Determination of bone age and bone mass in modern United Kingdom and Saudi Arabian populations.

By

Khalaf Alshamrani

Thesis submitted for the degree of Doctor of Philosophy

Department of Oncology and Metabolism
Medical School

Abstract

The Greulich & Pyle (G&P) and Tanner & Whitehouse (TW) methods are frequently used to determine bone age. The applicability of these methods for populations who are of different ethnicity or socioeconomic status to the reference standard has been subjected to questions. Therefore, a systematic review was undertaken to evaluate the applicability of the G&P atlas for four major ethnicities. The G&P standard appeared imprecise when applied to Asian male and African female populations.

The applicability of the G&P and TW3 to modern population from the United Kingdom and Saudi Arabia was assessed. The automatic software called (BoneXpert) which calculates bone age beside on the G&P and TW3, was used. The software can eliminate observer variability and provide timesaving solution. In total 821 hand radiographs (426 males) were included on the analysis. In the UK, the G&P atlas appeared to be applicable while The TW3 consistently underestimates the age of females by an average of 5 months. Furthermore, significant differences between BA and CA were apparent in Saudi Arabian males when using the G&P atlas and TW3 method.

The added advantage of BoneXpert is that bone mass can be assessed from left hand radiographs. However, results from 291 patients, in which their dual energy x-ray absorptiometry (DXA) and hand radiographs have been acquired on the same day, showed weak correlation between DXA and bone mass calculated by the software.

The BoneXpert performance with regard to images taken using modalities other than conventional radiography was evaluated. Nevertheless, another advantage is that children are more likely to expose to much lower radiation dose from hand-wrist DXA compare to left hand radiographs. The low quality of DXA prohibits the use of BoneXpert software for the automatic determination of bone age while the TW3 cannot be determined manually from the hand-wrist DXA.

Acknowledgements

In the name of Allah, the most Gracious, the most Merciful.

Firstly, I would like to express my sincere gratitude to my primary supervisor Dr Amaka C Offiah for the continuous support during my PhD study, for her patience, motivation, and encouragement. Her guidance helped me in all the time of research and writing of this thesis. I would like also to thank Professor Nick Bishop, and all the co-authors for their support and useful suggestion.

I take this opportunity to thank the endocrine clinic team at Sheffield Children's hospital. Special thanks to Dr Charlotte Elder, Dr Neil Wright, Dr Nils Krone and Dr Paul Dimitri for their help in patients' recruitment (Chapter 6).

I also place on record, my sincere of gratitude to my brothers and sisters for their care. My thanks go also to my colleagues and friends; Hassan, Mohammad, Fawaz and Hani who were the best companions in Sheffield. I truly appreciate and would like to sincerely thank Najran University for giving me the scholarship opportunity and funding me PhD study.

Dedication

This thesis is dedicated to my parents Mr. Awod and Mrs. Shareefa Alshamrani for their love, encouragement and prayers. I also dedicate this thesis to my wife and my two daughters for their endless support throughout the years. I could not have completed this effort without their assistance, patience, tolerance, and enthusiasm.

The Author

My background includes a BSc in Radiography Science from King Saud University and an MSc in Radiography from the Cardiff University.

Table of Contents

List of publications arising from this thesis

Accepted:

Alshamrani K, Messina F & Offiah A. Is the Greulich and Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis. Accepted (*European Radiology*). DOI: 10.1007/s00330-018-5792-5

Alshamrani K, Messina F, N Bishop & Offiah A. Estimating bone mass in children: can bone health index replace dual energy x-ray absorptiometry? (*Pediatric Radiology*) DOI: 10.1007/s00247-018-4309-3. (Available online from: https://link.springer.com/article/10.1007/s00247-018-4309-3)

Under review:

Alshamrani K, Offiah A. Applicability of Two Commonly Used Bone Age Assessment Methods to 21st Century UK Children. (Submitted to *European Radiology*)

Alshamrani K, Hewitt A, Offiah A. Applicability of Two Bone Age Assessment Methods to Children from Saudi Arabia. (Submitted to *Clinical Radiology*)

Alshamrani K, Elzene K, Offiah A. Bone Age Determination using Dual-Energy X-ray Absorptiometry. (Submitted to *Skeletal Radiology*)

Conferences' Abstracts:

Alshamrani K,Offiah A. Applicability of the Greulich and Pyle bone age atlas to the United Kingdom children born in the 21st century. Pediatr Radiol (2018) 48(Suppl 2):299. http://doi.org/10.1007/s00247-018-4151-7.

Alshamrani K, Russell J, Bishop N & Offiah. A comparison of bone health index and bone mineral density as indicators of bone mass in children. Pediatr Radiol (2018)48 (Suppl2):299. http://doi.org/10.1007/s00247-018-4151-7.

Alshamrani K, Offiah A. Applicability of the Greulich and Pyle bone age atlas to Children in Saudi Arabia. Pediatr Radiol (2018) 48(Suppl 2):299. http://doi.org/10.1007/s00247-018-4151-7.

Alshamrani K, Messina F, Bishop N, Offiah A. Estimating bone mass in children: a comparison between DXA and hand radiographs analysed using BoneXpert software. UKRC 2018 conference. 2018 July.

Alshamrani K, Offiah A. Bone age assessment using Greulich and Pyle and Tanner-Whitehouse methods: a systematic review. In the 8th International Conference on Children 2017 Jul 11 (Vol. 6). BioScientifica. DOI: 10.1530/boneabs.6.P080

Alshamrani K, Offiah A. Bone age determination using dual-energy X-ray absorptiometry. In8th International Conference on Children 2017 Jul 11 (Vol. 6). BioScientifica. DOI: 10.1530/boneabs.6.P081.

Alshamrani K, Kruger E, Offiah A. Developing a Computerised Bone Age Tool. UKRC 2016 conference. 2016 July.

List of Figures

List of Tables

Abbreviations:

BA	Bone age
BHI	Bone health index
BHI SDS	Bone health index standard deviation scores
DXA	Dual-Energy X-ray Absorptiometry
EV	External validity
G&P	Greulich & Pyle
IMD	Index of multiple deprivation
IV	Internal validity
NICE	The National Institute for Health and Care Excellence
RUS	Radius, ulna and short bone
TW1	Tanner Whitehouse 1 (published in 1962)
TW2	Tanner Whitehouse 2 (published in 1975)
TW3	Tanner Whitehouse 3 (published in 2001)
UK	United Kingdom

Definitions:

- Reliability: reliability refers to the overall consistency of a measure. Measures that have a high reliability usually produce similar results under consistent conditions. Diagnostic test that are highly reliable are accurate, reproducible, and consistent from one testing occasion to another. In Radiology, the reliability is classified into several class including the inter-rater (inter-observer) reliability which assesses the degree of agreement between two observers' results when they use the same type of measure. Furthermore, the other type is called intra-rater (intra-observer) which assesses the degree of the agreement among repeated diagnostic tests performed by the same single observer. However, reliability does not imply validity, which the later refers to the extent of which a diagnostic test is well-founded and corresponds accurately to what is measured.
- Accuracy: refers to the ability of a diagnostic test to discriminate between the target condition and health. This discriminative potential can be quantified by the measures of diagnostic accuracy in different ways such as sensitivity and specificity, predictive values, likelihood ratios and diagnostic odds ratio (Šimundić, 2009). Each of these measures is related to some specific aspect of the diagnostic procedure. For example, assessing the ability to detect or exclude a disease requires a different accuracy measure than assessing its predicative ability. Accuracy is different to the precision, which is an indication to how close diagnostic test results are to each other (Harper and Reeves, 1999). When repeated

analyses on the same sample produces similar results then the diagnostic test is deemed to have a high precision.

- Coefficient of variation (CV): the coefficient of variation is a measure of the availability with in a given population. In other words, it shows the extend of the variability in relation to the mean of the population. It is collected by dividing the standard deviation (SD) divided by the mean of the sample and multiplied by 100.
- Mean absolute error: is a measure of differences between two continues variables. Additionally, it measures the average size of the errors in a set of predictions, without considering their direction. To measure the differences between values which are predicted compare to the observed values, the root mean-square error is used. Therefore, both measures are related to quantifying the errors associated with a predicated model when compared to the observed values. However, the mean absolute error is usually preferred as it is easiness to interpret.
- Reproducibility: refers to the variations in measurements made on a subject under different conditions. These differences can be due to different instruments used, measures are produced by different observers or measures that have been taken at over a period of time. However, systematic bias can be associated with measurements that have been made by different observers. To test for the systematic bias and quantify for the agreement, Bland Altman is usually constricted as well as the use of Kappa coefficient which adjust the agreement for excepted chance of agreement.

CHAPTER 1

Introduction, Aims and Objectives

1.1 Introduction

Bone age (BA) determination is the estimation of the skeletal maturation of an individual in relation to healthy population. The BA is usually required to investigate whither the individual is experiencing delayed or an advancement in growth. The left hand radiograph, have been used widely to capture the change in skeletal maturation. In clinical practice, two methods are well known for BA determination; Greulich & Pyle (G&P) and Tanner Whitehouse method (TW). Most of the data which were used to establish the G&P atlas and the TW3 came from healthy children, who originally were from of North America and European origin and some of them lived 60 or 70 years ago. These methods have been subjected to criticism with regards to their applicability to a different population. Whenever a particular atlas is utilised to determine BA, the concern to be raised “Is it suitable to compare the bone age of this child who might be of different ethnicity or socioeconomic status to the standard?”

Nevertheless, the assessment of BA has some technical and methodological aspects that should be considered. One is the use of a subjective method, which is more likely to suffer from inter- and intra-observer variations. This has been overcome by the introduction of automated system called BoneXpert, which can automatically calculate BA based one G&P and TW method. The automated system has an advantage of calculate bone mass, derived from the cortical thickness of the three middle metacarpals and metacarpal width and length. This could be advantage, particular in those children required to capture the change in bone mass as a result of treatment. Another aspect is the radiation dose from left hand radiograph

involved in BA assessment. Children diagnosed with certain chronic diseases require regular BA monitoring, which means repeated hand radiographs numerous times throughout their childhood. Therefore, considering alternative modality which involves less radiation such as Dual Energy X-ray Absorptiometry (DXA) would be preferred.

1.2 Aims

This thesis aims to determine the applicability of two bone age assessment methods currently used in relation to modern populations using both manual and automated approaches. The thesis also aims to evaluate the relationship between bone mass calculated by BoneXpert to those measured by DXA.

1.3 Objectives

1- To systematically review and summarise the findings of the published literature in regard to the applicability of the G&P atlas to children and adolescents who are of a different population from the original standard.

2- To evaluate the applicability of the G&P and TW3 methods to UK children born in the 21st century, using an automated software programme, thereby eliminating any effect of observer variability.

3- To assess the applicability of the G&P and TW3 methods to children from Saudi Arabia.

4- To evaluate the use of BoneXpert to determine BA in children from Saudi Arabia, in which the software has not been validated.

5- To compare bone mass measured by BoneXpert and expressed as the bone health index with bone mineral density as measured by DXA.

6- To assess whether hand-wrist DXA can replace radiographs for BA assessment using the G&P and/or TW3 methods.

7- To assess the possibility of using the Bonexpert software to determine BA from hand-wrist DXA scans

1.3 Thesis structure

This thesis has been written and presented in the format of the scientific papers, which are suitable for publication in a peer-reviewed journal. This alternative PhD thesis format has been recently supported by The University of Sheffield to prepare PhD candidates for academic publishing. In fact, two chapters have been published while the other three chapters has been submitted to peer-reviewed journals. Additionally, the alternative format can reduce time when rewriting publications into thesis chapters and can improve the writing skills required in publication. Approval from faculty of Medicine, Dentistry and Health has been obtained to write and present this thesis in the alternative format as following:

Chapter Two: Literature review in relation to the need for bone age assessment and the commonest approaches used as well as factors that can affect skeletal development.

Chapter Three: Is the Greulich and Pyle Atlas Applicable to All Ethnicities? A Systematic Review and Meta-Analysis.

Chapter Four: Applicability of Two Commonly Used Bone Age Assessment Methods to 21st Century UK Children.

Chapter Five: Applicability of Two Bone Age Assessment Methods to Children from Saudi Arabia

Chapter Six: Bone Age Determination using Dual-Energy X-ray Absorptiometry

Chapter Seven: Estimating bone mass in children: can bone health index replace dual energy x-ray absorptiometry?

Chapter Eight: Overall discussion and conclusions.

CHAPTER 2

Literature review

2.1 The need for bone age assessment

The determination of bone age (BA) and understanding growth in children are critical for medical and psychological purposes. While the chronological age (CA) is the actual time in years and months starting from a child's date of birth, BA refers to the level of biological maturation of a child's bones. In clinical practice, BA is frequently requested to rule out any delay or advancement in terms of skeletal development of a child (Acheson R M, 1954). A significant delay in BA can be an indication of chronic illness, a constitutional delay of growth, growth hormone deficiency or malnutrition (David D Martin, Jan M. Wit, *et al.*, 2011). Indeed, when there is a delay, the child will require more time to reach the end of the growth process. BA determination is also important in children with congenital adrenal hyperplasia, as well as when monitoring the response of the skeleton to certain treatments, such as hydrocortisone (Speiser *et al.*, 2010). Moreover, BA may be required when planning for orthopaedic surgery e.g. leg lengthening or epiphyseal closure surgery to estimate the remaining years of growth (Moseley, 1977).

In forensic and legal contexts, BA is used to estimate CA in situations where the CA is undocumented or unable to be proven (Black, 2010). In such situations, the authorities need to judge whether a person is considered a child or an adult (6–8). In the United Kingdom (UK) as well as most other European countries, an individual is considered to be a child if under the age of 18 chronological years, where access to education, medical care and social care differ to an adult (Hm Government, 1989). With the increase number of immigrants around the world (United Nation, 2017), age

estimation becomes essential for those in which their CA cannot be prove. In Europe, it has been estimated that approximately 160,000 unaccompanied children entered between 2015 and 2016, of which approximately 2,300 were in the UK (Eurostat, 2017; Hm Government, 2018)

Although there is no precise figure available regarding the number of those children with a valid documented age, authorities have faced challenges in estimating some of their ages because many have lost their documents or may have falsified their age. Being 18 years of age is crucial in legal situations, but other ages are legally relevant to children, for instance, the age of criminal responsibility is deemed as 10, 13 and 14 years old in the UK, France and Germany, respectively (Baumann *et al.*, 2009).

