Original Article

Development and Validation of a Novel Mobile Application for Bone Age Assessment

Patel R; Bajpai A*; Mendpara H; Dave C; Mehta S; Mendpara P; Shukla R**

Department of Pediatric Endocrinology, Regency Center for Diabetes Endocrinology & Research and Department of Pediatric Critical Care, Regency Hospital Limited, Kanpur, India

***Corresponding author: Bajpai A**

Department of Pediatric Endocrinology, Regency Center for Diabetes Endocrinology & Research, Regency City Clinic, Opposite PPN Market, Kanpur 208001, India.
Tel- +919454081769
Email: dranurag.bajpai@gmail.com

Abstract

Background: Despite its pivotal role, the use of bone age is limited due to a lack of physician expertise, time, and access.

Objective: To develop and validate a mobile application-based bone age assessment tool.

Study Design: The study involved the selection of standardized images (from 307 radiographs), delineation (90 radiographs), evaluation (200 radiographs) and incorporation of regions of interest on the application, and validation against the Tanner-Whitehouse 3 method by an expert (252 radiographs) and non-expert users (110 radiographs).

Results: The application-based assessment by expert users had an absolute standardized difference of 4.7 months (95% confidence interval; 4.2-5.2 months), a relative standardized difference deviation of 4.4% (3.9-4.9%), and similar intraindividual [2.8 (1.9-3.7) months versus 3.6 (2.8-4.5) months] and interindividual variation [4.2 (3.3-5.0) versus 4.2 (3.3-5.2) months] compared to the TW3 method. Non-expert user assessment had an absolute standardized difference of 6.7 months (5.6-7.7 months) and a relative standardized difference deviation of 5.8% (4.7-6.9%). The mean test time was lower for application than TW3 for both expert [1.3 (0.2) versus 3.1 (1.0) minutes, p<0.001] and non-expert users [2.8 (1.0) versus 5.0 (1.4) minutes, p<0.001].

Discussion: Our study confirms the accuracy of mobile application-based bone age assessment. This, along with good precision, reduced complexity, and lower time requirement, suggests a potential for its widespread implementation.

Keywords: Bone age; Mobile application; Tanner whitehouse 3; Validation

Introduction

Bone age assessment is integral to Pediatric Endocrine evaluation [1]. Despite its significance, the use of bone age is limited in pediatric practice due to a lack of physician expertise, time, and accessibility. Bone age assessment methods use automated (BoneXpert) or manual comparison of the entire non-dominant hand (Greulich-Pyle) or specific regions of interest (Tanner-Whitehouse III method) with age and gender-specific standards [2-4].

Manual bone age assessment methods compare the whole radiograph (holistic approach; GP atlas) or a combination of regions of interest (analytical approach; TW3) to age and gender standards. The epiphyseal maturation of regions of inter-est varies in an individual making the perfect match between radiographs challenging. The analytical approach has higher precision than the holistic approach, as indicated by lower intra-observer variability for the TW method than the GP atlas [5].

These methods are complicated by the need for multiple sites that increase procedure time and variability. The need for morphological grading of each region of interest makes the TW3 method challenging for physicians with limited exposure to bone age assessment. Inconsistent interpretations of the capitate, hamate, the first distal, and fifth middle phalanx, part of the TW3 RUS method, have been reported [6]. Including epiphysis from similar regions contributes little to the diag-

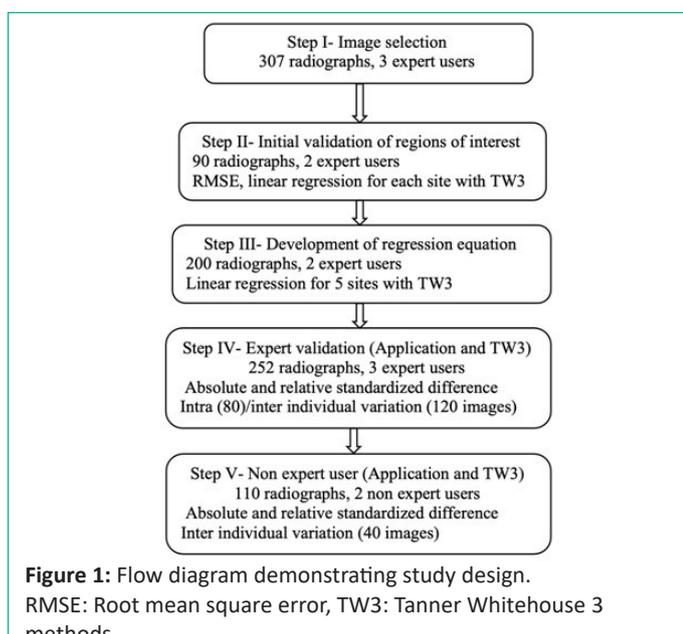**Table 1:** Predictive accuracy of most appropriate sites predicting TW3 bone age.

