JF Griffith
JCY Cheng 鄭振耀
E Wong 黃銘聰

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

Key Messages

1. Reference standards serve only as an estimation of skeletal age. Considerable physiological variation in hand and wrist bone maturation exists and therefore bone age estimates should not be applied too rigidly.
2. There is reasonable concordance between the Tanner and Whitehouse (TW3) [carpal score or radius, ulna, short-bone (RUS) score] standards and Greulich and Pyle (G&P) standards across the applicable age range.
3. The most notable deviation is that G&P standards tend to underestimate chronological age (by about 1 year) for boys aged 3 to 8 years.
4. The current practice of using G&P or RUS skeletal age standards to assess bone age in Hong Kong should continue. Greater accuracy could be achieved by adjusting designated G&P or RUS standards for different age-groups and genders.

Hong Kong Med J 2007;13(Suppl 3):S28-32

The Chinese University of Hong Kong, Shatin, Hong Kong:
Department of Diagnostic Radiology and Organ Imaging
JF Griffith
Department of Orthopaedics and Traumatology
JCY Cheng
Centre for Epidemiology and Biostatistics, School of Public Health
E Wong

HSRF project number: 932006

Principal applicant and corresponding author:
Prof James F Griffith
Department of Diagnostic Radiology and Organ Imaging, Prince of Wales Hospital, Shatin, NT, Hong Kong SAR, China
Tel: (852) 2632 2270
Fax: (852) 2636 0012
E-mail: griffith@ruby.med.cuhk.edu.hk

Introduction

Skeletal age standards provide a measure of the expected degree of skeletal maturation for a given chronological age. As skeletal maturation occurs earlier in girls than boys, a different set of standards apply to each gender. Skeletal/bone age should normally closely parallel chronological (biological) age. In several paediatric medical conditions, skeletal maturation may be accelerated or retarded, leading to appreciable discrepancy between skeletal and chronological ages.

Skeletal age is used to evaluate the effect of chronic disease on skeletal maturation, to time orthopaedic interventions, to predict adult height, and to monitor the effects of hormonal replacement. Skeletal age is also used to assign a chronological age to young persons whose birth date is uncertain.

There are several ways of evaluating skeletal maturity. The two most commonly applied methods currently used worldwide were devised by Greulich and Pyle (G&P) and Tanner and Whitehouse (TW).[1,2] The G&P method (developed in the 1930s) compares a radiograph of the patient's left wrist and hand to an atlas of reference radiographs of subjects aged 6 months to 18 years.[1] It is easy to use and allows bone age to be assigned in a few minutes. The second (TW2) method was developed in the United Kingdom in the 1950s. It scores the radiographic maturity of 20 individual bones in the hand and wrist[2]; the sum of such scores allows assignment of the overall skeletal maturity. Although more time-consuming (each assessment taking about 10 minutes), it is considered more accurate and reliable than the G&P method. To reflect the trend towards more rapid skeletal maturation in many countries, in 2001 Tanner et al[2] published new reference values based on American and European data obtained between the 1960s and the 1990s. These updated TW2 reference values were termed TW3[2] and emphasised the radius, ulna, short-bone (RUS) score rather than the carpal score (based on all carpal bones except the pisiform) or 20-bone maturity score (which is a composite of RUS and carpal methods), since "in most circumstances, the RUS score is all that is required".[2] A RUS score is quicker to obtain than a TW2 (ie 20-bone) maturity score. However, the applicability of these reference standards to modern day Hong Kong inhabitants is unclear.

Aims and objectives

This study was designed to (1) assess whether skeletal age standards currently used are applicable to modern day Hong Kong children/adolescents, and (2) to compare the G&P and TW3 methods of skeletal age assessment with reference to modern day Hong Kong children/adolescents.

