Bone 170 (2023) 116725

Contents lists available at ScienceDirect

Bone

journal homepage: www.elsevier.com/locate/bone

Full Length Article

Evaluation of two methods of bone age assessment in peripubertal children in Zimbabwe

Farirayi Kowo-Nyakoko [a,b,c,*], Celia L. Gregson [d,e], Tafadzwa Madanhire [b], Lynda Stranix-Chibanda [f], Ruramayi Rukuni [b,g], Amaka C. Offiah [h], Lisa K. Micklesfield [e], Cyrus Cooper [a], Rashida A. Ferrand [b,g], Andrea M. Rehman [i], Kate A. Ward [a,e]

[a] MRC Lifecourse Epidemiology Centre, University of Southampton, Southampton General Hospital, Tremona Road, SO16 6YD Southampton, UK
[b] Biomedical Research and Training Institute, 10 Seagrave Road, Avondale Harare, Zimbabwe
[c] Department of Medical Physics and Imaging Sciences, University of Zimbabwe- Faculty of Medicine and Health Sciences, Parirenyatwa Group of Hospitals, Mazowe Street, Harare, Zimbabwe
[d] Musculoskeletal Research Unit, Bristol Medical School, University of Bristol, Southmead Hospital, Bristol BS10 5NB, UK
[e] SAMRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, University of the Witwatersrand, Johannesburg, South Africa
[f] Child and Adolescent Unit, University of Zimbabwe-Faculty of Medicine and Health Sciences, Parirenyatwa Group of Hospitals, Mazowe Street, Harare, Zimbabwe
[g] Clinical Research Department, London School of Hygiene and Tropical Medicine, London, UK
[h] Department of Oncology & Metabolism, University of Sheffield, Damer Street Building, Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield S10 2TH, UK
[i] MRC International Statistics and Epidemiology Group, Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

A R T I C L E I N F O

Keywords:
Age
Radiograph
Skeletal maturity
Africa
Children
Growth

A B S T R A C T

Objectives: Bone age (BA) measurement in children is used to evaluate skeletal maturity and helps in the diagnosis of growth disorders in children. The two most used methods are Greulich and Pyle (GP), and Tanner and Whitehouse 3 (TW3), both based upon assessment of a hand-wrist radiograph. To our knowledge no study has compared and validated the two methods in sub-Saharan Africa (SSA), and only a few have determined BA despite it being a region where skeletal maturity is often impaired for example by HIV and malnutrition. This study aimed to compare BA as measured by two methods (GP and TW3) against chronological age (CA) and determine which method is most applicable in peripubertal children in Zimbabwe.
Methods: We conducted a cross-sectional study of boys and girls who tested negative for HIV. Children and adolescents were recruited by stratified random sampling from six schools in Harare, Zimbabwe. Non-dominant hand-wrist radiographs were taken, and BA assessed manually using both GP and TW3. Paired sample Student *t*-tests were used to calculate the mean differences between BA and chronological age (CA) in boys and girls. Bland-Altman plots compared CA to BA as determined by both methods, and agreement between GP and TW3 BA. All radiographs were graded by a second radiographer and 20 % of participants of each sex were randomly selected and re-graded by the first observer. Intraclass correlation coefficient assessed intra- and inter-rater reliability and coefficient of variation assessed precision.
Results: We recruited 252 children (111 [44 %] girls) aged 8.0–16.5 years. The boys and girls were of similar mean ± SD CA (12.2 ± 2.4 and 11.7 ± 1.9 years) and BA whether assessed by GP (11.5 ± 2.8 and 11.5 ± 2.1 years) or TW3 (11.8 ± 2.5 and 11.8 ± 2.1 years). In boys BA was lower than CA by 0.76 years (95 % CI: −0.95, −0.57) when using GP, and by 0.43 years (95 % CI: −0.61, −0.24) when using TW3. Among the girls there was no difference between BA and CA by either GP [−0.19 years (95 % CI: −0.40, 0.03)] or TW3 [0.07 years (95 % CI: −0.16, 0.29)]. In both boys and girls, there were no systematic differences between CA and TW3 BA across age groups whereas agreement improved between CA and GP BA as children got older. Inter-operator precision was 1.5 % for TW3 and 3.7 % for GP ($n = 252$) and intra-operator precision was 1.5 % for TW3 and 2.4 % for GP ($n = 52$).

* Corresponding author at: Biomedical Research and Training Institute, 10 Seagrave Road, Avondale, Harare, Zimbabwe.
E-mail address: F.Nyakoko@soton.ac.uk (F. Kowo-Nyakoko).

https://doi.org/10.1016/j.bone.2023.116725
Received 27 October 2022; Received in revised form 27 January 2023; Accepted 28 February 2023
Available online 3 March 2023

Conclusion: The TW3 BA method had better precision than GP and did not systematically differ from CA, meaning that TW3 is the preferred method of assessment of skeletal maturity in Zimbabwean children and adolescents. TW3 and GP methods do not agree for estimates of BA and therefore cannot be used interchangeably. The systematic differences in GP BA assessments over age means it is not appropriate for use in all age groups or stages of maturity in this population.

1. Introduction

Bone age (BA) is a measure of the development of the skeleton incorporating the size, shape and degree of mineralization of the epiphyses and physeal plates of a bone to define their proximity to full maturity [1]. BA differs from chronological age (CA), which is calculated from the date of birth and does not necessarily reflect an individual's stage of puberty. For example, two eleven-year-olds can differ vastly in their developmental stage but CA is the same. Methods to assess BA should be reliable and practical, being adaptable for use in different populations. The two most widely used methods are the Greulich and Pyle (GP) and the Tanner and Whitehouse 3 (TW3) methods, both based upon assessment of hand/wrist radiographs [2,3]. The GP uses a technique which matches a given radiograph with an atlas of standards representing different sexes and ages and the radiograph is then assigned a skeletal age equal to that of the nearest age. In contrast, TW3 uses an individual bone specific scoring technique where the appearance of each bone is evaluated and compared to a set of bones at different stages of maturation. The GP method is more commonly used, as it is straightforward and quick, though it was founded on the assumption that the skeleton matures in a uniform way, assigning a global score to the hand and wrist [4]. GP was developed in 1959 based upon the left hand radiographs of a reference population of white North American children collected in the 1930's [3]. The TW method, first known as TW1 was also developed in the 1930's using radiographs of white European children, it was later modified to TW2 in 1983 using additional data collected in the 1950s and 1960s in western Europe. The TW1 and TW2 methods originally scored 20 bones (radius, ulna, short bones and carpal bones) of the hand before a further update in 2001, namely TW3, when scoring of the radius, ulna and short bones (RUS) was separated from the carpal bones [2]. The TW method is more complex and requires more time for assessment than GP (7.9 min and 1.4 min respectively) [5,6]. The bones of the hand ossify at different rates and the ability of TW3 to assess ossification centres separately makes it ideal for use hence being described as more objective and reproducible than GP [7].