Region of interest	RMSE	Standardized coefficient (Beta)	p-value
Proximal phalanx of the middle finger	1.1	0.24	<0.001
Radius	0.9	0.22	<0.001
Ulna	2.2	0.22	<0.001
Proximal phalanx of thumb	1.1	0.21	<0.001
Middle phalanx of the middle finger	1.2	0.13	0.028
Distal phalanx of the little finger	1.4	0.233	0.086
Distal phalanx of the middle finger	1.3	-0.126	0.283
Metacarpal of thumb	1.3	0.033	0.530
Distal phalanx of thumb	1.4	0.081	0.354
Middle phalanx of the little finger	1.5	0.015	0.890
Proximal phalanx of the little finger	1.2	0.093	0.462
Metacarpal of the middle finger	1.2	0.117	0.357
Metacarpal of the little finger	1.3	0.043	0.678
Top 5 sites	0.68	1.09	<0.001
All 13 sites	1.32	1.41	<0.001

nostic accuracy of a method while increasing its complexity. A higher number of regions of interest increases the test time and complexity of the method; making a reduction in the number of sites desirable. This highlights the need for an accessible tool allowing rapid bone age assessment with reduced complexity.

The availability of smart mobile phones provides an opportunity to allow point-of-care bone age assessment. We have developed mobile application tools guiding the evaluation and management of children with short stature and diabetic ketoacidosis [7,8]. We, therefore, aimed to develop a mobile application-based tool with a reduced number of regions of interest to allow simplified, reproducible, and valid bone age assessment.

Material and Methods

The study involved the selection of age and gender-specific standard images, delineation, evaluation, and incorporation of regions of interest on the mobile application, development of the regression equation for bone age interpretation, validation against the gold standard, and comparison with the automated method of bone age assessment (Bone Xpert) (Figure 1).

Figure 1: Flow diagram demonstrating study design.
RMSE: Root mean square error, TW3: Tanner Whitehouse 3 methods

The anonymized radiographs of the left hand and wrist of children and adolescents presenting to the Pediatric Endocrine Clinic of our hospital were accessed after approval from the Institutional Ethics Committee. Radiographs with poor quality, improper hand orientation, and bone anomalies (skeletal dysplasia or metabolic bone disease) were excluded. Pediatric Endocrinologists with experience in bone age assessments (Expert users) selected and validated the images.

Image Selection

Three Expert users (AB, CD, RP) selected six-monthly age and gender-specific images of thirteen TW3 RUS regions of interest (metacarpal, proximal, middle, and distal phalangeal epiphysis of middle and little fingers; distal, proximal phalangeal, and metacarpal epiphysis of thumb; radial and ulnar epiphysis) from 307 radiographs [152 boys; age 10.9 (3.1) years, 2-18 years].

Site Selection

Two Expert users (CD, RP) assessed the bone age of additional 90 radiographs [52 boys; age 9.8 (3.4) years, 2-18 years] for each region of interest and TW3 method. The predictive accuracy of individual regions of interest for TW3 measurements was assessed using Root Mean Square Error (RMSE) and linear regression. The combined predictive value of the regions of interest found significant on linear regression was compared to that of all sites for six-monthly and annual images.

All regions of interest showed reasonable predictive accuracy for the TW3 method (RMSE 1.1 to 2.2, Table 1). Five sites (middle and proximal phalanx of middle finger, proximal phalanx of thumb, radius, and ulna) had significant predictive value on linear regression. The predictive accuracy of the combination of these five sites for the gold standard was similar to that for all 13 sites (RMSE 0.68 as against 1.32; R^2 96.1% as against 96.4%). Annualized images had better diagnostic accuracy than six-monthly images (RMSE 0.9 as against 1.5, p 0.04). Therefore, annualized images of these five regions of interest were included for further analysis.

Development of Equation

The equation for bone age computation from these regions was developed using linear regression for TW3 assessment in an additional set of 200 radiographs [97 boys; 9.9 (3.2) years, 2-18 years] by two expert users (RP, SM). The five regions of interest explained 96.1% variation in the TW3 results for 200 distinct images on linear regression. The regression equation was used to compute mobile application-based bone age.