The use of BA to estimate CA is a controversial area, with the European Society Pediatric Radiology stating, "*It is impossible to determine whether a person is over or under 18 years based on BA of the hand/wrist*" (European Society of Paediatric Radiology, 2018). The Royal College of Paedatrics and Child Health in the UK argued that there current methods are not reliable from making precise age estimation (Royal College of Paediatrics and Child Health, 2018). However, the Study Group for Forensic Age Estimation (Berlin, Germany) considered the radiographic examination of the left hand among the most suitable methods currently available beside the physical assessment stating that methods should be used to gather to increase the diagnostic accuracy (A Schmeling *et al.*, 2008). However, this accurate assessment requires an understanding of the factors that affect skeletal development, which impact on BA. These factors will be discussed in the next section.

2.2 Factors affecting BA determination

Many factors may affect growth and maturation, which can be divided into inherited factors, such as the sex and ethnicity of the individual and external factors, which include environmental aspects, nutrition and health status. All these factors influence the rate and timing of maturational stages, making it difficult to differentiate between them (Cameron, 1997). Additionally, some factors may change over time, which can affect skeletal maturity of both the individual and the population, so it is complicated to determine the relationship between BA and CA. When these factors impact on the population, the effect is known as 'secular change'. Johnston defined the secular change as "the change over time in the characteristic pattern of growth of the children of a population" (Johnston, Francis E., 2002). Therefore, the influence of secular change on the applicability of these reference standards cannot be neglected.

The factors influencing skeletal maturation will now be discussed. The effect of sex and ethnicity will be included as inherited factors, while the impact of the environment, nutritional intake and health status will be discussed under socioeconomic factors.

2.2.1 Inherited factors

The ossification centres on the hand are regarded as indicators for BA. The first presence and the fusion of these ossification centres differs between females and males, therefore, separate maturational standards have been developed for females and males (Greulich and Pyle, 1959). The difference in timing can range from weeks to years depending on the age of individual,

being weeks in young infants but increasing to months and years during middle childhood and adolescence. In general, females are more advanced in terms of maturational changes compared to males. Consequently, males have a longer period of growth that allows them to gain height and weight before growth cessation (Humphrey, 1998). Although sex has been found to influence the timing of the maturational changes, the actual pattern of these changes and the order of their appearance are not affected by the sex (Greulich and Pyle, 1959; Tanner *et al.*, 2001)

The influence of ethnicity on skeletal maturation rate has been studied widely. The word “ethnicity” has been a subject of much debate in health research. There is no precise definition for ethnicity, but it may indicate one or more of several factors including; shared origin, shared culture and tradition and similar physical or genetic characteristics (Senior and Bhopal, 1994; Afshari and Bhopal, 2002; Bhopal, 2004). When it comes to reporting ethnicity, it should be mentioned that studies have used different classifications and sub-divisions. In the UK for example, seven major ethnic groups have been suggested for use in health related research (Simpson and Akinwale, 2007)However, different parts of the world may have different views on how to classify ethnic groups (Crews and Bindon, 1991; Bhopal and Donaldson, 1998). This is important because one of the main criteria when reporting research is to produce results that can be compared easily to previous and future research. In relation to BA assessment, several studies have classified their research population into four major ethnic groups as follows; African, Asian, Caucasian and Hispanic (Loder, 1993; Ontell *et al.*,

1996; Schmeling *et al.*, 2000; A. Zhang, James W. Sayre, *et al.*, 2009; Thodberg and Sävendahl, 2010; Mansourvar *et al.*, 2014)

Numerous studies have assessed the variability in the maturation process of different populations, arguing that the differences between ethnic groups in maturation rate cannot be attributed to environmental factors alone, as ethnic groups living under similar conditions have showed differences in maturational rate (Loder, 1993; Ontell *et al.*, 1996; Schmeling *et al.*, 2000; A. Zhang, James W. Sayre, *et al.*, 2009). Loder et al. assessed the maturation rate from left hand radiographs in 841 children aged between 0 and 18 years living in Ohio, United States (US), of which 461 were of African descent . African females were skeletally advanced by between 0.4 to 0.7 years compared to Caucasian females, whereas males aged 3 to 7 years showed delayed BA of around 0.9 years, which was not observed in Caucasian males. Another study was conducted by Ontell et al. who assessed the skeletal maturation of four ethnic groups. The authors cautioned that the maturation rates were different among these four groups, particularly in African and Hispanic females as well as in Asian and Hispanic males, as BA tended to exceed CA by between 9 and 11 months. Furthermore, Zhang et al. concluded that BA of Asian and Hispanic children was advanced when compared to African-American and Caucasians, particularly between 10 and 15 years of age.

Although these studies did not report other factors that can affect the maturation rate, such as environmental factors, children within each study lived within the same geographical area, thereby supporting the concept that the ethnicity of individuals can influence the rate of skeletal maturation. This

was discussed by Sutow et al., who found that Japanese children living in Japan were skeletally delayed in comparison to the Caucasian children who lived in Cleveland (US) at all age groups (Greulich-Pyle, 1957). However, Greulich argued that this was not due to ethnicity, as Japanese children living in California were skeletally delayed only between the ages of 5 and 7 years, which was attributed to less favourable environmental conditions, i.e. low socioeconomic status.

2.2.2 Socioeconomic factors

The term “socioeconomic status” refers to number of environmental factors such as the nutrition status, the health condition and social class of an individual. These factors can have a positive as well as a negative influence on the skeletal development, which makes it crucial to understand how the maturational rate of children performs in relation to these factors.

Full growth potential is achieved under ideal living conditions, when sufficient food, access to healthcare and housing are available (Cameron, 2002). Even in most modern societies, these conditions were not always available for older generations. Although there are people living in poverty in developed countries today, there has been an improvement in living standards over the past few decades, in which economic progress has played a major role (Easterlin, 2000). This improvement in socioeconomic status in developed countries has been suggested to be the main drive behind earlier skeletal maturation (Schmeling *et al.*, 2006; Schmidt *et al.*, 2008; Hackman and Black, 2013). Evidence suggests that current Caucasian American adolescents are significantly advanced in BA compared to healthy children of the 20th century (Calfee *et al.*, 2010).

Additional studies have indicated that BA is advanced in girls up to 13 years old and in boys aged 10 years and above as a result of improved socioeconomic status (Calfee *et al.*, 2010; Hackman and Black, 2013; Zabet *et al.*, 2014a). In contrast, children of lower socioeconomic status, where access to food and healthcare might be limited, are more likely to experience a delay in BA. Hawley et al. observed that the skeletal maturity of South Africans was delayed in 1962 but then accelerated during 2001 due to improvements in nutrition and access to healthcare.

Nutrition is one of the main factors influencing the growth of a child and contributes to the definition of socioeconomic status. Although both males and females are affected by malnutrition, evidence suggests that males are more prone to the negative effect than females. The World Health Organisation defines malnutrition as “deficiencies, excesses or imbalances in a person’s intake of energy and/or nutrients”. Therefore, malnutrition may refer to undernutrition or overnutrition, both of which can influence BA.

With regard to undernutrition, the body retards growth as a response to inadequate nutrition (Johnston, Francis E., 2002). This has been reported in several populations, such as in Guamanian Indians, in whom undernutrition resulted in delayed BA compared to a well-nourished population (Greulich, 1951; Gulati *et al.*, 1991). Additionally, children from a low socioeconomic class who experienced undernutrition also had delayed skeletal maturation (Pathmanathan and Raghavan, 2006; Hawley *et al.*, 2012a). Currently, the United Nations Children’s Fund (UNICEF) estimates that approximately 200 million children around the world do not have adequate nutrition, which can result in delayed or stunted growth (UNICEF, 2018). Approximately 65% of

those children live in developing countries in South Asia and Africa. Moreover, access to healthcare is variable within these developing countries, which may also significantly influence children's growth (Drèze and Sen, 2011; Zaidi *et al.*, 2017).

In contrast, overnutrition which leads to obesity has also been shown to affect skeletal maturation rate, with obesity in children linked to BA advancement (Giuca *et al.*, 2012; Johnson *et al.*, 2012). Indeed, Gicua et al. showed that obese children tend to be advanced in skeletal development compared to normal children (Giuca *et al.*, 2012). Studies have shown that modern societies are more likely to experience overnutrition and it should be noted that the prevalence of obesity in children has increased by 5% worldwide (Ng *et al.*, 2014). In the UK in particular, recent figures suggest that almost a third of children are overweight or obese (Van Jaarsveld and Gulliford, 2015).

In summary, inherited and socioeconomic factors can affect the speed of the maturational process, with none of these factors acting in isolation. Some factors can change over time, which makes assessing the relationship between skeletal maturity and CA a complicated and complex process.

2.3 Approaches to assess BA

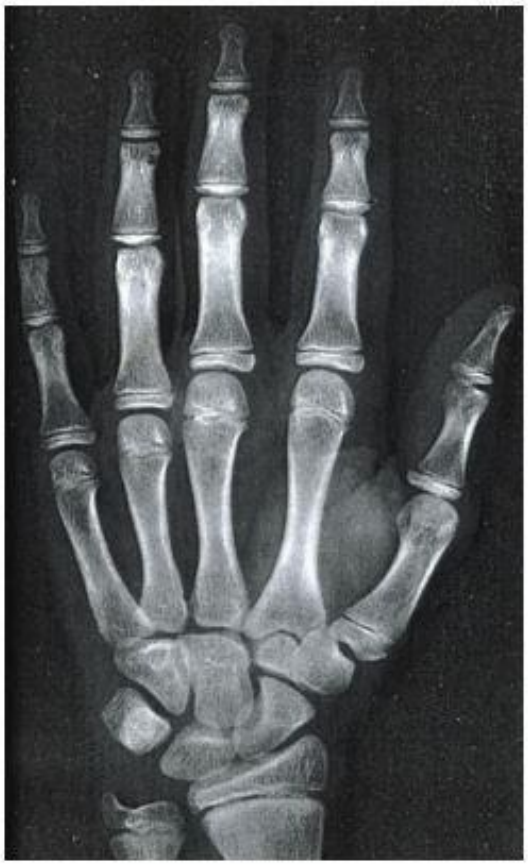
The determination of BA should be carried out using a reliable and suitable method. Currently, several radiographic methods are available which utilise different parts of the body, such as the hand and wrist. These techniques are generally based on the assumption that the ossification centres of the examined area appear and mature at different stages of development in a

consistent manner. Each stage reflects the CA; hence this method usually reflects the level of individual growth development. The hand-wrist is the most frequent region of the body used to assess BA, as it consists of many ossification centres that appear, change in shape and size, and fuse in a consistent pattern. The process of BA determination from the left hand is dependent on matching the acquired hand-wrist radiograph to a defined maturity stage that has been previously illustrated by a reference sample of hand radiographs of children of known sex and age. These reference standards describe and graphically illustrate the most important morphological and developmental changes that occur within the hand-wrist depending on CA. As these standards were primarily designed to indicate the normal development of a child at a known CA, children who formed these standards were healthy with no history of disorders that can affect growth. These reference data are typically used to identify children in whom normal growth is not experienced and to plan any necessary medical intervention through assessing skeletal maturity. Currently, two commonly used methods that utilise the hand-wrist radiograph to determine BA are Greulich and Pyle (G&P) and Tanner and Whitehouse (TW).

2.3.1 G&P atlas

The method used to construct the atlas depended on identifying changes that occur in the hand-wrist which reflect the process of maturation (Greulich and Pyle, 1959). In each maturational stage, 100 radiographs reflective of that maturational stage were chosen and arranged in the order of their relative skeletal status from the least to the most mature. Then, the radiograph that mostly reflected of that stage was included in the atlas

(Figure 2.1) and the CA was assigned. As bone growth differs depending on the individual's sex, the atlas contains 31 and 28 standard plates constructed for males and females, respectively.



Skeletal Age of Individual Bones

The skeletal age assigned to each bone in this standard is 13 years

The radial epiphysis and the epiphyses of the second to fifth metacarpals are now as wide as the adjacent margins of their shafts.

The ossification centre of the sesamoid in the tendon of the adductor pollicis is now visible, just medial to the head of the first metacarpal.

The epiphyses of the proximal phalanges of the second, third, fourth and fifth fingers have increased somewhat in thickness and their radial margins end in distally directed tips. The epiphysis of the middle phalanx of the fifth finger is now as wide as its shaft. The tips of the epiphyses of the distal phalanges of the second to fifth fingers are bent slightly distally and the distal ends of the corresponding middle phalanges are now slightly concave.

Figure 2.1: Standard No. 23 "Male" corresponding BA of 13 years (G&P atlas)

The population which formed the G&P atlas were originally enrolled in the "Brush study", also called "the Cleveland study", which is a longitudinal study that began in 1926. The data were collected in the period 1926 to 1947 from North American children whose physical, psychological, nutritional, medical health and socioeconomic status were recoded. The children were deemed to be of high socioeconomic status and were examined every 3 months until the age of 1 year, then every 6 months until the age of 5 years, and then every 12 months throughout adolescence.

The process of assigning BA using the G&P atlas begins by identifying the standard that most closely matches the child's radiograph, usually beginning with the standard for that child's CA. Then, each of the hand-wrist bones and sesamoids on the child's radiograph are compared systematically with the chosen standard radiograph to confirm a match. The G&P method is considered straightforward and quick, therefore widely used. CA and sex-matched standard deviation tables exist, and BA is said to be delayed or advanced if it falls below or above 2 standard deviations for the CA.

Many studies have assessed the applicability of the G&P atlas in relation to different populations, which may be of different ethnicity and/or socioeconomic status to those children on which the standards were based. Quantitative and qualitative synthesis approaches have been used to assess the reliability of the G&P atlas. A systematic review and meta-analysis conducted to assess the reliability and applicability of the G&P atlas when applied to different populations is detailed in Chapter Three.

2.3.2 TW method

In contrast to the G&P atlas, the TW method depends on assessing and scoring the skeletal maturity of each individual bone of the hand and wrist rather than analysing overall maturational status. The TW method is based on the Harpenden Growth study, which was established in the UK by Tanner and Whitehouse between 1948 and 1971 (Tanner, J. M., Whitehouse, R. H., Marshall, W. A., Healy, M. J. R. & Goldstein, 1975). The Harpenden longitudinal study involved a left-hand and wrist radiograph of 420 children aged between 3 and 18 years every 3 months. Subsequently, the method was revised (TW2) in 1975, with some changes to the description of the stages, although both versions were based on the same left hand radiographs. However, due to evidence suggesting that full skeletal maturity was reached sooner than was thought in the 1960s, Tanner et al. then developed the TW3 method, which used different reference data collected during the First Zurich Longitudinal Growth study as well as data from Japan, America and England (Tanner *et al.*, 2001).