The applicability of these methods in varying environments, populations with varying genetic ancestry, socio-economic status, and disease patterns is questionable as all these factors have an influence on the process of skeletal maturation. Sub-Saharan Africa (SSA) has a high prevalence of HIV and malnutrition, leading to delayed puberty. Given that these BA methods were developed decades ago in European and American populations, it is not known how applicable they are to African populations and previous studies are limited. None have compared the performance of the two methods, despite the fact that SSA is a region where skeletal development is commonly impaired [8,9]. Therefore, this study aimed to determine whether GP or TW3 is most applicable for use in peripubertal children in Zimbabwe. The agreement between the two methods, inter- and intra-rater reliability and intra- and inter-operator precision for each was determined.

2. Methods

This cross-sectional study used hand/wrist radiographs collected from IMVASK (The **IM**pact of **V**ertical HIV infection on child and **A**dolescent **SK**eletal development), a prospective cohort study conducted in Harare, Zimbabwe [10]. Children without HIV were recruited by stratified random sampling within three age and sex strata from six government schools within the same catchment area as the main public-sector hospitals in Harare (Parirenyatwa and Sally Mugabe hospitals) from where children with HIV were recruited. For the purposes of this methodological evaluation manuscript, participants were restricted to those testing negative for HIV. Participants were also excluded if they had CA above 15.0 years in girls and 16.5 years in boys because TW3 only assesses BA up to these ages.

Sociodemographic and clinical data were collected using a questionnaire administered by trained research staff. Anthropometric measurements included height (sitting and standing) and weight, taken in duplicate by two independent staff members. If the two measurements differed by >0.5 kg or 0.5 cm, a third reading was taken by an additional reader. The mean of the two or three measurements was recorded as the final figure. Weight was measured to the nearest 0.1 kg using a Seca 875 weight scale, and height to the nearest 0.1 cm using a stadiometer (Seca, Hamburg Germany). The equipment was calibrated annually over the 3 years of IMVASK.

A nurse and/or doctor carried out Tanner pubertal staging using testicular volume (assessed using an orchidometer) penile size and pubic hair growth (quality, distribution, and length) in boys and breast development (size and contour), age of menarche and pubic hair growth in girls for assessment [11]. Grading of penile, testicular and breast growth was from I to V as per tanner descriptions. In the event of disagreement of the pubertal stage testicular and breast development for boys and girls respectively were used to assign tanner stage. Socio-economic status was determined from indicators in a principal component analysis that included an asset list which showed the head of household age, highest maternal and paternal education levels, monthly household income, number in the household, house ownership, access to amenities and household item ownership list [12]. Socio-economic status was then split into three groups (low, middle, and high). Dietary calcium and vitamin D intake were assessed using a tool developed for the Zimbabwean context based on the validated dietary diversity and food frequency tool from India and Malawi, and international guidelines applicable to SSA [13,14]. The dietary intake was classified into three groups: daily dietary calcium intake as very low (<150 mg/d), low (150–299 mg/d), and moderate (300–450 mg/d) and daily dietary vitamin D intake as very low (<4.0 μg/d), low (4.0–5.9 μg/d), and moderate (6.0–8.0 μg/d). Physical activity was self-reported using the International physical activity questionnaire which has been validated in multiple countries. The amount of physical activity was assessed as multiples of the resting metabolic rate (MET) in MET minutes/week and categorised to low (MET minutes <600/week), moderate (MET minutes = 600 to 3000/week), and vigorous (MET minutes>3000/week) [15,16]. The same thresholds of physical activity were used in this study regardless of age.

Radiographs were taken of the non-dominant hand/wrist by a trained radiographer. A digital radiography (Siemens, Germany) system was used with images stored in DICOM format. The non-dominant hand was positioned with the palm facing downwards in contact with the imaging plate. The axis of the middle finger aligned with the axis of the forearm. The upper arm and forearm were in the same horizontal plane. The central ray of the X-ray beam was directed on the distal end of the third metacarpal and a tube to film distance of 76 cm was used. A lead apron shielded the gonads. Standard exposure parameters for a hand radiograph were used. Radiation dose to the participant was estimated to be less than 2 μSV (micro-Sieverts).

BA was assessed using both GP and TW3 (RUS) methods by a single radiographer. The TW3 (RUS) was based on comparison to a sexmatched TW3 atlas to determine the level of maturity for 13 selected regions of interest representing specific bones of the hand and wrist [2]. The specific bones included radius, ulna, metacarpals I, III, V; proximal phalanges I, III, V; middle phalanges III, V; distal phalanges I, III, V. The atlas categorises each bone into specific stages of development labelled from A to I. A numerical score is assigned to each stage of development for each individual bone in the atlas. The numerical scores were summed to give a total maturity score ranging from 0 to 1000. Each maturity score matches a certain skeletal age via a lookup table. The GP method of assessment involves comparison of the radiographs to the nearest matching sex-specific reference radiograph provided in the GP atlas [3]. The assessor was aware of the sex but not the CA of the children. Estimated BA was compared to CA, calculated from the date of birth to the date of examination of the radiograph.

To calculate intra-rater reliability and precision (co-efficient of variation), 52 (20 %) radiographs were randomly selected and rescored by the first assessor at least three months later than the initial score. All radiographs ($n = 252$) were re-scored by a second trained radiographer to determine inter-rater reliability and precision.

Ethics approval was granted by the Medical Research Council of Zimbabwe (Harare; reference MRCZ/A/2297) and the University of Southampton, UK (ERGO II 62773). Written informed consent was provided by the parents/guardian, and the children provided written assent.