Development of the Mobile Application

The panel of gender-specific annualized images of the five selected regions of interest was loaded on the mobile application platform. The mobile application displays unlabeled annualized images of these regions of interest after entering gender and date of birth. The user selects the image closest to the test radiograph. The application provides instantaneous bone age readings (2-15 years in girls and 2-16.5 years in boys) based on the image selection and the pre-loaded regression equation (Figures 2A & 2B).

Validation

Three expert users (RP, CD, SM) validated the results of the mobile application against the gold standard TW3 method on 252 novel sets of radiographs (143 boys; 10.1 (3.4), 2-16 years
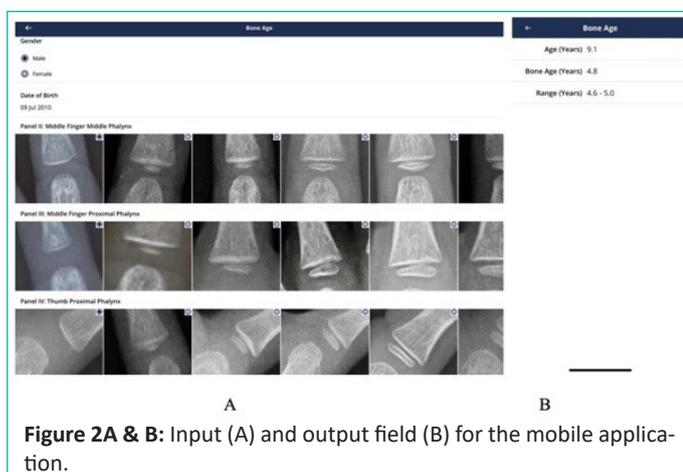
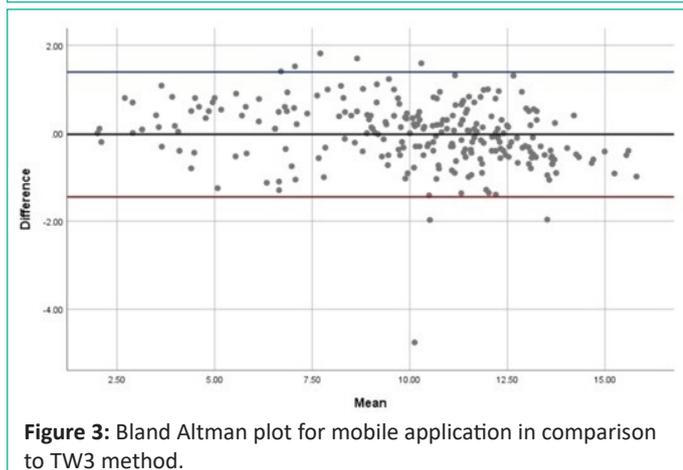**Figure 2A & B:** Input (A) and output field (B) for the mobile application.

Figure 3: Bland Altman plot for mobile application in comparison to TW3 method.

of age). The precision and accuracy were assessed using absolute and relative standardized differences and Bland Altman Plot. Intraindividual variation was measured in 80 radiographs assessed twice by the same user, while the interindividual variation was studied on 120 radiographs reported by two expert users. The experts were blinded for the radiographs and were provided images for assessment on different days. The gender of the individual was the only information disclosed to the raters with no disclosure about chronological age or diagnosis.

Validation in Non-Expert Users

The diagnostic accuracy of the application amongst non-expert users was determined on 110 radiographs [50 boys, 9.9 (2.1), 2-18 years] assessed by two physicians with no prior exposure to bone age assessment. Their bone age assessments by mobile application and TW3 method were compared with TW3 bone age readings of expert users.

Statistical Analysis

The data was analyzed using IBM Statistical Package for Social Sciences (SPSS version 25.0, SPSS, Inc., Chicago, IL, USA) for Macintosh. Data is expressed as mean (standard deviation) or mean (95% confidence interval). Accuracy and precision were assessed by root mean square error, absolute and relative standardized difference, and the Bland-Altman plot. Linear regression analysis was used to rank regions of interest in order of their predictive value and develop regression equations to calculate the bone age. The time taken by both expert and non-expert users in assessing bone age by the mobile application and the TW3 method was compared using the Student's t-test.

Results

The mean mobile application assessed bone age by expert users on 252 novel radiographs was similar to that of the gold standard TW3 method [10.1 (2.9) as against 10.1 (3.1) years, p=0.517). The mobile application had an absolute standardized difference of 4.7 months (4.2-5.2 months) and a relative standardized difference deviation of 4.4 % (3.9-4.9%) compared to the TW3 method. The Bland-Altman plot showed agreement between the Mobile application and the gold standard results (Figure 3). The mobile application-based bone age assessment had similar intraindividual [2.8 (1.9-3.7) as against 3.6 (2.8-4.5) months, p=0.14] and interindividual variation [4.2 (3.3-5.0) versus 4.3 (3.3-5.2) months, p=0.87] compared to the TW3 method.