The TW method divides the ossification centres within the hand into two groups, the RUS (radius, ulnar and short bones) which involves the radius, ulnar, metacarpals and phalanges and the carpals, which include all the carpal bones, except the pisiform. However, assessment of the carpals is limited due to their being less useful for diagnosing or monitoring treatment in children (David D Martin, Jan M Wit, *et al.*, 2011). Johnston argued that the carpals are more likely to cause significant variable error when re-rating a left hand radiograph by the same observer (Johnston and Jahina, 1965). Additionally, the carpal maturation rates vary among individuals and

populations, with the carpals being less mature compared to the rest of the hand bones (Acheson, Vicinus and Fowler, 1966; Krailassiri, Anuwongnukroh and Dechkunakorn, 2002; Al-Hadlaq *et al.*, 2007).

In the TW3 method, the assessment process starts with assessing each ossification centre individually; the maturational stage of the ossification centre is assigned a letter from A to H, where A indicates no ossification of the bone, while H means that the bone is fully mature. For each stage, there are written criteria and a diagram to reflect the typical appearance of the bone. These criteria must be met before assigning the bone's stage. For instance, if there are one or two criteria specified for a particular stage, then at least one criterion must be met. However, when there are three criteria, then *two* of them must be met. Each stage is assigned a numerical score, the sum of which will result in a skeletal maturity score, which ranges between 0 (meaning the bones have not begun to ossify) and 1000 (indicating full maturity).

In females, the maximum score, which is 1000, corresponds bone age of 15 years or 13 years when using RUS or carpal method, respectively. However, in males the same maximum score, correspond BA of 16.5 years when using RUS and 15 years when using carpal method. The final score can either be the sum of the scores from the RUS alone, which are the scores from radius, ulnar and short bones, or the sum of the scores from the identified carpal bones. The third option, which is the combined score of the RUS and carpal bones "also called 20-bone", was disregarded in the TW3 method, as the authors believed that it was unnecessary. In relation to the reliability of the TW3 method, many studies have been conducted to evaluate its use in

different populations. Some of these populations have shown advanced skeletal maturation compared to the TW3 standard. The findings of these studies, (particularly the mean between BA and CA and the conclusion) are summarised in Table 2.1 (following page). From these studies, the BA of Asian children appeared to be advanced after the age of 6, whereas full maturity was reached around half a year earlier than the TW3 standard (Ashizawa *et al.*, 2005; Griffith, Cheng and Wong, 2007; S.-Y. Zhang *et al.*, 2009; Kim, Lee and Yu, 2015). These differences were more likely to be due to differences in genetics and living conditions. In contrast, African children showed delayed skeletal maturation compared with the TW3 standard (Cole, 2015). Cole et al. and Hawley et al. argued that these variations in skeletal maturation were due to less favourable environments. In Caucasian and Hispanic populations, skeletal maturation seems to conform to the TW3 standard (Caldas, Ambosano and Haiter-Neto, 2007; Schmidt *et al.*, 2008; Büken *et al.*, 2009; Freitas *et al.*, 2012; Pinchi *et al.*, 2014), although one study showed that BA was advanced in females between the ages of 7 and 15 years (Caldas, Ambosano and Haiter-Neto, 2007).

It should be mentioned that TW is a complex technique and more time consuming than the G&P method, which is one of its major drawbacks. Assessing skeletal maturity using the TW method usually requires 7.9 minutes in comparison to 1.4 minutes when using G&P (Cox, 1996). The significantly longer time required for the TW method and the higher level of complexity may have limited the practicability of this method.

Table 2.1: Summary of studies that assessed the reliability of the TW3 method.

Study	Origin/ Ethnicity	Age (years)	Sex	Mean BA-CA (years)	Authors' Conclusion
Ashizawa et al, 2005	Beijing	6-16	M=631 F=642	M=0.07 F=0.11	Full maturity was reached at the age of 16, i.e. half a year earlier than TW3 standard.
Haiter-Neto et al, 2006	Brazilian	7-15	M=180 F=180	M=-0.2 F=0.10	TW3 can be used within the Brazilian population.
Caldas et al, 2007	Brazilian	7-15.9	M=110 F=128	M=-0.2 F=0.40	There was a statistically significant difference between BA and CA in females but not males.
Griffith et al, 2007	Hong Kong	0-18	M=645 F=329	M=0.22 F=0.3	Greater accuracy could be achieved by adjusting designated standards for different age groups and genders.
Schmidt et al, 2008	Germany	1-18	M=48 F=40	M=0.61 F=0.23	There was a strong correlation between TW3 BA and CA.
Zhang et al, 2008	Chinese	1-20	M= 8685 F= 8716	Not reported	Advanced skeletal maturity was observed in males and females, after the age of 6 and 10 years, respectively.

Buken et al, 2009	Turkish	11-16	M=169 F=164	M=-0.18 F=-0.21	TW3 can be used in girls 11-15 years and boys 11-16 years old.
Freitas et al, 2012	Portugal	4-17	M=1412 F=1444	M=-0.23 F= -0.12	No trend of advancement in skeletal maturity was observed.
Hawley et al, 2012	South African	9-10	M=131 F=113	M=-0.66 F=-1	Children showed delayed skeletal maturation in comparison with the TW3 standard.
Pinchi et al, 2014	Italian	6-20	M=162 F=145	Not reported	The TW3 appears to be reliable in males and females.
Cole et al, 2015	South African	9-20	M=374 F=287	Not reported	BA was delayed in black south African (by 7 months) but not in white south African. The standard was applicable to the girls.
Kim et al, 2015	Korean	7-12	M=135 F=77	M=0.41 F=0.12	TW3 showed good reliability in the evaluation of BA of prepubertal healthy Korean children.

A positive value of the mean difference between BA and CA indicates advanced while a negative value indicates delayed BA compared to chronological age, M= male, F= female

2.4 Technical aspects in BA assessment

There are some technical aspects that need to be considered when assessing BA. One of these is that the determination of BA, being a subjective technique, is likely to suffer from variations in rating among assessors due to different levels of training, with its reliability largely dependent on a well-trained assessor (Cox, 1996; Thodberg *et al.*, 2009). These variations regarding accuracy may have negatively impacted on which methods are regarded as acceptable for use in clinics. As BA assessment plays an important role in the clinical and forensic environment, it is very important that results are accurate.

When using the G&P method, Berst et al. reported that the inter-observer and intra-observer variability varied from 0.10 to 1.05 years and 0.09 to 1.20 years respectively (Berst *et al.*, 2001). In another study, the average intra-observer variability for the G&P method was 0.96 years, which was then reported to be statistically significant at the 5% level (Bull *et al.*, 1999). To eliminate the role of the observer and to increase the accuracy of ratings, automated BA systems have been considered.

During the last three decades, several attempts have been made to assess BA automatically. Unfortunately, the accuracy of most of these automatic systems was considered sub-optimal, hence, they were not clinically useful. However, BoneXpert, an automated software for BA assessment was introduced in 2009. The BoneXpert system determines BA from left hand radiographs based on the G&P and TW3 methods. The software has a stand-alone version and also a version module that can integrate with PACS

system. The input is a DICOM format image, which includes information on the subject's gender, the image resolution, and the date of the x-raystudy.

The BoneXpert approach for BA assessment consists of three computational layers (Thodberg *et al.*, 2009): 1. The first layer reconstructs the borders of 15 bones – the five metacarpals, the phalanges of fingers 1, 3 and 5, and the radius and ulna. This allows the software to determine to what extent the bone appears normal. Hence, abnormal bones are automatically rejected. 2. The second layer determines bone maturity values, which called intrinsic bone ages, for 13 of these 15 bones based on the appearance of the bone. If a BA deviates more than 2.4 years from the average of all the bones, the BA is then considered unacceptable. Additionally, the image is rejected and no BA value is reported when fewer than eight bones are accepted.

3. The third layer convert the computed intrinsic bone ages to conform on average with G&P BA based on a training set of images with manual ratings.

The most complicated layers are the first two layers, which w

The first and the second layers were established biased on radiographs from Danish and Belgian children (age range 7–17 years), enhanced by radiographs from multiple sources; in total 1,678 images (Thodberg, 2009). BoneXpert's accuracy is recognized to be poor for boys above 17 years and girls above 15 years, so the intended age range for the clinical use of BoneXpert v1.0 is GP BA 2.5–17 years for boys and 2–15 years for girls.

The third layer was developed by combing three datasets in order to average over several manual raters, which serve as an adjustment of BoneXpert to G&P BA. These studies are Erasmus study, a study performed

in Tübingen (Lequin *et al.*, 2000; Van Rijn *et al.*, 2001; van Rijn, Lequin and Thodberg, 2009), and the G&P atlas. Then, a nonlinear transformation of the intrinsic BA into the BoneXpert G&P BA was established from these three studies.

The software has shown adequate accuracy with a precision of 0.18 years compared to 0.58 for manual rating (Thodberg *et al.*, 2009). In terms of re-rating errors, the software shows superiority to the manual method, with a precision SD of zero compared to between 0.25 to 0.85 years for manual rating. BoneXpert works in the BA range of 2.5 to 17 years for boys and 2 to 15 years for girls, with a rejection rate of approximately 2% for poor quality. The software provides standard deviation scores (SDS) for each hand radiograph, thus assisting in the comparison of a child's BA with healthy children of the same sex and age. BoneXpert has been validated to be used as an automated BA assessment tool for Caucasians, African-Americans, Hispanics, and Asians in the US (Thodberg and Sävendahl, 2010). However, the validity of the software has not been evaluated for other ethnic groups, such as Africans and Asians living in Africa and Asia.

BoneXpert can also calculate bone mass, presented in the form of a bone health index (BHI) and derived from the cortical thickness and width and length of the three middle metacarpals. The manual equivalent of this technique was first established during the 1960s and was later called radiogrammetry (Virtama and Mahone, 1960; Rosholm *et al.*, 2001). In this method, cortical thickness is manually measured as an indication of bone strength; with bone loss comes a thinning of the cortex which can be assessed by radiogrammetry. Although this method has been deemed to be

inexpensive and easily acquired, its reproducibility is questionable with a coefficient of variation (CV) approximately 10% due to the subjectivity in assessing bone mass (Adams, Davies and Sweetnam, 1969). The method became even less popular with the availability of DXA, which can measure bone mineral content. However, soon after digital radiography was introduced, automatic software to quantify bone mass from hand radiographs was developed. Among these, BoneXpert was shown to have much better precision (1.42%) than manual radiogrammetry (Thodberg, 2009). Additionally, the software also provides SDS for the BHI values, which facilitates the evaluation of bone mass in comparison to age-matched healthy children. These SDS are based on values from a large cohort of Caucasian children. The bone mass values acquired by BoneXpert have recently been evaluated in relation to values obtained using DXA. There was moderate correlation between the two techniques, but some limitations of these studies include the time interval between the DXA and the radiograph of the left hand and the small sample sizes (Nusman *et al.*, 2015; Schündeln *et al.*, 2016)

BoneXpert offers the advantage of monitoring BHI at no additional radiation, as left hand and wrist radiographs are frequently requested in paediatrics to determine BA for many clinical indications. All radiogrammetry methods, including BoneXpert, are dependent upon cortical bone. The software might be valuable when monitoring the change in cortical bones following certain treatment plans. For example, bisphosphonates are commonly used to treat paediatric patients with low bone mass (e.g. those with osteogenesis imperfecta) and have been shown to increase the cortical thickness of the

metacarpal bones (Glorieux *et al.*, 1998). BoneXpert may be an effective way of capturing this change in cortical thickness as the method depends on changes in cortical bone thickness. Interestingly, comparison of DXA and BoneXpert results have not so far been published in patients treated with bisphosphonates.

Another technical aspect in relation to the determination of BA is the radiation dose associated with the left hand and wrist radiograph. The effects of ionising radiation have been extensively studied (De González and Darby, 2004; Hall and Brenner, 2008; Ramsthaler *et al.*, 2009), emphasising the main principle in medical imaging, which is ensuring that the dose delivered to the patient is as low as reasonably practical (Uffmann and Schaefer-Prokop, 2009). Children who suffer from some chronic diseases require regular BA monitoring, which means repeated hand-wrist radiographs numerous times throughout their childhood (91). Hand-wrist radiographs are routinely required for age estimation in children seeking asylum (A. Schmeling *et al.*, 2008; Black, 2010). In both medical and legal situations, BA should be justified and the radiation dose optimised where possible, especially when the procedure involves children (Hall, 2009). The effective dose of a left hand radiograph is on average 1 µSv, while people in the UK on average are exposed to approximately 2.7 millisieverts of radiation annually (Hall and Brenner, 2008; Oatway *et al.*, 2010). Therefore, one hand radiograph equals to around 25 minutes of exposure to natural background radiation. Although the radiation dose involved in left hand radiographs is relatively low, any exposure to radiation is not without risk (Damilakis *et al.*, 2010).

To reduce exposure to ionising radiation, other imaging modalities involving less radiation are usually preferred, hence the assessment of BA based on left hand-wrist DXA scans (Pludowski, Lebiedowski and Lorenc, 2004; Heppe *et al.*, 2012). Pludowski et al. assessed BA based on the G&P atlas using hand-wrist DXA images and compared the results with those obtained from radiographs. They concluded that the DXA approach produced comparable results to radiographs, with the advantage of using a less potent radiation dose. The radiation dose was 10-fold less than left hand radiographs, but the use of inappropriate statistical analysis in their study render their conclusions questionable. Later studies showed that using DXA hand-wrist images to assess BA produced similar results to conventional methods, concluding that DXA could be an alternative method for assessing BA (Heppe *et al.*, 2012; Romann and Fuchslocher, 2016). However, these studies only determined BA based on the G&P atlas, so it is still unclear whether DXA left hand scans are appropriate for the more detailed TW3 method.

2.5 Summary

Assessing BA from left hand radiographs is important in clinical practice to monitor the maturation and growth of children. The assessment also has a role in forensic science to determine CA. There are two common methods currently used to assess BA from hand and wrist radiographs; the TW3 and G&P methods. Several factors influence BA which may limit the applicability of these two methods.

To determine whether these standards are still appropriate for use in modern populations, the applicability of these methods should be reviewed in light of economic progress and elevation in living standards. Additionally, the applicability of these methods should be reviewed in relation to other populations who are of different ethnicity to those of the original standards.

