Kim S, Lee J, Lee MS, Kim YJ, Kan KS. Factors associated with advanced bone age in overweight and obese children. *Ped Gastroenterol Hepatol Nutri* 2020;**23**:89–97.

15. Chen W, Chang MH. New growth charts for Taiwanese children and adolescents based on World Health Organization standards and health-related physical fitness. *Pediatr Neonatol* 2010;**51**:69–79.

16. Bland JM, Altman DG. Statistical methods for assessing agreement between 2 methods of clinical measurement. *Lancet* 1986;**1**:307–10.

17. Akoglu H. User's guide to correlation coefficients. *Turk J Emergent Med* 2018;**18**:91–3.

18. Buckler JM. How to make the most of bone ages. *Arch Dis Chil* 1983;**58**:761–3.

19. Rijn RV, Thodberg HH. Bone age assessment: automated techniques coming of age? *Acta Radiol* 2013;**54**:1024–9.

20. Chow JC, Chou TY, Tung TH, Yuh YS. Recent pubertal timing trends in Northern Taiwanese children: comparison with skeletal maturity. *J Chin Med Assoc* 2020;**83**:870–5.

21. Miltenburg Caspersen L, Sonnesen L. Secular trend of the skeletal maturation in relation to peak height velocity-a comparison between 2 groups of children born 1969-1973 and 1996-2000. *Eur J Ortho* 2020;**42**:612–8.

22. Eckert-Lind C, Busch AS, Petersen JH, Biro FM, Butler G, Bräune EV, et al. Worldwide secular trends in age at pubertal onset assessed by breast development among girls: a systematic review and meta-analysis. *JAMA Ped* 2020;**174**:e195881.

23. Duren DL, Nahhas RW, Sherwood RJ. Do secular trends in skeletal maturity occur equally in both sexes? *Clin Orthop Relat Res* 2015;**473**:2559–67.

24. Ontell FK, Ivanovic M, Ablin DS, Barlow TW. Bone age in children of diverse ethnicity. *Am J Roentgenol* 1996;**167**:1395–8.

25. Zhang SY, Liu G, Ma CG, Han YS, Shen XZ, Xu RL, et al. Automated determination of bone age in a modern Chinese population. *ISRN Radiol* 2013;**2013**:874570.

26. Zhang SY, Liu LJ, Wu ZL, Liu G, Ma ZG, Shen XZ, et al. Standards of TW3 skeletal maturity for Chinese children. *Ann Hum Biol* 2008;**35**:349–54.

27. Zhang SY. *The standards of skeletal age in hand and wrist for Chinese: China 05 and its application*. Beijing: Science Press; 2015. Available at http://www.sciencep.com/.

28. Kim JR, Lee YS, Yu J. Assessment of bone age in prepubertal healthy Korean children: comparison among the Korean standard bone age chart, Greulich-Pyle method, and Tanner-Whitehouse method. *Korean J Radiol* 2015;**16**:201–5.

29. Wang YM, Tsai TH, Hsu JS, Chao MF, Wang YT, Jaw TS. Automatic assessment of bone age in Taiwanese children: a comparison of the Greulich and Pyle method and the Tanner and Whitehouse 3 method. *Kaohsiung J Med Sci* 2020;**36**:937–43.

30. Büken B, Şafak AA, Büken E, Yazici B, Erkol Z, Erzengin ÖU. Is the Tanner-Whitehouse (TW3) method sufficiently reliable for forensic age determination of Turkish children? *Turk J Med Sci* 2010;**40**:797–805.