The bone age application assessment by non-expert users (HM, PM) had an absolute standardized difference of 6.7 months (5.6-7.7 months) and a relative standardized difference deviation of 5.8% (4.7-6.9%). Interindividual variability was similar for mobile application and TW3 methods [6.7 (5.7-7.6) versus 6.5 (5.4-7.6) months, p=0.84]. The mean time for bone age assessment was lower for mobile application than TW3 method for both expert [1.3 (0.2) as against 3.1 (1.0) minutes, p<0.001], and non-expert users [2.8 (1.0) as against 5.0 (1.4) minutes, p<0.001].

Discussion

The findings of our study confirm the accuracy of mobile application-based bone age assessment for both expert and non-expert users. This, along with similar precision, lower time requirement than the gold standard, and offline availability across mobile platforms, makes it ideal for widespread bone age assessment across clinical settings.

The absolute standardized difference for expert users in our mobile application (4.7 months) is lower than that reported with manual (GP and TW3, 5-10 months), automated (Bone Xpert, 8.4-8.5 months), and deep learning-based artificial intelligence methods (TW3I, 6 months) indicating good diagnostic accuracy [5-6,9,10]. High intraindividual variation has been a cause of concern with manual methods (3-10 months) [5]. The intraindividual variation for our mobile application is lower than that reported for manual measures (TW3 and GP 2.9 months) and at par with automated methods (Bone Xpert, 2.1 months) [5,11]. The absolute standardized difference of 6.7 months for non-expert users aligns with automated measures and indicates a high potential for implementation in physicians with limited exposure to bone age assessment.

Including epiphysis from similar regions contributes little to the diagnostic accuracy of a method while increasing its complexity. Only five of the 13 regions included in the TW3 RUS method had significant predictive value in our study. These represent four distinct bone groups (thumb, radius, ulna, and middle finger). The reduction in the region of interest (from 13 to five) lowered the test time by 60% compared to the TW3 method without affecting precision. Inconsistent interpretations of the capitate, hamate, the first distal, and fifth middle phalanx, part of the TW3 RUS method, have been reported [6]. These sites did not achieve statistical significance in our study and were, therefore, not included in the application.

The study's findings may not be applicable to general pediatricians; good diagnostic accuracy in non-expert users, however, suggests its generalizability for the setting. Mobile application-based bone age assessment represents a rapid, accurate, and precise tool that can be implemented across resource settings.

More studies across settings and comparisons with other methods are needed to confirm this potential.

Author Statements

Acknowledgements

Ethics Approval

Institutional ethics committee, Regency Hospital**,** RHL-IEC-16039, September 11, 2019.

References

1. Agarwal N, Bajpai A. Bone age assessment. In: Bajpai A, Dave C, Agarwal N, Patel R, editors. MedEclasses basic pediatric endocrinology. Kanpur: Grow Society Publications. 2019; 45-8.

2. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. IEEE Trans Med Imaging. 2009; 28: 52-66.

3. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. Stanford, CA: Stanford University Press. 1959.

4. Tanner JM, Healy MJR, Goldstein H, Cameron N. Assessment of skeletal maturity and prediction of adult height (TW3 method). London, New York: W B Saunders. 2001.

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

6. Zhou XL, Wang EG, Lin Q, Dong GP, Wu W, Huang K et al. Diagnostic performance of convolutional neural network-based tanner-Whitehouse 3 bone age assessment system. Quant Imaging Med Surg. 2020; 10: 657-67.

7. Patel RV, Bajpai AT, Mendpara HV, Dave CC, Mehta SS, Dixit S et al. Development and validation of a mobile application for point of care evaluation of growth failure. J Pediatr Endocrinol Metab. 2022; 35: 147-53.

8. Mendpara H, Bajpai A, Patel R, Shukla R, Kapoor R. The development and validation of a mobile application for guidance for management of severe diabetic ketoacidosis. Indian J Pediatr. 2022; 89: 1251-6.

9. Thodberg HH, Sävendahl L. Validation and reference values of automated bone age determination for four ethnicities. Acad Radiol. 2010; 17: 1425-32.

10. Martin DD, Deusch D, Schweizer R, Binder G, Thodberg HH, Ranke MB. Clinical application of automated Greulich-Pyle bone age determination in children with short stature. Pediatr Radiol. 2009; 39: 598-607.

11. Thodberg HH. Clinical review: an automated method for determination of bone age. J Clin Endocrinol Metab. 2009; 94: 2239-44.