The availability of validated automated software has eliminated inter- and intra-observer variation, providing a timesaving solution for the determination of BA. The BoneXpert software assesses BA based on the G&P and TW3 methods, providing a G&P SDS for each hand radiograph, enabling comparison of a child's BA with healthy children of the same sex and age. The added advantage of BoneXpert is that bone mass can be assessed from left hand radiographs, so there is no need for further radiation exposure. However, the use of the software to monitor bone mass has not been fully evaluated, and as far as could be determined, never in patients who have increased cortical thickness as a result of medical treatment. Furthermore, the software performance with regard to images taken using modalities other than conventional radiography has not been evaluated. The radiation dose of left hand DXA is much lower in comparison to that of a left hand radiograph, therefore DXA scans could be an alternative for children in whom the determination of BA is required numerous times during their childhood. The results from studies using DXA to assess BA are encouraging, suggesting that BA can be determined using DXA with the advantage of less radiation. However, none of these studies has assessed the role of BoneXpert or whether TW3 BA can be determined from left hand DXA scans.

CHAPTER 3

Is the Greulich and Pyle Atlas Applicable to All Ethnicities? A Systematic Review and Meta-Analysis

Alshamrani K[1], Messina F[2, 3], Offiah A.[1,4]

[1] Department of Oncology & Metabolism, University of Sheffield, UK

[2]School of Health and Related Research, University of Sheffield, UK

[3]Leeds Institute of Clinical Trials Research, University of Leeds, UK

[4]Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield, UK

Correspondence: Khalaf Alshamrani, Academic Unit of Child Health, Damer Street Building, Sheffield, S10 2TH

kalshamrani1@sheffield.ac.uk

In Press (European Radiology) DOI: 10.1007/s00330-018-5792-5

3.1 Abstract

Objective: To determine whether the Greulich & Pyle (G&P) atlas is applicable when applied to populations of different ethnicity.

Methods: A systematic review of studies published between 1959 and 15th April 2019 identified from the Embase, MEDLINE and Cochrane databases was undertaken. Quality of the studies was assessed using the National Institute for Health and Care Excellence tool. Meta-analysis used mean differences and standard deviations as summary statistics for the difference between bone age (BA) and chronological age (CA).

Results: A total of 49 studies were included of which 27 (55%) were related to Caucasian populations. Of the 49 eligible studies, 35 were appropriate for further meta-analysis. In African females, meta-analysis showed a significant mean difference between BA and CA of 0.37 years (95% CI: 0.04,0.69). In Asian males, meta-analysis showed significant differences between BA and CA of -1.08, -1.35, -1.07, -0.80 and 0.50 years for chronological ages of 6,7,8,9 and 17 years respectively. Meta-analysis showed no significant differences between BA and CA in African males, Asian females, Caucasians and Hispanics.

Conclusions: The G&P standard is imprecise and should be used with caution when applied to Asian male and African female populations, particularly when aiming to determine chronological age for forensic/legal purposes.

3.2 Introduction

Determining maturity and understanding growth in a child is critical for medical and psychosocial purposes. Assessing bone age is important to investigate whether the maturity of bones is occurring at the same rate as the chronological ageing process. Furthermore, bone age assessment has a role in forensic and legal investigations when the individual's chronological age is in doubt. For example, in asylum-seekers and unaccompanied minors without valid documents to prove their ages (Menjívara and Krista M. Perreirab, 2017). It is important to assess bone age using a reliable and suitable method (A. Schmeling *et al.*, 2008). Incorrectly assessing a child as an adult leaves the child with limited access to education, healthcare and other support provided to children.

There are two approaches widely used to determine bone age from a left hand radiograph; the Greulich and Pyle (G&P) and Tanner and Whitehouse (TW) methods (Greulich and Pyle, 1959; Tanner *et al.*, 2001). The population which formed the G&P standard atlas were North American Caucasians of good socioeconomic status. The assessment process is typically based on comparing a hand-wrist radiograph of a child with the age-matched standard radiographs as contained in the atlas. The G&P method depends on comparing the overall maturational status and is known to be straightforward and quick, therefore widely used. In contrast, the TW method depends on assessing and scoring the skeletal maturity of each individual bone of the hand, hence taking a longer time than the G&P method. Since the establishment of the G&P atlas, many studies have been conducted in different parts of the world to determine whether it is applicable to different

populations. This question is important, particularly given the increasing legal and illegal influx of immigrants to certain parts of Europe. This systematic review and meta-analysis aim to provide a better understanding of the applicability of the G&P atlas to children and adolescents who are of a different population from the original standard.

3.3 Materials and Methods

3.3.1 Search Strategy

A systematic search of the MEDLINE, Embase, and Cochrane databases was conducted. We searched MEDLINE using keywords, ((Greulich and Pyle)) OR Greulich Pyle, ((bone age assessment OR bone age determination)) AND left hand and refined the search to include articles in English published between 1st January 1959 and 15th April 2019. No free text was used in this search. For Embase, we used the term (Greulich and Pyle) and refined the search to include articles in English published between 1st January 1959 and 15th April 2019. We also searched the Cochrane library using the keywords, (Greulich and Pyle), and the MeSH term (Age Determination by Skeleton). The search was refined to include articles in English published between 1st January 1959 and 15th April 2019. Each study's title and abstract were screened to determine whether it presented data correlating bone age assessed by the G&P with chronological age. The full text was retrieved when the reviewers could not decide on the study's eligibility from the title and abstract alone. The following exclusion criteria were then applied:

1. Health status of participants could not be confirmed from the article or participants with developmental disorders or subjected to nutritional

supplementation (these represent unhealthy children expected to show delayed or advanced bone age)

2. Using a modified method of G&P and/or using modalities other than conventional radiography
3. Full text not available within the resources available to the reviewers
4. Full text not in English
5. Review Articles
6. When the mean difference between bone age (BA) and chronological age (CA) was not reported or could not be calculated by the reviewers based on the study results presented.

The search was independently carried out by two reviewers (KA, ACO), followed by a consensus meeting to agree the final selection of studies for inclusion in this review.

3.3.2 Quality assessment

Two reviewers KA and ACO independently assessed the quality of included studies using the tool developed by the National Institute for Health and Care Excellence (NICE) (5). Discrepancies were resolved by discussion. The tool considers five aspects of a study; population, method of participant selection, outcomes, analysis and generalisability of the study. Then, an overall quality grading is given to each study for internal validity (IV) and a separate grading for external validity (EV) as follows:

- ++ All/most of the checklist criteria have been fulfilled and the conclusions are unlikely to alter
- + some of the checklist criteria have been fulfilled, the conclusions are unlikely to alter even when they have not been fulfilled

- - few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter

3.3.3 **Data Extraction**

A single reviewer (KA) extracted and recorded the following data from eligible studies; Sample size (males and females), ethnicity or country of origin, mean difference and standard deviation (SD) between bone age and chronological age (BA-CA), mean and SD of bone age, mean and SD of chronological age, authors' conclusions and applicability of the standard.

Given the review question, studies were divided into four groups based on major ethnic groups; African, Asian, Caucasian and Hispanic. Data for each major ethnic group were summarised and analysed separately. Some studies reported the place/country from which participants were recruited, and in such cases the study was grouped under the major ethnicity of that country. The mean differences between BA and CA are to be interpreted as follows: a positive value indicates that the child's bone age exceeds the child's chronological age; a negative value indicates delayed bone age compared to chronological age.

Additionally, we defined four categories to reflect the applicability of the G&P standard to the studied population as follows: (a) applicable; (b) not applicable (determined by the authors' use of words identical or similar to "applicable" or "not applicable" respectively in the study's discussion or conclusion); (c) needs some modification (authors use phrases such as, "can be used with caution" or when the standard was found to be applicable to a certain age group but not others); (d) not clear (when the study failed to

mention whether the standard was applicable, not applicable or needed modification).

2.3.4 Statistical Analysis

A combination of random effect meta-analyses by ethnicity (African, Asian, Caucasian, Hispanic) and sex was conducted using R Software (6). Overall meta-analysis of all ethnicities was also determined. Additionally, meta-regression with covariates analysis (including sex and ethnicity as explanatory variables) was determined. Yearly interval sub-analysis of Asians aged 6 to 17 years and Caucasians aged 10 to 17 years were carried out in males and females. Other ethnicities were excluded from interval sub-analysis as the age groups were not constant between studies.

In total, 50 meta-analyses were performed using mean differences and standard deviations as summary statistics for the difference between bone age and chronological age. When a study examined more than one ethnicity, each ethnicity was treated as a separate study (only for the meta-analysis). Heterogeneity was assessed between 0% (no heterogeneity) and 100% (maximum heterogeneity) using the I-squared statistic. A funnel plot was determined to assess bias or the present of any systematic heterogeneity.

3.4 Results

This systematic review identified 931 studies of which 48 were eligible for inclusion (Figure 3.1–page 54). Four additional studies were identified from the reference lists of the initial 48 extracted papers, therefore the total number of included studies was 52 (Demisch and Wartmann, 1956; Hansman and Maresh, 1961; Johnston, 1963; Andersen, 1971; Roche,

Davila and Eyman, 1971; Wenzel, Droschl and Melsen, 1984; So and Yen, 1990; Y So and Lisa So, 1991; Loder, 1993; Kullman, 1995; Jiménez-Castellanos *et al.*, 1996; Ontell *et al.*, 1996; Koc *et al.*, 2001; Mora *et al.*, 2001; Van Rijn *et al.*, 2001; Krailassiri, Anuwongnukroh and Dechkunakorn, 2002; Lewis, Lavy and Harrison, 2002; Garamendi *et al.*, 2005; Chiang *et al.*, 2005; Haiter-Neto *et al.*, 2006; Griffith, Cheng and Wong, 2007; Büken *et al.*, 2007, 2009; Schmidt *et al.*, 2007b; A. Zhang, James W. Sayre, *et al.*, 2009; Zafar *et al.*, 2010; Santos *et al.*, 2011; Moradi, Sirous and Morovatti, 2012; Patil *et al.*, 2012; Santoro *et al.*, 2012; Soudack *et al.*, 2012; Cantekin *et al.*, 2012; Dembetembe and Morris, 2012; Hackman and Black, 2013; Paxton, Lamont and Stillwell, 2013; Shilpa *et al.*, 2013; Suri *et al.*, 2013; Awais *et al.*, 2014; Mansourvar *et al.*, 2014; Mughal, Hassan and Ahmed, 2014; Rai, 2014; Zabet *et al.*, 2014b; Gungor *et al.*, 2015a; Kim, Lee and Yu, 2015; Mohammed *et al.*, 2015; Patel *et al.*, 2015; Maggio *et al.*, 2016; Öztürk *et al.*, 2016; Chaumoitre *et al.*, 2017; Govender and Goodier, 2018), of which 28 (57%) were related to Caucasian populations. The total number of children in the included studies was 24,735 (13,237 boys), comprising 14,021 Caucasians (7,526 boys); 6,776 Asians (3,731 boys); 1,851 Africans (1,137 boys) and 2,087 Hispanics (1,081 boys).

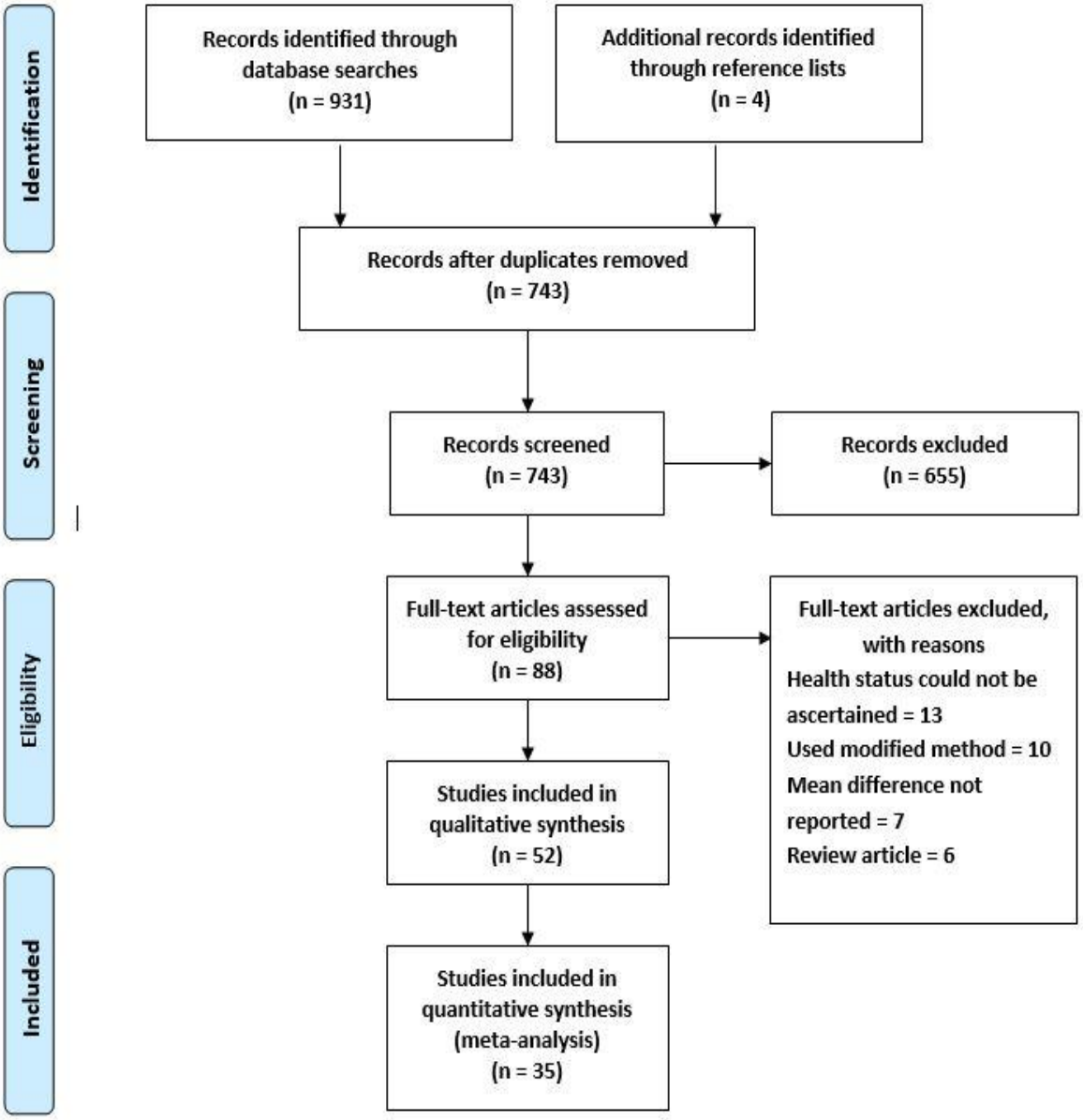


Figure 3.1: Flow chart to show article selection process

As summarised in Table 3.1 (page 57), there was minimal risk of bias for internal validity alone in two studies (Zafar *et al.*, 2010; Alcina *et al.*, 2018) for external validity alone in five studies (Jiménez-Castellanos *et al.*, 1996; Ontell *et al.*, 1996; Garamendi *et al.*, 2005; Soudack *et al.*, 2012; Kim, Lee and Yu, 2015) and for both internal and external validity in 13 studies (Roche, Davila and Eyman, 1971; Van Rijn *et al.*, 2001; Mora *et al.*, 2001; Krailassiri, Anuwongnukroh and Dechkunakorn, 2002; Büken *et al.*, 2007, 2009; Cantekin *et al.*, 2012; Hackman and Black, 2013; Paxton, Lamont and Stillwell, 2013; Awais *et al.*, 2014; Zabet *et al.*, 2014b; Mohammed *et al.*, 2015; Chaumoitre *et al.*, 2017). There was significant risk of bias for internal validity alone in 0 studies, for external validity alone in two studies (Hansman and Maresh, 1961; Rai, 2014) and for both internal and external validity in 2 studies (Lewis, Lavy and Harrison, 2002; Schmidt *et al.*, 2007b). Sources of bias in these four studies requiring that their results be interpreted with caution include: absent documentation of statistical criteria such as p values and /or observer reliability (Hansman and Maresh, 1961; Lewis, Lavy and Harrison, 2002), insufficient detail about the source of the study population (Schmidt *et al.*, 2007b) and non-representative samples (Rai, 2014).