2.1. Statistical analysis

Statistical analyses were conducted using Stata 16 (StatCorp, Texas, USA). Weight for age *Z*-scores (WAZ), height for age Z-scores (HAZ) and body mass index (BMI) Z-scores were calculated using the 1990 UK reference data for children [17] with a Z-score of −2.0 denoting stunting (HAZ) and underweight (WAZ).

The primary outcome was the difference between BA measured using two methods, and CA in years. Quantitative data were examined for normality using the Shapiro Wilkes test and histograms. Quantitative data are presented as mean ± standard deviation (SD) and categorical data as proportions and frequencies. Chi-squared tests were used to compare categorical variables by sex. Paired-student *t*-tests were used to compare quantitative variables in (1) the whole study population, (2) girls and boys separately, and (3) stratified by sex and age (in bands of 8–10, 11–13 and 14–16 years); this allowed determination of whether the performance of the BA method differed by maturity level. The mean difference was determined by subtracting CA from BA (BA-CA), presented with 95 % confidence intervals. A negative value indicates BA less than CA, and a positive value indicates BA greater than CA indicating delayed and advanced maturity respectively. Scatter plots for BA by both TW3 and GP method vs CA were examined with the line of best fit plotted.

Bland-Altman plots [18] were generated to assess agreement between CA and BA for each method, and between GP BA and the TW3 BA. Linear regression analysis was performed to determine the direction of the relationship between the average and differences in the methods, β coefficient and 95 % confidence intervals are presented. Inter- and intra-rater reliability were assessed using the intraclass correlation coefficient. The intra- and inter-operator coefficient of variation assessed BA reproducibility, i.e., precision. The formula used to calculate CV was (sample SD / mean) × 100 [19,20].

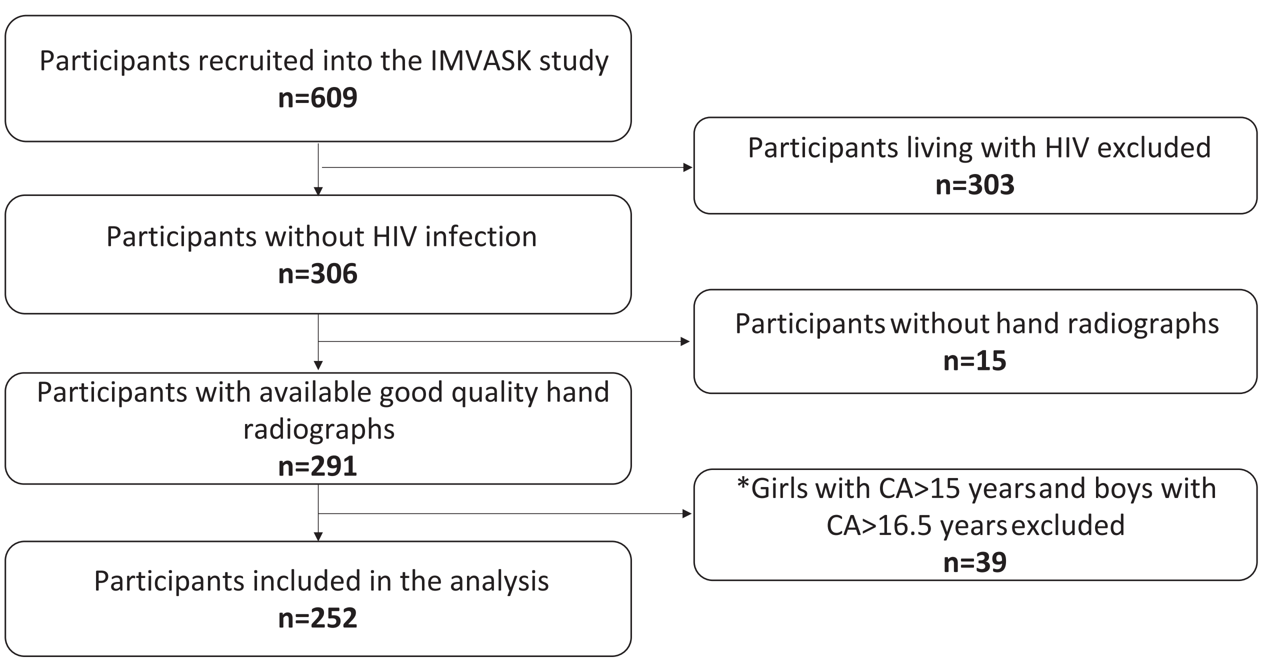
3. Results

3.1. Participant characteristics

Of the 306 children without HIV enrolled in IMVASK, 54 were excluded; 15 had missing hand radiographs, 33 girls were above 15 years of age, and 6 boys were above 16.5 years of age (Fig. 1).

In total 252 (82 %) were eligible for these analyses, with 141 (56 %) being boys. There was no difference in chronological age between the boys and girls, however mean WAZ and BMI-z scores was lower in boys than girls (Table 1).

Although there were no sex differences in BA measured using both methods more boys than girls had lower BA than CA for GP BA (77 % vs 57 %, Chi^2 $p = 0.001$), and TW3 BA (65 % vs 43 %, Chi^2 $p < 0.001$).



Footnotes: *IMVASK: The IMpact of Vertical HIV infection on child and Adolescent SKeletal development in Harare, Zimbabwe, CA: chronological age. *Threshold for TW3 analysis is at 15 years for girls and 16.5 years for boys*

Fig. 1. Flow chart showing participants included in the analysis.

Table 1
Demographic, lifestyle, anthropometry, and pubertal status characteristics of study participants by sex.