Methods

This study was conducted from October 2000 to December 2002. The hand and wrist radiographs of 1020 Chinese subjects aged 18 years or less, attending the

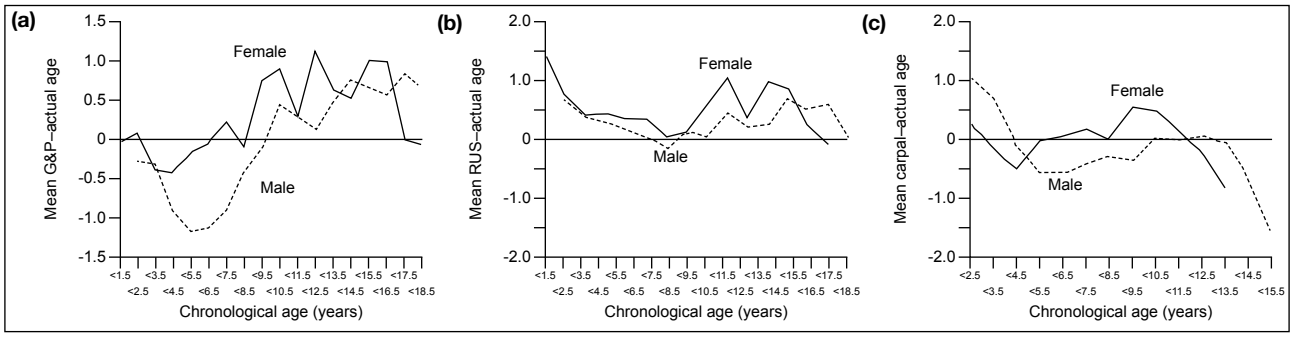

Fig. Differences in mean bone age estimation by using different standards: (a) Greulich and Pyle (G&P), (b) Tanner and Whitehouse (TW3) radius, ulna, and short-bone (RUS) scores, and (c) TW3 carpal scores

Zero baseline represents no difference between mean designated age and mean chronological age for different age-groups

accident and emergency departments of Prince of Wales, North District, and Tuen Mun hospitals with suspected fractures were reviewed. Patient age, gender, and race were obtained from the radiological records. Age and sex distribution are shown in the Table. Radiographs of both the left or right sides were used (as good side-to-side symmetry exists).[1,2] Only radiographs adequately displaying the necessary growth plates and ossification centres were used. Skeletal maturation was assessed by TW3 (RUS and carpal scores) and G&P methods. All skeletal score estimations were undertaken by one experienced rater, without prior knowledge of the patient's age. Films from 20 different subjects were assessed simultaneously with images being displayed on a single large viewing panel. For TW3 scoring, the radius was scored first on each of the 20 images followed by the ulna and remaining hand and wrist bones until all 20 bones were scored. Estimations by G&P were made at the same session. When bone developmental status was intermediate between two G&P standards, an age proportional to the relative likeness of the two most closely resembling standards was applied.[1]

Statistical methods

The estimated number of subjects overall, and for each subgroup the number needed to yield a reasonably reliable estimate of skeletal maturity was determined using the method described by Healy and based on reported variations of each subgroup.[1,3] Accuracy of individual scoring systems (G&P, RUS, and carpal vs chronological age) was evaluated by the paired *t* test. Statistical tests were performed using the Statistical Package for Social Sciences (Windows version 11.0; SPSS Inc, Chicago [IL], US) with a level of significance set at 5%.

Results

Accuracy of individual scoring systems

For both females and males, respective mean G&P skeletal age (Table, Fig a) tended to be underestimates of chronological age during early years of life (up to age 5 and 9 years) but tended to be overestimates in older subjects (from about 6 and 10 years to 18 years). For females and to a lesser extent males, mean RUS score (Table, Fig b) tended to overestimate chronological age at almost all ages (up to about 16 years). Mean female carpal scores (Table, Fig c) tended to minimally underestimate chronological age in younger subjects (up to 5.5 years), minimally overestimate for those of intermediate age (4.5-11.5 years). Almost the reverse was true for males.