Studies included in this systematic review reported the mean difference between bone age and chronological age in different forms. Thirty studies (60%) (Hansman and Maresh, 1961; Johnston, 1963; Andersen, 1971; Roche, Davila and Eyman, 1971; Wenzel, Droschl and Melsen, 1984; So and Yen, 1990; Y So and Lisa So, 1991; Jiménez-Castellanos *et al.*, 1996; Koc *et al.*, 2001; Krailassiri, Anuwongnukroh and Dechkunakorn, 2002; Chiang *et al.*, 2005; Haiter-Neto *et al.*, 2006; Büken *et al.*, 2007, 2009;

Griffith, Cheng and Wong, 2007; Patil *et al.*, 2012; Cantekin *et al.*, 2012; Dembetembe and Morris, 2012; Hackman and Black, 2013; Shilpa *et al.*, 2013; Suri *et al.*, 2013; Mughal, Hassan and Ahmed, 2014; Awais *et al.*, 2014; Zabet *et al.*, 2014b; Kim, Lee and Yu, 2015; Patel *et al.*, 2015; Gungor *et al.*, 2015a; Öztürk *et al.*, 2016; Alcina *et al.*, 2018) presented the mean difference for each year of age for each sex. Thirteen studies (Loder, 1993; Ontell *et al.*, 1996; Mora *et al.*, 2001; Garamendi *et al.*, 2005; A. Zhang, James W. Sayre, *et al.*, 2009; Zafar *et al.*, 2010; Soudack *et al.*, 2012; Moradi, Sirous and Morovatti, 2012; Santoro *et al.*, 2012; Mansourvar *et al.*, 2014; Rai, 2014; Govender and Goodier, 2018) divided their sample into subgroups, where each subgroup contains up to 5 age groups e.g. children aged between one and 5 years old. For each subgroup, the overall mean difference for each sex is reported. Nine studies (Demisch and Wartmann, 1956; Kullman, 1995; Van Rijn *et al.*, 2001; Lewis, Lavy and Harrison, 2002; Schmidt *et al.*, 2007b; Calfee *et al.*, 2010; Paxton, Lamont and Stillwell, 2013; Maggio *et al.*, 2016; Chaumoitre *et al.*, 2017) only reported the overall mean difference between bone age and chronological age, limiting the applicability of their results to individual age groups. Data relating to ethnicity or country of origin, sample size, mean BA-CA and the authors' conclusions are summarised for each study in Tables 3.2 to 3.5.

Table 3. 1: Quality assessment of the included studies (after agreement between the two assessors)

Study [Reference]	IV*	EV**	Study	IV*	EV**
Demish & Wartmann, 1956	+	+	Santos et al, 2011	+	+
Hansman 1961	+	-	Cantekin et al, 2012	++	++
Johnston, 1963	+	+	Dembetmbe & Morris, 2012	+	+
Andersen 1971	+	+	Moradi et al, 2012	+	+
Roche et al, 1971	++	++	Patil et al, 2012	+	+
Wenzel et al, 1984	+	+	Santoro et al, 2012	+	+
So & Yen 1990	+	+	Soudack et al, 2012	+	++
So & Yen 1991	+	+	Suri et al, 2012	+	+
Loder et al, 1993	+	+	Hackman et al, 2013	++	++
Kullman 1995	+	+	Paxton et al, 2013	++	++
Ontell et al, 1996	+	++	Shilpa et al, 2013	+	+
Jimenez, 1996	+	++	Awais et al, 2014	++	++
Koc et al, 2001	+	+	Rai et al, 2014	+	-
Mora et al, 2001	++	++	Mansourvar et al, 2014	+	+
Van Rijn et al, 2001	++	++	Mughal et al, 2014	+	+
Krailassiri et al, 2002	++	++	Gungor et al, 2015	+	+
Lewis et al, 2002	-	-	Kim et al, 2015	+	++
Chiang et al, 2005	+	+	Mohammed et al, 2015	++	++
Garamendi et al, 2005	+	++	Ozturk et al, 2015	+	+
Haiter-Neto et al, 2006	+	+	Patel et al, 2015	+	+
Buken et al, 2007	++	++	Zabate 2015	++	++
Griffith et al, 2007	+	+	Maggio et al, 2016	+	+
Schmidt et al, 2007	-	-	Chaumoitore et al 2017	++	++
Buken et al, 2009	++	++	Alcina et al 2018	++	+
Zhang et al, 2009	+	+	Govender et al 2018	+	+
Calfee et al, 2010	+	+	Zafar et al, 2010	++	+

*IV: internal validity, **EV: external validity

++	Indicates that the study has been designed or conducted in such a way as to minimise the risk of bias
+	Indicates that either the answer to the checklist question is not clear from the way the study is reported, or that the study may not have addressed all potential sources of bias
-	Reserved for those aspects of the study design in which significant sources of bias may persist

Table 3.2: Summary of studies that assessed the reliability of the G&P atlas in Caucasian children

Study	Origin/ ethnicity	Age (years)	N	Mean BA-CA (years)	Authors' Conclusion	Applicability
Demish & Wartmann 1956	White	9-15	M=81 F=70	M=0 F=0.5	There is a high positive correlation between BA and CA.	Applicable
Hansman & Maresh, 1961	White	0-18	M=27 F=36	M=-0.33 F=0.75	The mean BA for both sexes is equal to CA during infancy but less than CA toward adolescents.	Applicable
Johnston 1963	White	7-17	M=388 F=405	M=0.40 F=0.20	Children show significant differences between CA and BA.	Needs some modification
Andersen 1971	Danish	7-18	M=477 F=535	M=0.49 F=-0.43	BA is lower than CA, indicating that the American children mature earlier than the Danish.	Needs some modification
Roche et al, 1971	British	2-13	M=62 F=82	M=0.01 F=0.07	Children matured skeletally at about the same as the G&P standard.	Applicable
Wenzel et al, 1984	Austrian	7-16	M=459 F=178	M=-0.2 F=-0.13	Major deviations between BA and CA were at and after puberty.	Not clear
Loder et al, 1993	White	0-18	M=203 F=177	M=-0.1 F=0.07	The G&P atlas is applicable to white girls at all ages and white boys in early childhood (less than 4 years old). BA of white boys was delayed during middle and late childhood but advanced during adolescence by 5 years.	Not applicable
Kullman 1995	Swedish	12-19	M=38 F=34	M=-0.4 F=-0.4	It is recommended to assess skeletal development using G&P.	Applicable
Ontell et al, 1996	White	3-18	M=208 F=130	M=-0.29 F=0.14	The G&P standard is applicable to white girls at all ages, while in boys it can only be applied in adolescence.	Applicable (for girls but not boys)
Koc et al, 2001	Southeast Turkey	7-17	M=225	M=-0.2	Mean BA was delayed between 7 and 13 years old and then advanced between 14 and 17 years. The atlas can be used with some modification.	Needs some modification

Mora et al, 2001	European American	0-19	M=130 F=130	M=0.09 F= -0.14	Prepubertal European-American children have significantly delayed BA when compared to African-American children. Post-pubertal European-American males have significantly advanced BA when compared with African-American males. A new standard is needed for reliable BA assessment.	Not applicable
Van Rijn et al, 2001	Dutch	5-20	M=294 F=294	M=-0.28 F=-0.14	Significant correlation between BA and CA in boys and girls. The G&P atlas is still applicable to Dutch Caucasian children and adolescents.	Applicable
Buken et al, 2007	Turkish	11-19	M=251 F=241	M=0.13 F=0.54	Mean skeletal ages were significantly advanced for boys and girls between 11 and 17 years old. The cause of this acceleration might be new social and cultural factors rather than economic conditions.	Needs some modification
Schmidt et al, 2007	Germany	1-18	M=303 F=303	M=-0.49 F=-0.39	The G&P atlas method overestimated the samples' age. This may be due to high acceleration in growth.	Applicable
Buken et al, 2009	Turkish	11-16	M=169 F=164	M=-0.02 F=-0.65	The G&P atlas is appropriate in girls 11-15 years old and boys 11-16 years old from the Black Sea region of Turkey.	Needs some modification
Zhang et al, 2009	White	0-18	M=164 F=163	M=0.01 F=-0.15	BA was relatively close to CA in white children.	Applicable
Calfee et al, 2010	Caucasian	12-18	M=62 F=76	M=0.98 F=0.66	American children between 12 and 18 years demonstrate BA exceeding CA. Females between 12 and 15 years old are most likely to demonstrate a discrepancy of at least 2 years between BA and CA, while males demonstrate this throughout adolescence.	Not clear
Cantekin et al, 2012	Eastern Turkish	7-17	M=342 F=425	M=-0.13 F=0.20	The mean differences between BA and CA are low enough to be of no practical significance, and thus, this method can be used in all age groups within the current study.	Needs some modification
Santoro et al, 2012	Italian	7-15	M=243 F=261	M=-0.1 F=0.40	The G&P method is accurate, particularly in the age ranges of 7-9 years and 10.4-11.5 years.	Applicable

Suri et al, 2012	White	9-18	M=311 F=261	M=0.50 F=0.50	Wide range of differences between BA and CA at each yearly age group from 9 to 18 years. Overall, the differences in skeletal and chronological age were positively correlated.	Not clear
Hackman & Black 2013	Scottish	1-20	M=249 F=157	M=-0.13 F=-0.16	The G&P atlas over-aged females from birth until 13 years of age and under-estimated males from birth until 13 years of age after which point it consistently over-aged boys between 13 and 17 years of age.	Needs some modification
Paxton et al, 2013	Australian	0-18	M=276 F=130	M=-0.12 F=-0.30	The G&P atlas is an accurate means of BA determination in Australian children.	Applicable
Mansourvar et al, 2014	White	10-16	M=46	M=0.04	The G&P is reliable in Caucasian males.	Applicable
Gungor et al, 2015	Turkish	10-18	M=259 F=276	M=0.64 F=-0.98	It is appropriate to use the G&P method in southern Turkish children: however, a revision is needed for better results and to minimise errors.	Needs some modification
Zabate et al, 2015	French	10-19	M=100 F=90	M=-0.19 F=-0.53	The G&P overestimated all males and females except boys who are 12 years and girls who are 11 and 18 years old. G&P can be used on French population but not without caution because of a tendency for this method to overestimate age.	Needs some modification
Ozturk et al, 2015	Central Turkey	9-17	M=186 F=249	M=-0.10 F=0.90	The G&P atlas was applicable to Caucasian boys of younger age groups and Caucasian girls of all ages. However, some improvement is needed.	Needs some modification
	Eastern Turkey	9-17	M=189 F=225	M=-0.90 F=-0.90		
Maggio et al, 2016	Western Australian	0-25	M=180 F=180	M=0.24 F=-0.14	The G&P standard is not suitable for the determination of legal majority.	Not clear
Chaumoitre et al 2017	France	1-21	M=1423 F=1191	M=-0.18 F= 0.06	The GP atlas is a reproducible and repeatable method that is still accurate for the present population, with a high correlation between BA and CA.	Applicable

A positive value of the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age, M = males, F = females

Table 3.3: Summary of studies that assessed the reliability of the G&P atlas in African children

Study	Origin/ ethnicity	Age (years)	N	Mean BA-CA (years)	Authors' conclusion	Applicability
Loder et al, 1993	African American	0-18	M=249 F=212	M=0.28 F=0.51	African girls were skeletally advanced by 0.4 to 0.7 years except during middle childhood. While BA for boys was only advanced during adolescence.	Applicable for boys but not for girls
Ontell et al, 1996	African American	3-18	M=95 F=65	M=0.28 F=0.55	African girls showed significant differences at all ages except middle childhood. G&P is applicable to African boys until adolescence.	Applicable for boys but not for girls
Mora et al, 2001	African American	0-19	M=135 F=139	M=-0.01 F=0.11	On average, the BA of 10% of prepubertal African American children was 2 SD above the normative data in the G&P atlas. The atlas is imprecise for African American children born after 1980.	Not applicable
Lewis et al, 2002	Malawian	1-28	M=93 F=46	M=-1.7 F=-1.5	The atlas is inaccurate for this group of children. Poor nutrition and chronic diseases such as malaria and diarrhoea which are endemic in Malawi are likely to be contributing factors.	Not applicable
Garamendi et al, 2005	Moroccan	13-25	M=144	M=-1.7	G&P has a high error rate and therefore should not be considered as an optimal diagnostic method.	Not applicable
Zhang et al, 2009	African American	0-18	M=179 F=170	M=-0.02 F=0.03	BA was relatively close to the CA in African American children.	Applicable
Dembetembe & Morris 2012	South African (black)	13-19 20-21	M=104 M=27	M=0.2 M= 2.1	Skeletal maturity as characterised by complete epiphyseal fusion occurred approximately 2.1 years later than G&P method. G&P is not directly applicable to African males.	Not applicable
Mansourvar et al, 2014	African American	8-15	M= 47	M=1.87	G&P is not reliable for assessment of children between 8 and 15 years.	Needs some modification
Govender et al 2018	South African	0-18	M= 61 F= 24	M= 0.3 F= 0.2	New SA assessment tools for South Africa is advised.	Needs some modification