	Boys (n = 141)	Girls (n = 111)	p-Value
Socio-demographics			
Chronological age, mean (SD)	12.2 (2.4)	11.7 (1.9)	0.065
Socio-economic status, n (%)			0.361
Low: group 1	34 (24.1)	33 (29.7)	
Middle: group 2	52 (36.9)	32 (28.8)	
High: group 3	55 (39.0)	46 (41.4)	
Anthropometry			
Height-for-age Z-score, mean (SD)	−0.60 (1.03)	−0.47 (1.16)	0.327
Standing height-for-age Z-score < −2, n (%)	9 (6.4)	8 (7.2)	0.796
Sitting height-for-age Z-score, mean (SD)	−1.31 (1.06)	−1.32 (1.18)	0.920
Sitting height-for-age Z-score < −2, n (%)	31 (22.0)	30 (27.0)	0.354
Weight-for-age Z-score, mean (SD)	−0.69 (1.07)	−0.27 (1.22)	0.004
Weight-for-age Z-score < −2, n (%)	15 (10.6)	7 (6.3)	0.226
BMI Z-score, mean (SD)	−0.51 (1.05)	−0.07 (1.20)	0.002
BMI Z-score < −2, n (%)	11 (7.8)	4 (3.6)	0.162
Bone age measures			
GP bone age, mean (SD)	11.5 (2.8)	11.5 (2.4)	0.859
TW3 bone age, mean (SD)	11.8 (2.5)	11.8 (2.1)	0.945
GP bone age < chronological age, n (%)	108 (76.6)	63 (56.8)	0.001
TW3 bone age < chronological age, n (%)	92 (65.2)	48 (43.2)	<0.001
Pubertal status			0.078
Tanner 1	45 (32.1)	23 (20.9)	
Tanner 2	34 (24.3)	30 (27.3)	
Tanner 3	22 (15.7)	28 (25.5)	
Tanner 4	35 (25.0)	22 (20.0)	
Tanner 5	4 (2.9)	7 (6.4)	
Lifestyle factors			
Daily vitamin D intake, n (%)			0.991
Very low, <4.0 μg/day	17 (12.1)	14 (12.6)	
Low, 4.0–5.9 μg/day	92 (65.2)	72 (64.9)	
Moderate, 6.0–8.0 μg/day	32 (22.7)	25 (22.5)	
Daily calcium intake, n (%)			0.914
Very low, <150 mg/day	62 (44.0)	51 (45.9)	
Low, 150–299 mg/day	31 (22.0)	25 (22.5)	
Moderate, 300–450 mg/day	48 (34.0)	35 (31.5)	
Physical activity level, n (%)			0.304
Low, <600 MET min/week	47 (33.3)	45 (40.5)	
Moderate, 600–3000 MET min/week	46 (32.6)	27 (24.3)	
High, >3000 MET min/week	48 (34.0)	39 (35.1)	

Foot notes: Student t-tests conducted on continuous variables and chi-squared tests on categorical variable. GP-Greulich and Pyle, TW3-Tanner Whitehouse 3, SD-Standard deviation.

Socio-economic status, physical activity, vitamin D and calcium intake did not differ between the boys and girls. Nevertheless, 24 % of boys and 30 % of girls were in the low socio-economic status.

3.2. Comparison of CA and BA by TW3 and GP

There was a positive linear relationship between CA and BA for each method (GP: $r = 0.93$ boys, $r = 0.92$ girls; TW3: $r = 0.91$ boys, $r = 0.88$ girls) (Fig. 2). The slopes for the two methods in both boys [TW3 β 0.95 (95 % CI: 0.87), 1.03; GP β 1.09 (95 % CI: 1.01, 1.17)] and girls [TW3 β 0.92 (95 % CI: 0.81, 1.03); GP β 1.09 (95 % CI: 0.98, 1.20)] differed from each other. In the boys the TW3 slope ($p = 0.179$) does not differ from the line of identity unlike the GP slope ($p = 0.030$). In the girls both slopes [TW3 ($p = 0.180$) and GP ($p = 0.110$)] do not differ from the line of identity.

In the whole sample of boys GP BA was lower than CA by 0.76 years (95 % CI: −0.95, −0.57) while TW3 BA was lower than CA by 0.43 years (95 % CI: −0.61, −0.24) (Fig. 3). The mean difference between CA and GP BA was a year in those aged 8–10 years, with an offset of half a year earlier in the older group. In contrast, the mean difference between TW3 BA and CA, was less in the younger age group compared to the older age groups (Fig. 4a). Among the girls, GP BA had a trend to be lower than CA in the 8–10- and 11–13-year age groups, whereas no differences were seen between TW3 BA and CA (Fig. 4b).

There was a positive bias in agreement between CA and BA as measured by GP in boys (β = 0.18, $p < 0.001$) and girls (β = −0.08, $p = 0.050$), with GP BA being much lower than CA at younger ages and less of a difference at older ages (Fig. 3a, b). For TW3 BA and CA, no systematic differences were observed for either sex. The slope for GP Bland-Altman plots was 0.18 in boys and 0.16 in girls and for TW3 plots was 0.05 in boys and −0.08 in girls.

3.3. Comparison of GP and TW3

Overall, GP BA was lower than TW3 BA by 0.33 years (95 % CI: −0.88, −0.23) in the boys and 0.25 years (95 % CI: −0.44, −0.07) in the girls. In both boys and girls, Bland-Altman plots show differential agreement between the TW3 and GP across all ages (boys: β = −1.47 $p < 0.001$ and girls: β = −1.17 $p < 0.001$) (Fig. 5). Differences between the two methods in younger ages (8–10 years) show that TW3 BA was higher relative to the GP BA. At the older ages GP BA was higher than TW3 BA.

3.4. Intra- and Inter-rater reliability

The intraclass correlation coefficient showed high inter-rater reliability of 0.98 for TW3 and 0.94 for GP ($n = 252$) and high intra-rater reliability of 0.97 for TW3 and 0.98 for GP ($n = 52$). There was good precision for both methods although the TW3 mean percentage coefficient of variation were better than for GP at (intra-operator 1.5 % vs2.4 %) and (inter-operator 1.5 % vs 3.7 %) respectively.

4. Discussion

To our knowledge, this is the first comparison of bone age assessment methods in a population of peripubertal children in sub-Saharan Africa. These data show that TW3 is a more applicable method, than GP, for assessment of skeletal maturity in children and adolescents from Zimbabwe. TW3 has better intra- and inter-operator precision, and no age-related bias. In contrast the age-related bias using the GP BA method, with greater differences at younger ages, indicates limitations for use in our study population. TW3 and GP methods did not agree in younger children and cannot be used interchangeably.