Comparative accuracy of different scoring systems

For females, G&P and carpal scores tended to underestimate, while RUS scores tended to overestimate chronological age in early childhood, with RUS being significantly ($P<0.0001$) less accurate in this age-group (up to and including 5 years). Thereafter, there was no difference in the two methods. For males, TW3 methods were significantly more accurate ($P<0.0001$) in those up to and including 8 years, whereas carpal scores were significantly more accurate ($P<0.05$) in late childhood. For both sexes, carpal and RUS scores were not applicable after the age of 12 and 16 years respectively.

Discussion

Skeletal maturation is one indicator of biological maturation that is often monitored by measuring 'bone age' or the level of maturation of particular bones (usually in the hands and wrists). Bone age studies are applied clinically to the investigation of unduly tall or short children, assessment of hormone replacement, the timing of orthopedic and orthodontic operations, and for attributing the likely chronological age for persons whose birth date is unknown.

Although bone/skeletal age standards are useful, they are merely imprecise estimates of chronological age.[1] Like height and weight, skeletal maturation varies considerably within the normal population, though unlike height and weight, it usually attains the same end-point in all adults. Due to the inherent variability in the maturation rate of normal children and observer variation in designating particular standards, skeletal maturation assessments should not be too rigidly applied.[1] However, more than two standard deviations beyond the designated G&P or TW3 age would

Table. Chronological age and designated skeletal age according to the Greulich and Pyle (G&P) standards, Tanner and Whitehouse (TW3) radius, ulna, and short-bone (RUS) scores and carpal scores, and the differences between these measures for different age-groups

Sex	Age-group (years)*	G&P standards						
		n	Chronological age (years)		G&P skeletal age (years)		G&P – chronological age (years)		P value
			Mean	SD	Mean	SD	Mean	SD	
F	<1.5	8	1.09	0.30	1.06	0.32	-0.03	0.51	0.869
	<2.5	19	2.11	0.23	2.18	0.48	0.06	0.40	0.510
	<3.5	39	3.05	0.27	2.65	0.59	-0.41	0.58	0.000
	<4.5	49	3.96	0.28	3.52	0.83	-0.43	0.74	0.000
	<5.5	32	4.88	0.29	4.71	1.02	-0.17	1.02	0.342
	<6.5	30	6.02	0.29	5.96	1.29	-0.07	1.23	0.774
	<7.5	28	6.97	0.29	7.20	1.07	0.22	1.04	0.270
	<8.5	29	7.94	0.28	7.83	1.29	-0.11	1.31	0.659
	<9.5	27	8.92	0.30	9.65	1.26	0.73	1.23	0.005
	<10.5	24	10.05	0.23	10.95	1.20	0.90	1.22	0.001
	<11.5	26	10.94	0.28	11.24	0.76	0.30	0.85	0.083
	<12.5	14	12.01	0.22	13.13	1.13	1.12	1.06	0.002
	<13.5	7	12.95	0.08	13.57	0.53	0.62	0.48	0.014
	<14.5	9	14.04	0.33	14.56	1.01	0.51	0.93	0.134
	<15.5	8	15.00	0.22	16.00	1.07	1.00	1.03	0.029
	<16.5	7	15.86	0.30	16.86	0.69	0.99	0.50	0.002
	<17.5	7	17.01	0.30	17.00	0.82	-0.01	0.79	0.979
	<18.5	3	17.74	0.21	17.67	0.58	-0.07	0.39	0.780
M	<2.5	17	2.16	0.21	1.88	0.72	-0.28	0.67	0.100
	<3.5	48	3.10	0.21	2.78	0.70	-0.32	0.66	0.001
	<4.5	49	3.90	0.29	2.99	0.76	-0.91	0.83	0.000
	<5.5	58	4.93	0.27	3.75	0.98	-1.18	0.92	0.000
	<6.5	43	5.99	0.33	4.86	1.16	-1.13	1.11	0.000
	<7.5	45	6.98	0.27	6.08	1.29	-0.90	1.23	0.000
	<8.5	56	8.01	0.30	7.61	1.60	-0.40	1.54	0.057
	<9.5	53	8.96	0.26	8.84	1.40	-0.12	1.39	0.540
	<10.5	50	9.97	0.28	10.41	1.50	0.44	1.50	0.042
	<11.5	59	10.95	0.32	11.21	1.20	0.26	1.12	0.076
	<12.5	49	11.95	0.25	12.08	1.08	0.13	1.01	0.379
	<13.5	48	12.93	0.25	13.40	1.40	0.47	1.35	0.021
	<14.5	27	13.97	0.27	14.72	0.96	0.75	0.95	0.000
	<15.5	12	14.94	0.28	15.58	1.56	0.64	1.49	0.164
	<16.5	19	15.96	0.32	16.53	1.33	0.56	1.36	0.086
	<17.5	8	16.93	0.32	17.75	0.46	0.82	0.38	0.000
	<18.5	9	18.02	0.27	18.67	0.50	0.64	0.64	0.017