A positive value of the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age, M = males, F = females

Table 3.4: Summary of studies that assessed the reliability of the G&P atlas in Asian children

Study	Origin/ ethnicity	Age (years)	N	Mean BA-CA (years)	Authors' Conclusion	Applicability
So & Yen 1990	Southern Chinese	11.9-12.3	F=117	F=0.6	Earlier skeletal maturation was demonstrated. Such a difference was contributed to by improved socioeconomic, nutritional and socio-hygienic conditions.	Not clear
So & Yen 1991	Southern Chinese	11.9-12.3	F=117	F=0.6	Earlier skeletal maturation was demonstrated. This is attributed to improved socio-economics condition.	Not clear
Ontell et al, 1996	Asian	3-18	M=63 F=30	M=-0.03 F=0.27	The G&P standard is applicable to Asian girls at all ages, while in boys it can only be applied from birth to 4 years old and from 7 to 13.3 years old.	Applicable (for girls but not boys.)
Krailassiri et al, 2002	Thai	7-19	M=139 F=222	M=0.8 F=0.8	Although the mean difference in BA and CA was equal in both sexes, males clearly differed from the G&P more frequently than females.	Not applicable
Chiang et al, 2005	Taiwan	7-19	M=230 F=140	M=0.82 F=-0.3	There is a discrepancy of more than one year between BA and CA in some age groups. We believe that some modification of the GP Atlas is necessary	Needs some modification
Griffith et al, 2007	Chinese	0-18	M=650 F=366	M=0.25 F=0.15	Hong Kong children appear to mature more slowly in the first decade but more quickly thereafter.	Needs some modification
Zhang et al, 2009	Asian	0-18	M=165 F=166	M=0.41 F=0.24	Asian children mature sooner than white children, especially between 10 and 13 years in girls and between 11 and 15 years in boy.	Not clear
Zafer et al, 2010	Pakistan	0-18	M=535 F=354	M=0.1 F=-0.19	This study suggests against the applicability of G&P in Pakistani children. Authors propose a cautious approach while employing G&P in this population to ensure appropriate clinical and medico-legal decisions.	Not applicable
Moradi et al, 2012	Iran	6-18	M=303 F=122	M=0.37 F=-0.04	Considering the possibility of a few months' difference, the G&P atlas can be used for the Iranian population.	Needs some modification

Soudack et al, 2012	Iseral	0-18	M=375 F=304	M=0.16 F=-0.04	There was no discrepancy between BA and CA in Israeli girls using G&P. There were discrepancies for boys, but these were small.	Applicable
Patil et al, 2012	India	1-19	M=194 F=181	M=0.69 F=0.64	G&P is not applicable to males, especially for age group 4 to 12 years. G&P is applicable to females except age groups 4-7 years, 9-10 years, 15-16 years. A new standard is needed for Indian children.	Not applicable
Shilpa et al, 2013	India	6-15	M=124 F=126	M=0.18 F=0.29	The G&P method of skeletal age estimation showed accuracy in only certain age groups in Bangalore South zone children.	Needs some modification
Awais et al, 2014	Pakistani	0-18	M=136 F=147	M=-1.3 F=0.06	G&P is reliable for girls in all age groups. However, G&P is not accurate for boys in whom it underestimated BA.	Not applicable
Mansourvar et al, 2014	Asian American	1-8	M=48	M=0.87	The delay in skeletal maturity was more than 2 years for the 4-6 years' age group. Some improvement is needed to enhance the precision of G&P.	Needs some modification
Mughal et al, 2014	Pakistan	4.5-9.5	M=139 F=81	M=-1.3 F=0.55	G&P standard significantly underestimates CA in Pakistani children between the ages of 4.5 to 9.5 years.	Not applicable
Rai et al, 2014	India	5-15	M=75 F=75	M=-0.07 F=-0.33	G&P atlas underestimates CA in children aged between 5 and 9 years.	Needs some modification
Kim et al, 2015	Korean	7-12	M=135 F=77	M=-0.48 F=-0.02	G&P is applicable to Korean children aged between 7 and 12 years.	Applicable
Mohammed et al, 2015	South India	9-20	M=330 F=330	M=-0.23 F=0.02	Mild underestimation of BA was noted in boys. G&P remains applicable to South Indian children.	Needs some modification
Patel et al, 2015	West India	6-16	M=90 F=90	M=-0.99 F=-0.40	G&P can be used in West Indian children aged between 6 and 16 years.	Applicable

A positive value of the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age, M = males, F = females

Table 3.5: Summary of studies that assessed the reliability of the G&P atlas in Hispanic children

Study	Origin Ethnicity	Age (years)	N	Mean Ba-CA (years)	Authors' Conclusion	Applicability
Jimenez et al, 1996	Spanish	0-14	M=139 F=100	M=-0.31 F=0.04	Boys show a delay of around 3 months with respect to the G&P atlas. Girls show a better fit to the corresponding (female) standard of the atlas.	Not clear
Ontell et al, 1996	Hispanic	3-18	M=105 F=69	M=0.28 F=0.38	The G&P atlas is applicable to boys aged between 4-13 years and to girls except during adolescence.	Needs some modification
Haiter-Neto et al, 2006	Brazilian	7-15	M=180 F=180	M=-0.2 F=0.1	The means of estimated and chronologic ages were similar in all age ranges. The standards can be used with some modification.	Needs some modification
Zhang et al, 2009	Hispanic	0-18	M=178 F=182	M=0.30 F=0.24	Hispanic children mature sooner than the G&P atlas, especially between 10 and 13 years of age in girls and between 11 and 15 years of age in boys.	Not clear
Santos et al, 2011	Portuguese	12-20	M=136 F=94	M=0.12 F=0.02	The G&P atlas can be used; however, caution must be taken at the end of the growing period.	Needs some modification
Mansourvar et al, 2014	Hispanic	15-18	M=43	M=0.3	The G&P method is reliable in Hispanic males.	Applicable
Alcina et al 2018	Spanish	0-19	M=590 F=560	M=0.33 F=0.01	Adjustment factors are proposed for each age and sex to reduce minimise systematic errors.	Not clear

A positive value of the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age, M = males, F = females

3.4.1 Meta-analysis based on ethnicity

1- Caucasian Females: 15 studies were included in the meta-analysis. These 15 studies presented moderate heterogeneity (I-squared 76%, Figure 3.2) but did not show any statistically significant results, with overall mean difference BA-CA of 0.13 years (95% CI: -0.17,0.43).

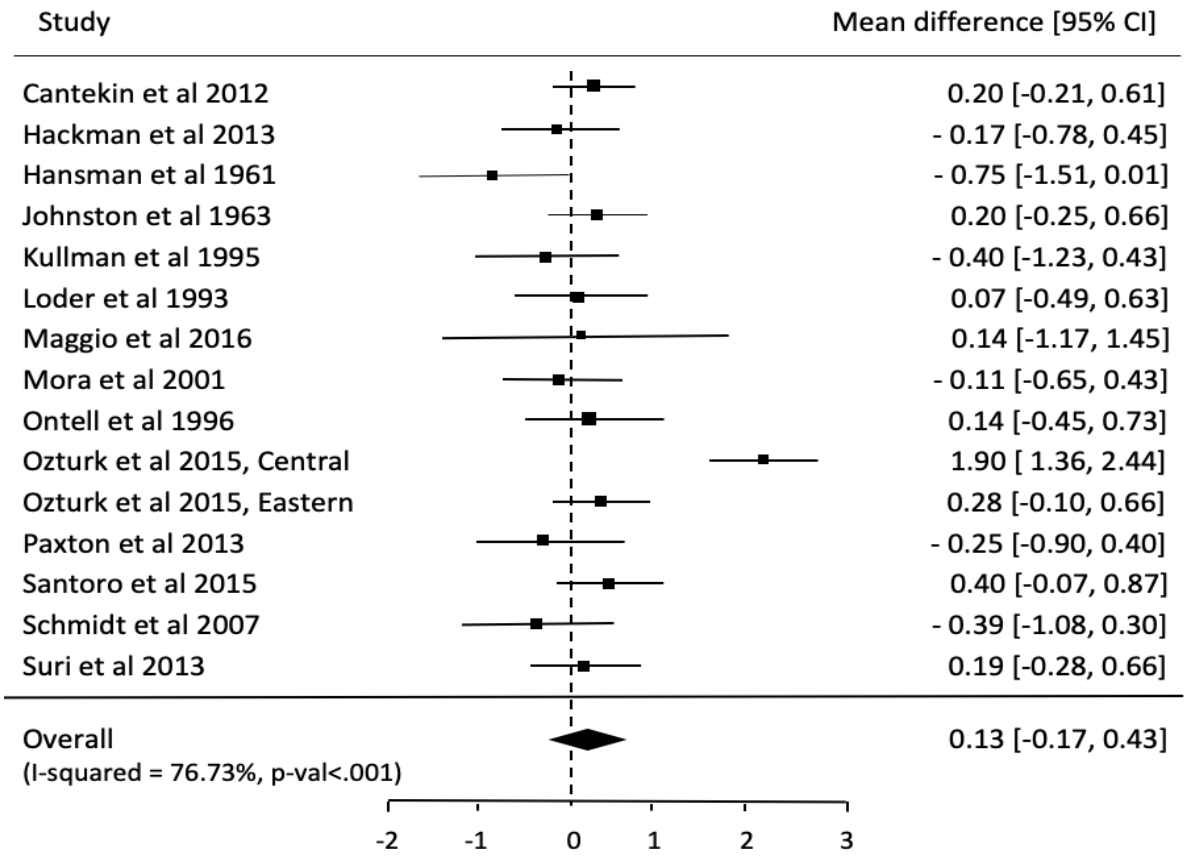


Figure 3.2: Forest plot of Caucasians females

2- Caucasian Males: 17 studies were included in the meta-analysis. These 17 studies presented low heterogeneity (I-squared 22%, Figure 3.3) and did not show any statistically significant results, with an overall mean difference BA-CA of -0.10 years (95% CI: -0.24,0.04).

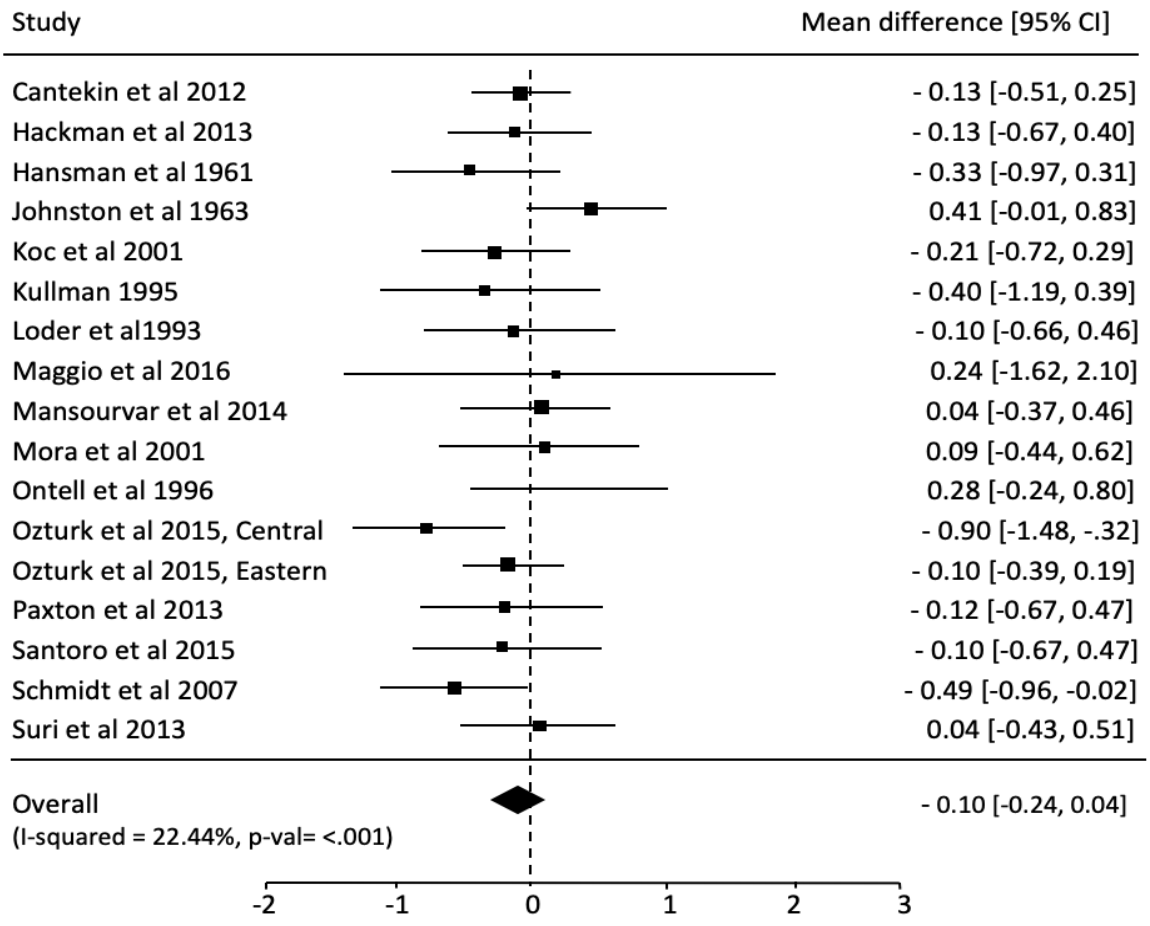


Figure 3.3: Forest plot of Caucasians males

3- African Females: only three studies were included in the meta-analysis. The three studies were homogeneous (I-squared 0%, Figure 3.4) and showed statistically significant results, with overall mean difference BA-CA of 0.37 years (95% CI: 0.04,0.69).