On evaluating the two bone age assessment methods, Bland-Altman analyses show that the two assessments differ especially in the young children, preventing interchangeable use. Similar results were found in a study in the UK where GP and TW2 were compared [21], and TW2 was recommended as the preferred method as it had better precision than GP. Otherwise, few studies have compared both methods [22–25]. Many more studies have looked at the applicability of the GP than TW3 method, potentially because TW3 is used less frequently in clinical practice due to time constraints [5,6]. Despite the TW3 taking longer to assess (7.9 min vs 1.4 min) [6], the consistency across age groups and sexes, and good intra- and inter-rater reliability, suggest that the extra time is justified. The bones of the hand ossify at different rates and the ability of the TW3 method to assess ossification centres separately makes it an ideal method for use.

The current study has demonstrated that both methods are reproducible with the TW3 being more reproducible than the GP method which is consistent with other studies [21,25–28], potentially explained by individual maturity indicator scoring, rather than the GP global grading approach. GP BA units are whole numbers – less precise than the TW3 approach with decimals. Cavallo et al. reported the evaluation of each bone segment in TW3 as an advantage minimising inter-rater variability [29]. Our study also found high inter- and intra-rater agreement supporting the reliability of these two methods in assessing

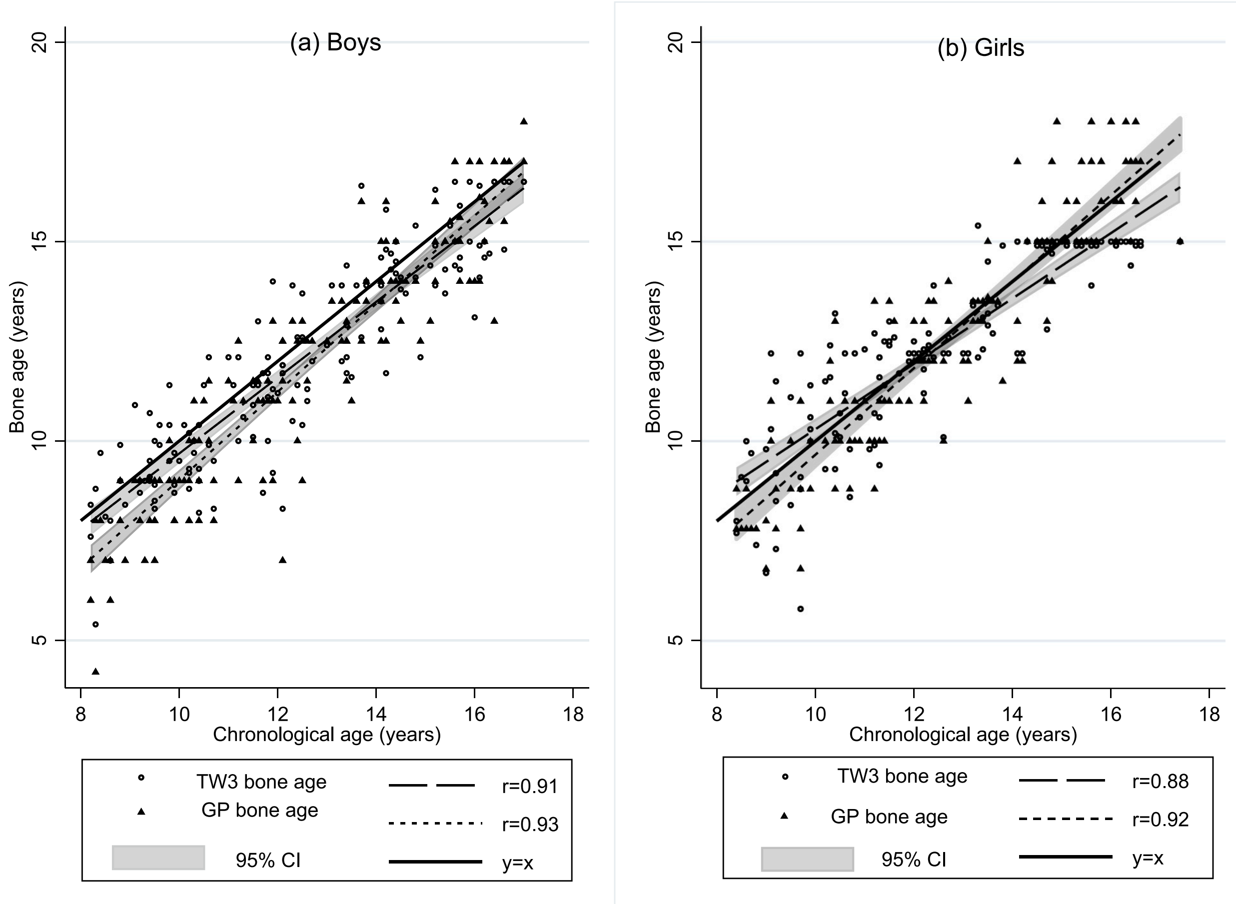


Fig. 2. Scatter plot showing the relationship between Chronological age and bone age by GP (Greulich and Pyle) and TW3 (Tanner Whitehouse 3) for (a) boys and (b) girls.