* <2.5 refers to the group between 1.5 and 2.5 years, while <3.5 between 2.5 and 3.5 years, and so forth. Few girls over the age of 12.5 years and few boys over the age of 14.5 years are required to test the standard, as there is little individual variation at these ages

make it probable that skeletal maturation was abnormally advanced or retarded.[1] A recommended 'normal range' for the G&P standards is given[1] and similar constraints apply to TW3 standards.[2]

In the current study, 1020 radiographs were scored. This sample size would be expected to yield, on average, a RUS score of ±2.7%, a carpal score of ±3.4%, and a G&P score of ±3.8% of the expected mean at each age.[3]

Greulich and Pyle standards

The G&P standards, the most commonly applied form of skeletal ageing in everyday practice, were determined from radiographs obtained from Cleveland, Ohio between 1931 and 1942. All subjects were of "somewhat above average social class"; each radiograph was selected from a subset of 100, that was considered representative of that particular age and sex.[1] Moreover, the resulting standards are considered easy to use, have acceptable accuracy, and be reproducible (0.02-0.04 being the intra- and inter-observer trainer error rates).[4] This study shows that for the most part, G&P standards for Hong Kong Chinese fairly closely resemble chronological age, although for both sexes they underestimate chronological age in early and middle childhood and overestimate it in late childhood and early adolescence. Thus, skeletally, Hong Kong children appear to mature more slowly in the first decade but more quickly thereafter, in keeping with trends reported in Turkish boys, Dutch Caucasians, and a previous study in local children.[5,6]

Tanner and Whitehouse standards

The TW reference standards were devised based on study of some 3000 normal British boys and girls examined longitudinally until the age of 12 years during the 1950s,[2] and was based on the premise that individual wrist and hand bones mature at different rates. Maturation scores are therefore applied to individual bones rather than to an entire group. To reflect changing trends towards more rapid maturation in many countries, in 2001 Tanner et al published a new set of reference data termed TW3.[2] The previously