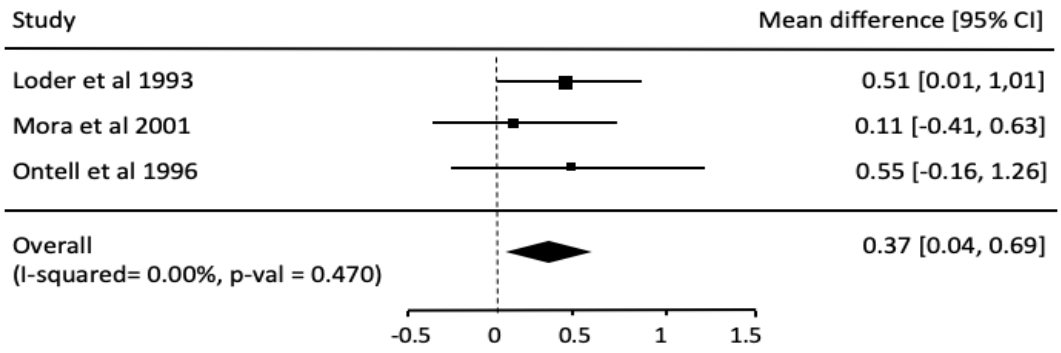


Figure 3.4: Forest plot of African females

4- African Males: only five studies were included in the meta-analysis. The five studies presented moderate heterogeneity (I-squared 78%, Figure 3.5) but did not show any statistically significant results, with overall mean difference BA-CA of 0.62 years (95% CI: -0.01,1.26).

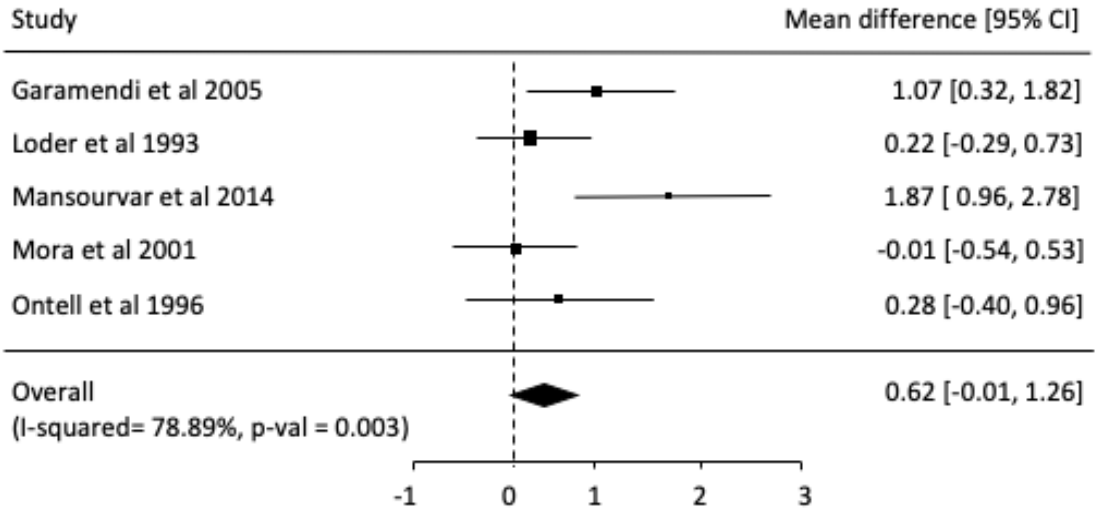


Figure 3.5: Forest plot of African males

5- Asian Females: only nine studies were included in the meta-analysis. These nine studies presented low to moderate heterogeneity (I-squared 27 %, Figure 3.6) but did not show any statistically significant results, with overall mean difference BA-CA of -0.10 years (95% CI: -0.32,0.12).

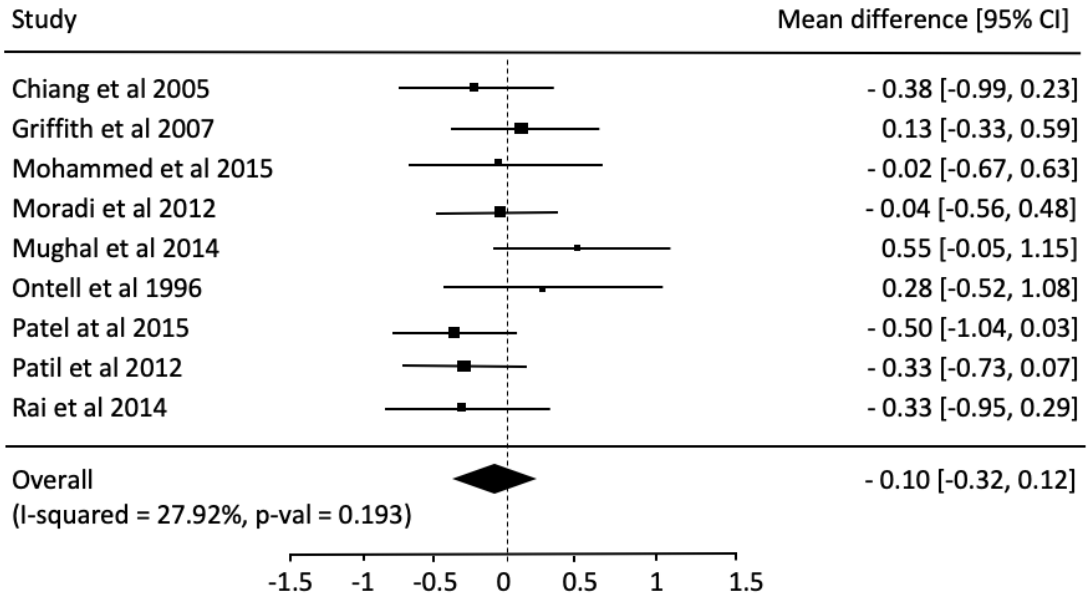


Figure 3.6: Forest plot of Asian females

6- Asian Males: 10 studies were included in the meta-analysis. The studies were highly heterogeneous (I-squared 82%, Figure 3.7) but did not show statistically significant results, with overall mean difference BA-CA of 0.15 years.

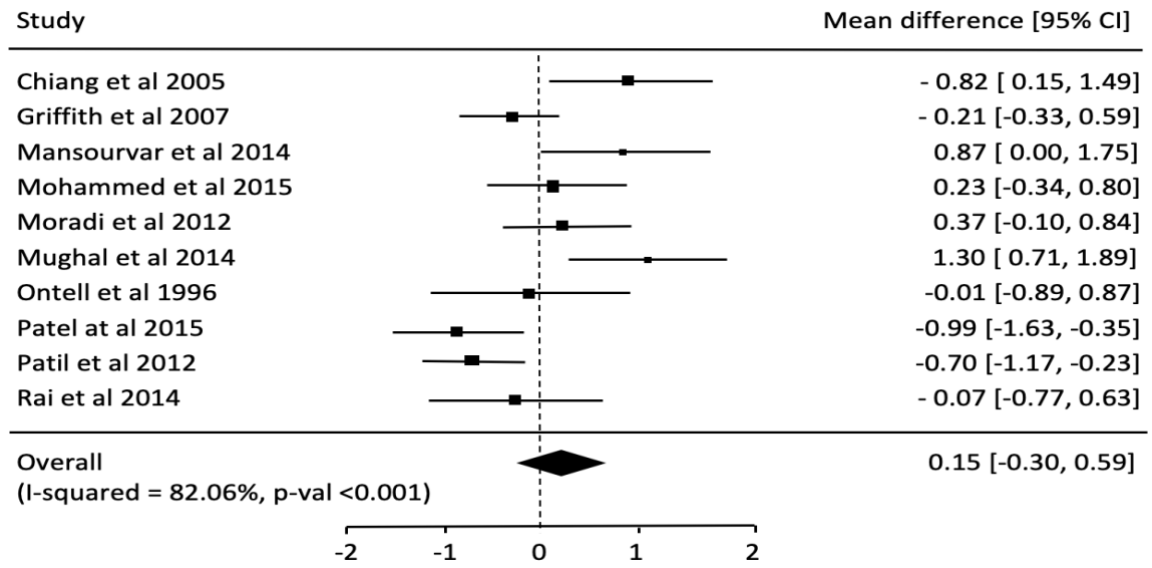


Figure 3.7: Forest plot of Asian males

7- Hispanic Females: only two studies were included in the meta-analysis. The two studies presented no heterogeneity (I-squared 0%, Figure 3.8) and did not show any statistically significant results, with overall mean difference BA-CA of 0.19 years (95% CI: -0.23,0.61).

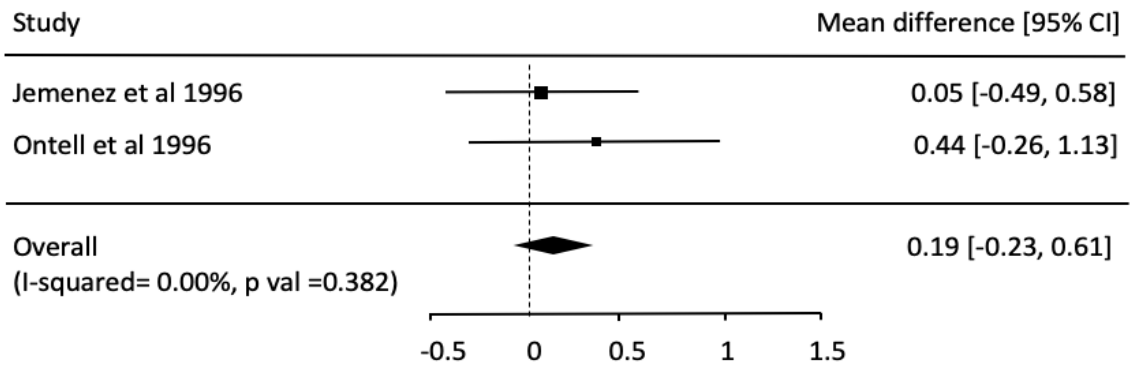


Figure 3.8: Forest plot of Hispanic females

8- Hispanic Males: only three studies were included in the meta-analysis. The three studies presented low heterogeneity (I-squared 11%, Figure 3.9) but did not show any statistically significant results, with overall mean difference BA-CA of -0.11 years (95% CI: -0.41,0.19).

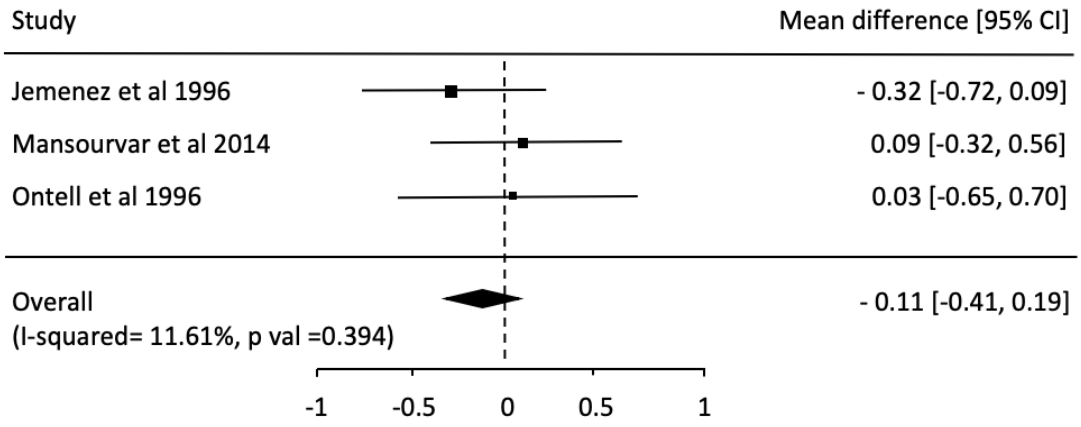


Figure 3.9: Forest plot of Hispanic males

A funnel plot was determined to assess bias or the present of any systematic heterogeneity. Large studies with higher power are placed toward the top, while lower powered studies are placed toward the bottom. Figure 3.10 shows minimal risk of publication bias within the included studies. The figure indicate that all of the studies are not large enough to be placed of the top. However, some studies have argued that visual interpretation of funnel plots is too subjective to be useful, as researchers had only a limited ability to correctly identify funnel plots from meta-analyses subject to publication bias

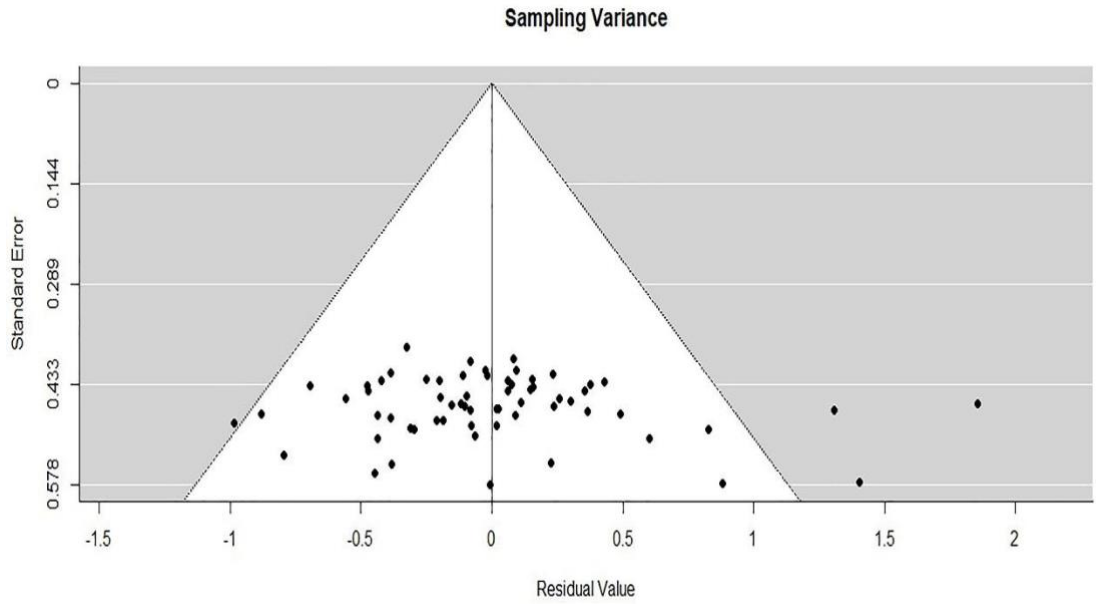


Figure 3.10: *Funnel plot of standard error plotted against residual value. Large studies with higher power are placed toward the top, while lower powered studies are placed toward the bottom.*

Additionally, meta-regression with covariates analysis (including sex and ethnicity as explanatory variables) was performed to further determine the effect of the ethnicity. In this regard, the coefficient for the Africans showed

statistical significance (p>0.05) with estimate being significantly different compare to Caucasian population (Table 3.6).