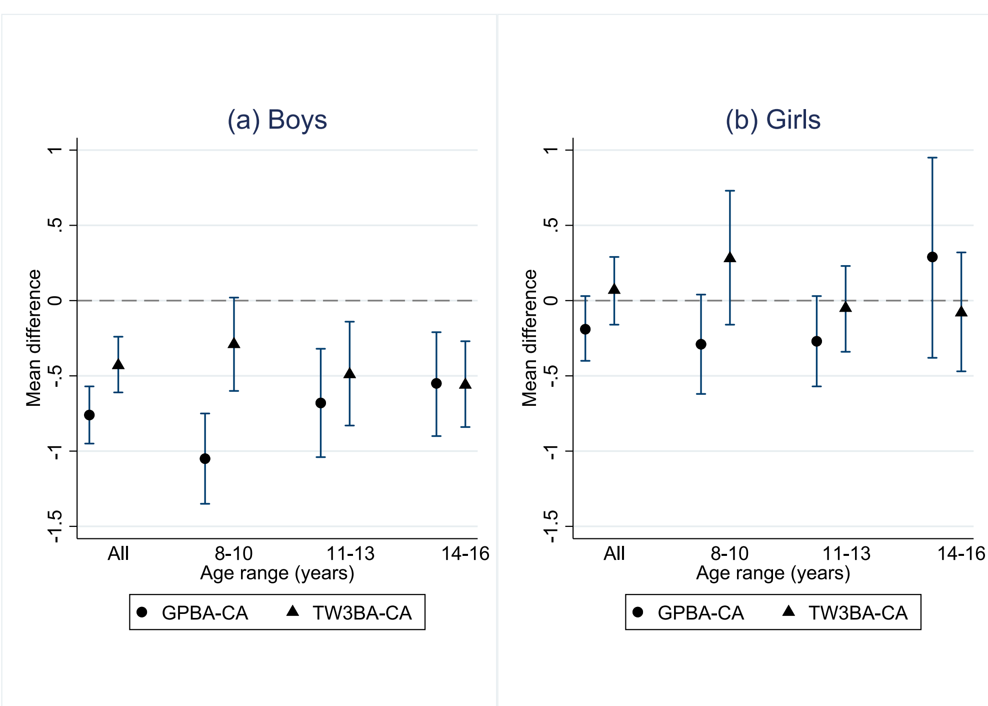


Fig. 3. Forest plot showing the mean differences (95 % Confidence Intervals) between CA and BA assessed by GP, and TW3. Foot notes: Student paired t-tests conducted on all continuous variables, CA-chronological age, GPBA-Greulich and Pyle bone age, TW3BA-Tanner Whitehouse 3 bone age. A negative value means delayed relative to the CA and a positive value means advanced relative to the CA. Sample size for age groups is Boys: 8–10 (n = 49); 11–13 (n = 49); 14–16 (n = 42) and girls: 8–10 (n = 44); 11–13 (n = 48); 14–16 (n = 20).

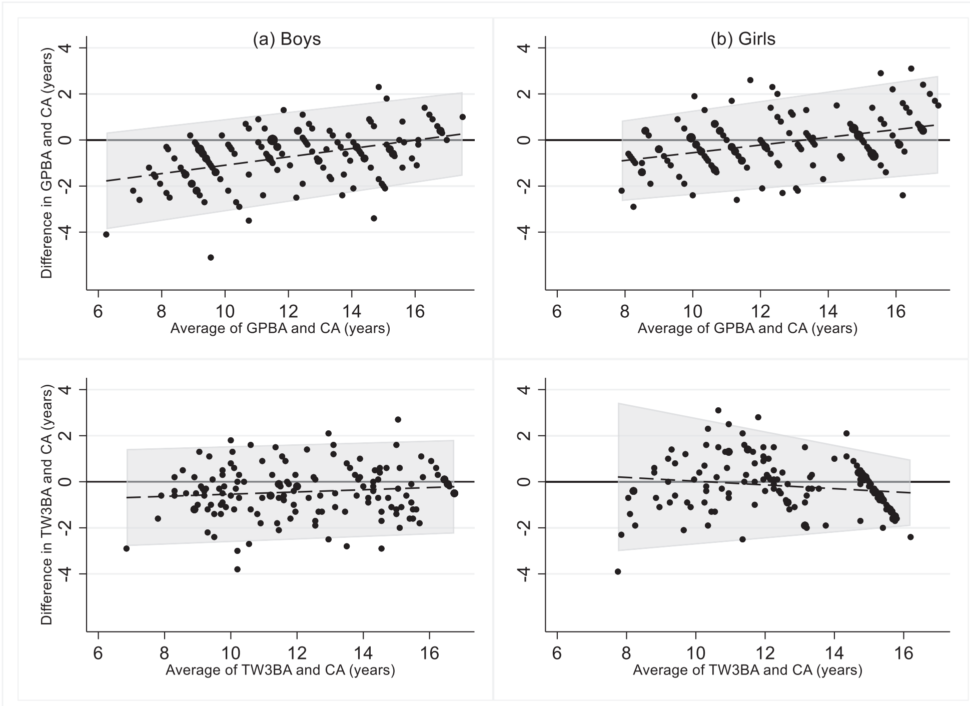


Fig. 4. Bland Altman plots showing mean differences between GPBA and CA (top) TW3BA and CA (bottom) and for (a) boys and (b) girls. Variance of the difference in measures was: top left 1.32; bottom left 1.24; top right 1.31 and bottom right 1.35.