TW3 RUS scores								TW3 carpal scores							
n	Chronological age (years)		RUS skeletal age (years)		RUS – chronological age (years)		P value	n	Chronological age (years)		Carpal skeletal age (years)		Carpal – chronological age (years)		P value
	Mean	SD	Mean	SD	Mean	SD			Mean	SD	Mean	SD	Mean	SD	
1	1.20		2.62		1.42										
17	2.14	0.23	2.92	0.42	0.78	0.41	0.000	11	2.19	0.24	2.46	0.35	0.27	0.35	0.029
39	3.05	0.27	3.47	0.38	0.42	0.41	0.000	28	3.07	0.28	2.91	0.68	-0.16	0.70	0.243
49	3.96	0.28	4.39	0.72	0.44	0.60	0.000	48	3.96	0.29	3.46	0.98	-0.50	0.89	0.000
32	4.88	0.29	5.24	0.80	0.35	0.78	0.015	32	4.88	0.29	4.87	0.98	-0.01	0.92	0.963
30	6.02	0.29	6.37	0.97	0.35	0.89	0.042	30	6.02	0.29	6.07	1.20	0.04	1.18	0.839
28	6.97	0.29	7.02	1.07	0.05	1.05	0.815	28	6.97	0.29	7.14	0.96	0.17	0.93	0.346
29	7.94	0.28	8.09	1.11	0.14	1.07	0.472	29	7.94	0.28	7.94	0.95	0.00	0.99	0.985
27	8.92	0.30	9.50	1.06	0.58	0.99	0.005	27	8.92	0.30	9.46	1.08	0.54	1.07	0.014
24	10.05	0.23	11.10	1.08	1.04	1.13	0.000	22	10.03	0.23	10.33	0.81	0.30	0.79	0.088
26	10.94	0.28	11.31	0.82	0.37	0.80	0.027	23	10.96	0.29	10.80	0.75	-0.15	0.84	0.393
14	12.01	0.22	12.98	1.03	0.97	0.98	0.003	10	11.97	0.21	11.28	0.60	-0.69	0.54	0.003
7	12.95	0.08	13.82	1.19	0.87	1.12	0.085	4	12.91	0.09	11.45	0.32	-1.46	0.27	0.002
9	14.04	0.33	14.28	0.74	0.24	0.68	0.322	8	14.11	0.28	10.98	0.75	-3.12	0.54	0.000
8	15.00	0.22	14.90	0.20	-0.10	0.29	0.341	4	14.88	0.22	11.39	0.50	-3.49	0.60	0.001
7	15.86	0.30	15.00	0.00	-0.86	0.30	0.000	3	15.74	0.12	11.36	0.38	-4.38	0.38	0.002
6	17.05	0.31	15.00	0.00	-2.05	0.31	0.000	4	16.88	0.26	11.19	0.62	-5.68	0.65	0.000
3	17.74	0.21	15.00	0.00	-2.74	0.21	0.002	1	17.51		10.93		-6.58		
12	2.22	0.21	2.88	0.63	0.66	0.60	0.003	4	2.08	0.30	3.13	0.53	1.05	0.62	0.043
48	3.10	0.21	3.48	0.82	0.38	0.82	0.002	25	3.09	0.21	3.78	0.71	0.69	0.66	0.000
49	3.90	0.29	4.20	0.71	0.30	0.68	0.003	23	3.93	0.30	3.84	0.85	-0.10	0.89	0.606
58	4.93	0.27	5.07	0.85	0.14	0.75	0.153	49	4.95	0.27	4.39	1.10	-0.56	1.06	0.001
43	5.99	0.33	6.06	0.93	0.07	0.94	0.632	42	5.99	0.34	5.42	1.08	-0.57	1.04	0.001
45	6.98	0.27	6.84	0.99	-0.14	0.94	0.327	45	6.98	0.27	6.55	1.29	-0.43	1.21	0.022
56	8.01	0.30	8.16	1.14	0.15	1.08	0.307	56	8.01	0.30	7.73	1.30	-0.28	1.25	0.102
53	8.96	0.26	9.00	1.24	0.04	1.27	0.817	53	8.96	0.26	8.59	1.15	-0.37	1.15	0.023
50	9.97	0.28	10.43	1.09	0.46	1.02	0.003	50	9.97	0.28	10.01	1.10	0.04	1.04	0.792
59	10.95	0.32	11.17	1.16	0.22	1.14	0.151	59	10.95	0.32	10.94	1.32	-0.01	1.24	0.965
49	11.95	0.25	12.21	1.05	0.27	1.00	0.068	49	11.95	0.25	12.01	1.19	0.06	1.16	0.725
48	12.93	0.25	13.63	1.25	0.69	1.20	0.000	44	12.93	0.26	12.68	1.07	-0.25	1.08	0.136
27	13.97	0.27	14.47	0.92	0.50	0.86	0.006	25	13.94	0.26	13.21	0.67	-0.74	0.70	0.000
12	14.94	0.28	15.54	1.36	0.59	1.34	0.153	9	14.95	0.29	12.74	0.89	-2.22	0.76	0.000
19	15.96	0.32	15.96	0.83	0.00	0.80	0.990	13	16.08	0.33	13.32	0.70	-2.76	0.70	0.000
8	16.93	0.32	16.44	0.12	-0.49	0.30	0.003	5	16.92	0.31	13.29	0.74	-3.63	0.72	0.000
9	18.02	0.27	16.48	0.04	-1.54	0.28	0.000	2	18.22	0.12	13.27	0.79	-4.95	0.67	0.061