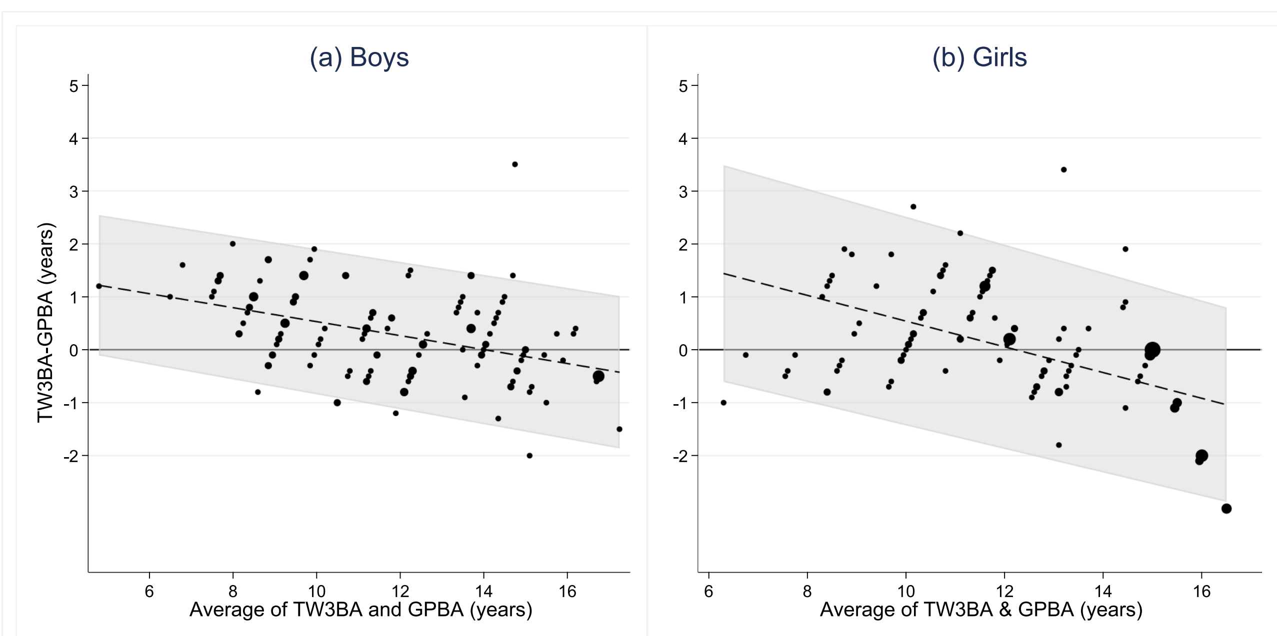


Fig. 5. Bland Altman plot showing mean differences between GP and TW3 bone age in years for (a) boys and (b) girls.

bone age, though notwithstanding the limitations attached to the use of GP in our population given the age dependence of the results.

One limitation of the TW3 method is that it has a cut off point for skeletal maturity of 15 years for girls and 16.5 for boys. In populations that may have later onset, or delayed, puberty, such as in our current study, this limits use, given skeletal maturity is not reached until at least 18 years for girls and 19 years for boys [3].

In the largest study of African children to date ($n = 607$), TW3 was used to assess and compare BA between black and white girls and boys in South Africa. Skeletal maturity was delayed by on average 7 months in black boys compared to white boys, whereas no ethnic differences in the progression of skeletal maturation were seen in girls [30]. Similarly, our data shows sex differences in the performance of TW3, suggesting on average 6 months delay in BA in the boys compared to CA and no differences in the girls.

Consistent with our findings, previous studies have commonly demonstrated lower GP BA than CA, in South Africa, Saudi Arabia, Turkey and Malawi [22,24,31,32]. Our findings are similar to those from a study in Pakistan which advised against using GP method in their population due to variations in BA depending on the age of the children (advanced BA during early childhood and delayed BA during middle and late childhood) [33]. Our study extends this work by testing the agreement between GPBA and CA and showing the GP method was dependent upon age, such that one would not be able to accurately determine whether a child is delayed in maturity using GP. Several other studies have found the applicability of GP differing by sex, ethnic group or by age [22,31,34].

The recruitment of the children from schools in Harare was designed to be representative of the general population in Zimbabwe, as about 98 % of the children in the whole country go to school. All schools in Harare were identified and the 6 studied were randomly chosen from this sampling frame [10]. Bone age assessment was repeated by a second assessor further strengthening the method employed. It is well-described that timing of skeletal maturation is affected by environment, nutrition, medication, and socioeconomic status [4,35,36]. Our population therefore contrasts with the ones in which the GP and TW3 methods were originally developed. The children used to create the GP atlas were from wealthy families and not suffering from any illnesses [3]. Generally, HAZ and WAZ were low in our study population, indicating a tendency towards stunting and underweight respectively. Most participants had low or middle socio-economic status. Calcium and vitamin D intakes were low, both micronutrients being critical for skeletal development [37]. Together these factors may explain why children in Zimbabwe were skeletally immature for their chronological age.

5. Conclusion

Of the two methods assessed, the TW3 is the most precise and appropriate for use in Zimbabwe as GP appears biased by age. The two methods (TW3 and GP) do not give equal estimates of bone age and therefore should not be used interchangeably. Hence the TW3 method would be the preferred method for use in determining skeletal maturity in peripubertal children in the region. This study is of relevance for low-income African populations in SSA who are faced with a high burden of communicable and non-communicable diseases and delays in skeletal maturation.

Funding

PhD funding support from the University of Southampton Global Challenges Research Fund and Faculty of Medicine, and part-funding from UK Medical Research Council Program U105960371. RAF is funded by the Wellcome Trust (206316_Z_17_Z). The IMVASK study was funded by the Wellcome Trust through a fellowship to RR (206764/Z/17/Z). For the purpose of open access, the author has applied a CC BY public copyright licence (where permitted by UKRI, 'Open Government Licence' or 'CC BY-ND public copyright licence'

CRediT authorship contribution statement

FK-N KAW CLG: conceptualisation, methodology, interpreting data, writing original draft, FK-N Investigation ACO LS-C RR RAF LKM AMR Writing-review and editing, CC KAW CLG AMR LS-C supervision, validation, AMR FK-N KAW CLG TM: formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Professor Ann Prentice for her support in the development of this work.