used 20-bone score was abolished on the grounds that it had little benefit over the more-easy-to-apply RUS score. For 10- to 12-year-olds, the new TW3 reference values are about a year ahead of those described in TW2, but are less at odds with earlier ages. As these revised standards were published only recently, other reports on their applicability and accuracy are unavailable.

In a Hong Kong Chinese population, TW3 and RUS standards tend to resemble chronological age fairly closely up to early adolescence. As carpal maturation is achieved around the age of 13 years for girls and 15 years for boys, for older individuals the carpal score is not applicable.[2]

Recommendations

Maturation of modern day Hong Kong Chinese children differs from currently applied standards, but the differences are small relative to the expected variation of the normal range. The G&P standards were less accurate than RUS or carpal scores in males between 3 and 8 years, but were comparable to TW3 standards (ie RUS and carpal score) for all other ages. Based on this study, we recommend that the practice of using either RUS or G&P skeletal age standards to assess bone age in Hong Kong be continued. Score of RUS is more accurate in males aged 3 to 8 years, while G&P is easier to apply. Greater accuracy could be achieved by adjusting the designated G&P or RUS standards according to the Table and Figure. If using G&P standards, for example, for a boy with a chronological age of 5 years, 1.2 years (14 months) should be added to the assigned skeletal age. Similarly, if a girl has a chronological age of 13 years, 0.6 years (7 months) should be subtracted from the assigned skeletal age. If the chronological age of a boy is unknown, and the G&P standards suggest an age of 6 years, then the actual age is probably 7 years. We recommend that two bone ages be reported: (i) the designated RUS or G&P age and (ii) the adjusted RUS or G&P age as per the working examples above. In addition, the normal variation (two standard deviations) for any particular chronological age should also be reported.[1,2]

Acknowledgements

This study was supported by the Health Services Research Fund (#932006). We thank and appreciate the help of Mr Jonathan Lau, Research Assistant, for the conscientious way in which he undertook data collection and archiving. We also thank Dr MK Yuen for help in allowing access to films from Tuen Mun Hospital.

References

1. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. California: Stanford University Press; 1959.
2. Tanner JM, Healy MJ, Goldstein H, Cameron N. Assessment of skeletal maturity and prediction of adult height (TW3 method). 3rd ed. London: WB Saunders; 2001.
3. Healy HJ. Statistics of growth standards. In: Falker F, Tanner JM. Human growth. A comprehensive treatise: methodology. Ecological, genetic, and nutritional effects on growth. Vol 3. 2nd ed. New York: Plenum Press; 1986:47-58.
4. King DG, Steventon DM, O'Sullivan MP, 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.
5. van Rijn RR, Lequin MH, Robben SG, Hop WC, van Kuijk C. Is the Greulich and Pyle atlas still valid for Dutch Caucasian children today? Pediatr Radiol 2001;31:748-52.
6. Low WD, Chan ST, Chang KS, Lee MM. Skeletal maturation of Southern Chinese children. Child Dev 1964;35:1313-36.