References

[1] J.M. Tanner, Growth at Adolescence, 1962.
[2] J.M. Tanner, R. Whitehouse, N. Cameron, W. Marshall, M. Healy, H. Goldstein, Assessment of skeletal maturity and prediction of adult height (TW2 method): Saunders London, 2001.
[3] W.W. Greulich, S.I. Pyle, Radiographic Atlas of Skeletal Development of the Hand and Wrist, Stanford University Press, 1959.
[4] K. Alshamrani, F. Messina, A.C. Offiah, Is the greulich and pyle atlas applicable to all ethnicities? A systematic review and meta-analysis, Eur. Radiol. 29 (6) (2019) 2910–2923.
[5] D. King, D. Steventon, M. O'sullivan, A. Cook, V. Hornsby, I. Jefferson, 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. 67 (801) (1994) 848–851.
[6] A.M. Mughal, N. Hassan, A. Ahmed, Bone age assessment methods: a critical review, Pakistan J. Med. Sci. 30 (1) (2014) 211.
[7] K. Khan, A.S. Elayappen, Bone growth estimation using radiology (Greulich–Pyle and Tanner–Whitehouse methods), in: Handbook of Growth and Growth Monitoring in Health and Disease, Springer, 2012, pp. 2937–2953.
[8] S.M. Arpadi, Growth failure in children with HIV infection, J. Acquir. Immune Defic. Syndr. 2000 (25) (1999) S37–S42.
[9] M. De Onis, M. Blössner, E. Borghi, Prevalence and trends of stunting among pre-school children, 1990–2020, Public Health Nutr. 15 (1) (2012) 142–148.
[10] R. Rukuni, C. Gregson, C. Kahari, F. Kowo, G. McHugh, S. Munyati, et al., The IMpact of vertical HIV infection on child and adolescent SKeletal development in Harare, Zimbabwe (IMVASK Study): a protocol for a prospective cohort study, BMJ Open 10 (2) (2020).
[11] P.M. Duke, I.F. Litt, R.T. Gross, Adolescents' self-assessment of sexual maturation, Pediatrics 66 (6) (1980) 918–920.
[12] S. Vyas, L. Kumaranayake, Constructing socio-economic status indices: how to use principal components analysis, Health Policy Plan. 21 (6) (2006) 459–468.
[13] S. Filteau, A.M. Rehman, A. Yousafzai, R. Chugh, M. Kaur, H. Sachdev, et al., Associations of vitamin D status, bone health and anthropometry, with gross motor development and performance of school-aged Indian children who were born at term with low birth weight, BMJ Open 6 (1) (2016), e009268.
[14] Infant WGo, Indicators YCF, Developing and Validating Simple Indicators of Dietary Quality and Energy Intake of Infants and Young Children in Developing Countries: Summary of Findings From Analysis of 10 Data Sets, FANTA, Project Washington, DC, USA, 2006.
[15] M. Hagströmer, P. Oja, M. Sjöström, The international physical activity questionnaire (IPAQ): a study of concurrent and construct validity, Public Health Nutr. 9 (6) (2006) 755–762.
[16] C.L. Craig, A.L. Marshall, M. Sjöström, A.E. Bauman, M.L. Booth, B.E. Ainsworth, et al., International physical activity questionnaire: 12-country reliability and validity, Med. Sci. Sports Exerc. 35 (8) (2003) 1381–1395.
[17] T.J. Cole, J.V. Freeman, M.A. Preece, British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood, Stat. Med. 17 (4) (1998) 407–429.
[18] D. Giavarina, Understanding bland altman analysis, Biochem. Med. 25 (2) (2015) 141–151.
[19] O. Shechtman, The coefficient of variation as a measure of sincerity of effort of grip strength, part II: sensitivity and specificity, J. Hand Ther. 14 (3) (2001) 188–194.
[20] C.-C. Glüer, G. Blake, Y. Lu, B. Blunt, M. Jergas, H. Genant, Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques, Osteoporos. Int. 5 (4) (1995) 262–270.
[21] R. Bull, P. Edwards, P. Kemp, S. Fry, I. Hughes, Bone age assessment: a large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods, Arch. Dis. Child. 81 (2) (1999) 172–173.
[22] K. Alshamrani, A. Hewitt, A. Offiah, Applicability of two bone age assessment methods to children from Saudi Arabia, Clin. Radiol. 75 (2) (2020) 156, e1–e9.
[23] K. Alshamrani, A.C. Offiah, Applicability of two commonly used bone age assessment methods to twenty-first century UK children, Eur. Radiol. 30 (1) (2020) 504–513.
[24] B. Büken, Ö.U. Erzengin, E. Büken, A.A. Şafak, B. Yazıcı, Z. Erkol, Comparison of the three age estimation methods: which is more reliable for Turkish children? Forensic Sci. Int. 183 (1–3) (2009) 103, e1–e7.
[25] V. Pinchi, F. De Luca, F. Ricciardi, M. Focardi, V. Piredda, E. Mazzeo, et al., Skeletal age estimation for forensic purposes: a comparison of GP, TW2 and TW3 methods on an Italian sample, Forensic Sci. Int. 238 (2014) 83–90.
[26] S. Mora, M.I. Boechat, E. Pietka, H. Huang, V. Gilsanz, Skeletal age determinations in children of European and African descent: applicability of the Greulich and Pyle standards, Pediatr. Res. 50 (5) (2001) 624–628.
[27] R.R. van Rijn, M.H. Lequin, S.G. Robben, W.C. Hop, C. van Kuijk, Is the Greulich and Pyle atlas still valid for dutch Caucasian children today? Pediatr. Radiol. 31 (10) (2001) 748–752.
[28] A. Maggio, A. Flavel, R. Hart, D. Franklin, Assessment of the accuracy of the greulich and pyle hand-wrist atlas for age estimation in a contemporary Australian population, Aust. J. Forens. Sci. 50 (4) (2018) 385–395.

[29] F. Cavallo, A. Mohn, F. Chiarelli, C. Giannini, Evaluation of bone age in children: a mini-review, Front. Pediatr. 9 (2021) 21.
[30] T.J. Cole, E.K. Rousham, N.L. Hawley, N. Cameron, S.A. Norris, J.M. Pettifor, Ethnic and sex differences in skeletal maturation among the birth to twenty cohort in South Africa, Arch. Dis. Child. 100 (2) (2015) 138–143.
[31] K.A. Dembetembe, A.G. Morris, Is greulich-pyle age estimation applicable for determining maturation in male Africans? S. Afr. J. Sci. 108 (9) (2012) 1–6.
[32] C. Lewis, C. Lavy, W. Harrison, Delay in skeletal maturity in Malawian children, J. Bone Jt. Surg. Br. Vol. 84 (5) (2002) 732–734.
[33] A.M. Zafar, N. Nadeem, Y. Husen, M.N. Ahmad, An appraisal of Greulich-Pyle Atlas for skeletal age assessment in Pakistan, JPMA J. Pak. Med. Assoc. 60 (7) (2010) 552.
[34] R.T. Loder, D.T. Estle, K. Morrison, D. Eggleston, D.N. Fish, M.L. Greenfield, et al., Applicability of the greulich and pyle skeletal age standards to black and white children of today, Am. J. Dis. Child. 147 (12) (1993) 1329–1333.
[35] A. Schmeling, W. Reisinger, D. Loreck, K. Vendura, W. Markus, G. Geserick, Effects of ethnicity on skeletal maturation: consequences for forensic age estimations, Int. J. Legal Med. 113 (5) (2000) 253–258.
[36] H.A. Delemarre-van de Waal, Environmental factors influencing growth and pubertal development, Environ. Health Perspect. 101 (suppl 2) (1993) 39–44.
[37] W. Koo, N. Walyat, Vitamin D and skeletal growth and development, Curr. Osteoporos. Rep. 11 (3) (2013) 188–193.