Table 3.6: Results of meta-regression with covariates (including sex and ethnicity as explanatory variable)

	Estimate	Standard error	Z value	CI% Lower	Upper	p
Intercept	-0.01	0.01	-0.18	-0.21	0.17	0.85
Sex	0.06	0.12	0.52	-0.17	0.30	0.59
Asian	0.01	0.14	0.08	-0.26	0.28	0.93
African	0.48	0.19	2.47	0.10	0.86	0.013
Hispanic	0.02	0.22	0.11	-0.42	0.47	0.90

Note: Caucasian is the reference group

3.4.2 Meta-analyses by yearly interval

The yearly interval meta-nalyses is expected to show a clearer picture as where the delay or/and advancement in BA usually occurred. The Mean difference between BA and CA for each year of chronological age was used to calculate perform the meta-analysis. For Caucasian males, seven studies were included (Johnston, 1963; Koc *et al.*, 2001; Büken *et al.*, 2007, 2009; Cantekin *et al.*, 2012; Suri *et al.*, 2013; Öztürk *et al.*, 2016). These studies did not show any statistically significant results. The mean difference BA-CA ranged from -0.32 years (at 13 years old) to 0.44 years (at 17 years old). For Caucasian females, six studies were included (Johnston, 1963; Büken *et al.*, 2007, 2009; Cantekin *et al.*, 2012; Suri *et al.*, 2013; Öztürk *et al.*, 2016). These studies did not show any statistically significant results, with mean

difference BA-CA ranging from -0.20 (at 10 years old) to 0.34 (at 14 years old).

For Asians, five studies were included (Chiang *et al.*, 2005; Griffith, Cheng and Wong, 2007; Patil *et al.*, 2012; Mohammed *et al.*, 2015; Patel *et al.*, 2015). The studies did not show any statistically significant results in females, with mean BA-CA ranging from -0.27 (at 6 years old) to 0.50 years (at 15 years old). In males however, the studies showed statistically significant results for the following ages; at 6 years with an overall mean difference BA-CA of -1.08 years (Figure 3.11), at 7 years with an overall mean difference BA-CA of -1.35 years (Figure 3.12– following page), at 8 years with an overall mean difference BA-CA of -1.07 years (Figure 3.13– following page), at 9 years with an overall mean difference BA-CA of -0.80 years (Figure 3.14– page 74) and at 17 years with an overall mean difference BA-CA of 0.50 years (Figure 3.15– page 74).

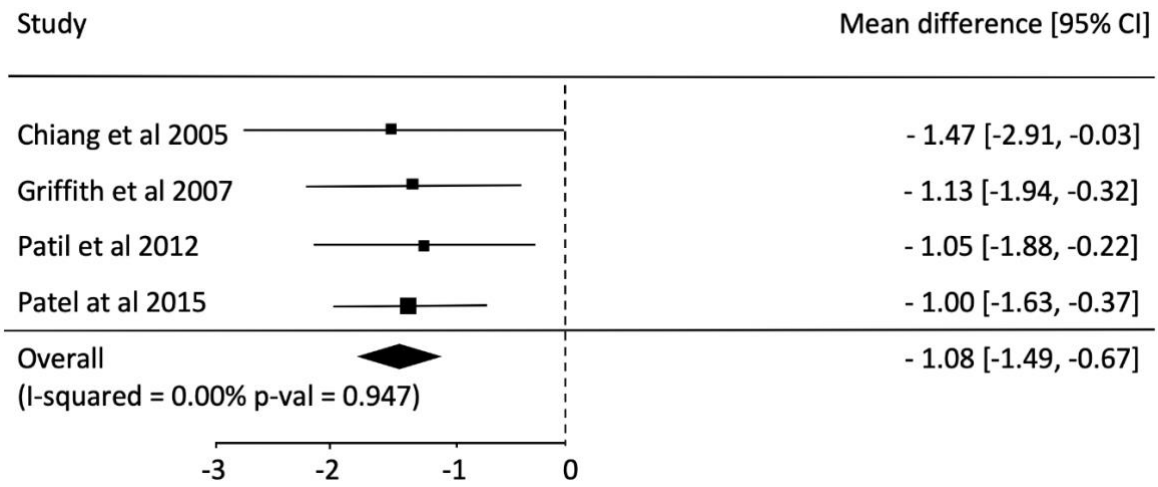


Figure 3.11: Forest plot of Asian males (6 years old) showing statistically significant results

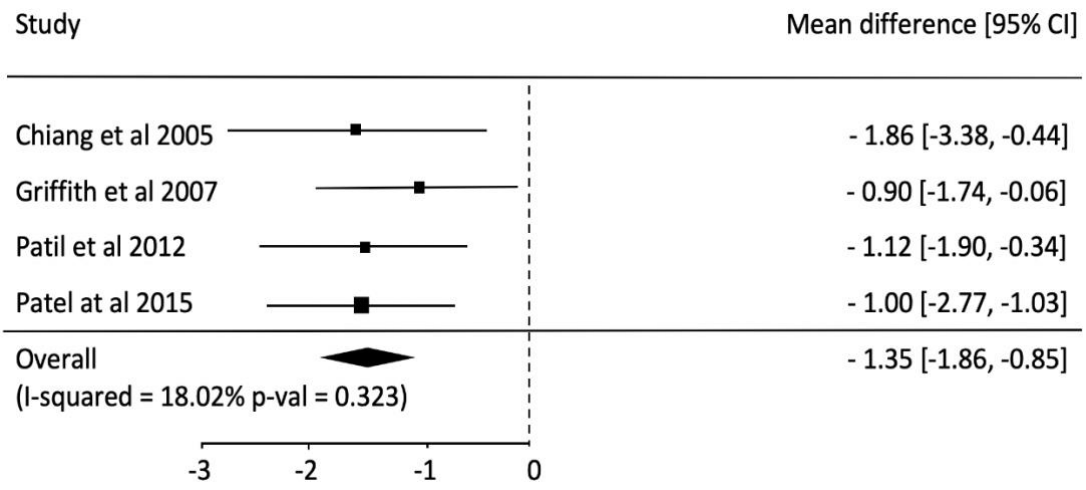


Figure 3.12: Forest plot of Asian males (7 years old) showing statistically significant results

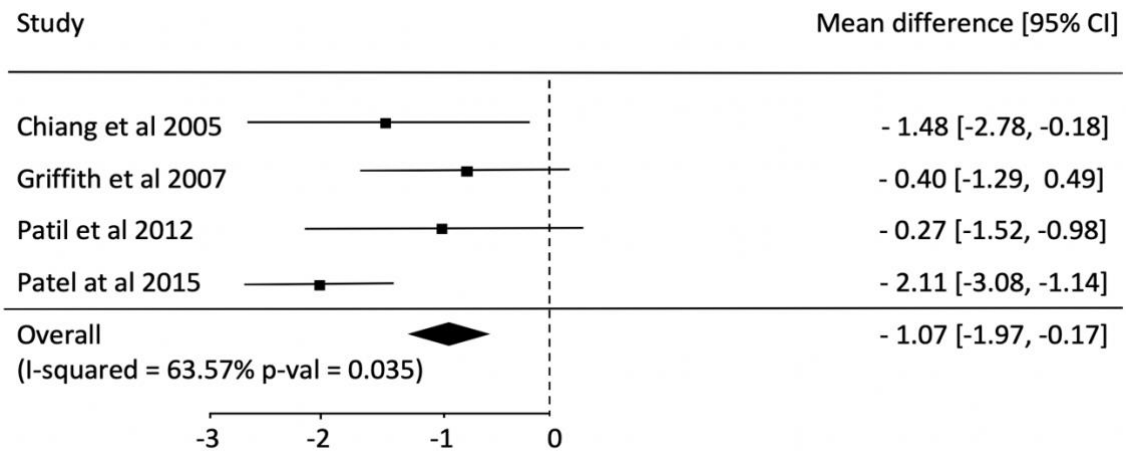


Figure 3.13: Forest plot of Asian males (8 years old)

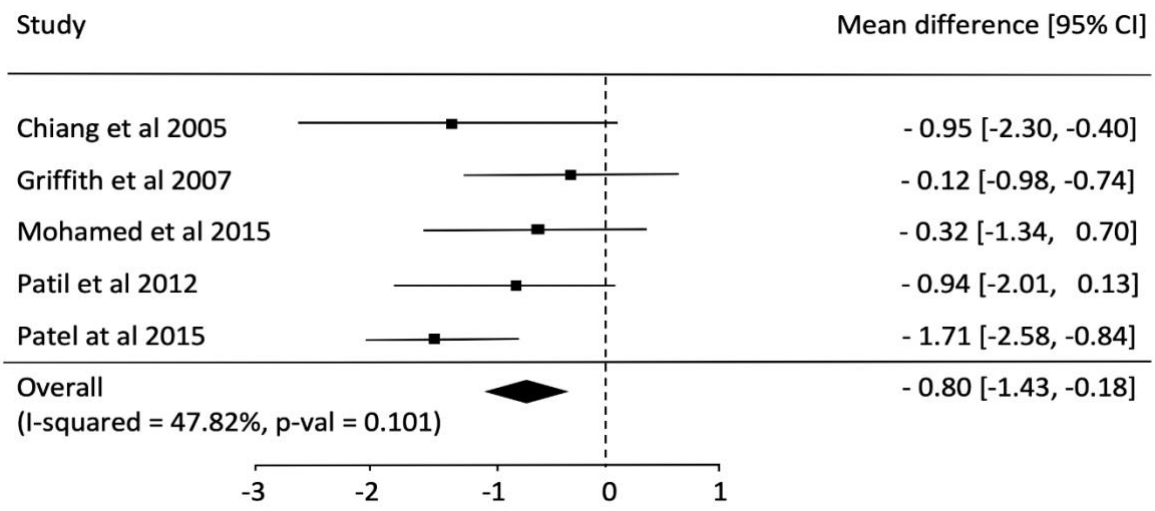


Figure 3.14: Forest plot of Asian males (9 years old)

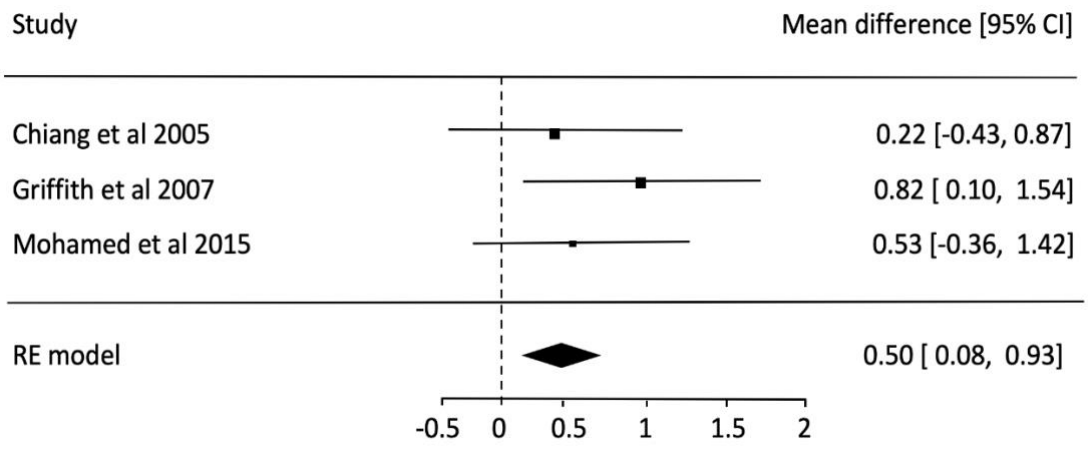


Figure 3.15: Forest plot of Asian males (17 years old)

Based on the results of the yearly interval meta-analysis, we produced graphs for Asians and Caucasians of both sexes (Figure 3.16, 3.17 – following page), which show BA according to our meta-analysis compared to BA as assessed by the G&P atlas. Maximum delay and advancement in BA compared to CA using the G&P atlas for Caucasians and Asians reported in the literature are summarised in appendix I, II, III, IV.

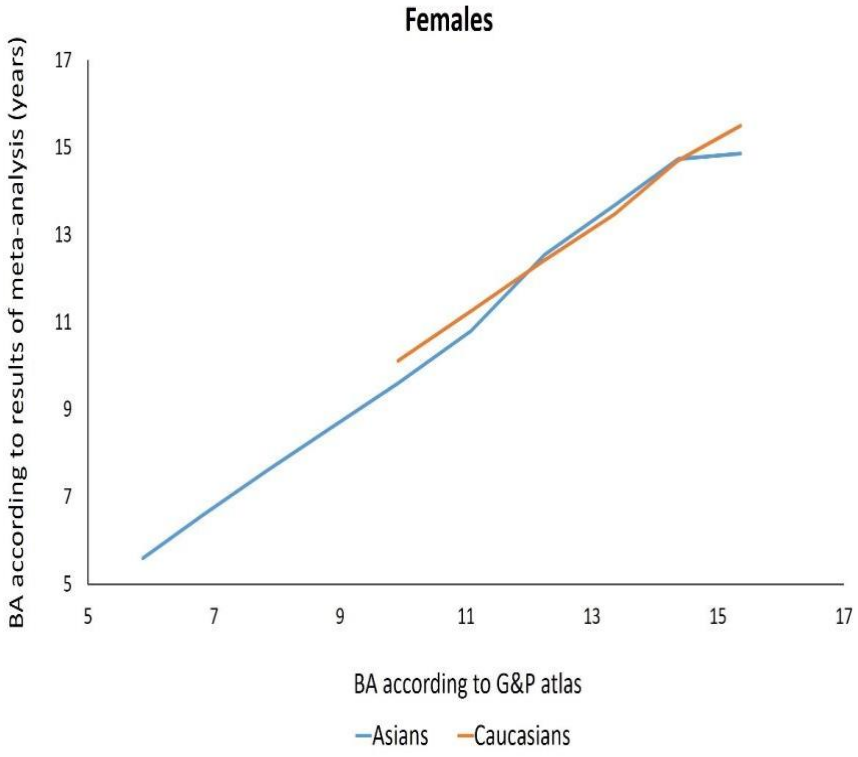


Figure 3.16: G&P bone age after adjustment based on meta-analysis (females)

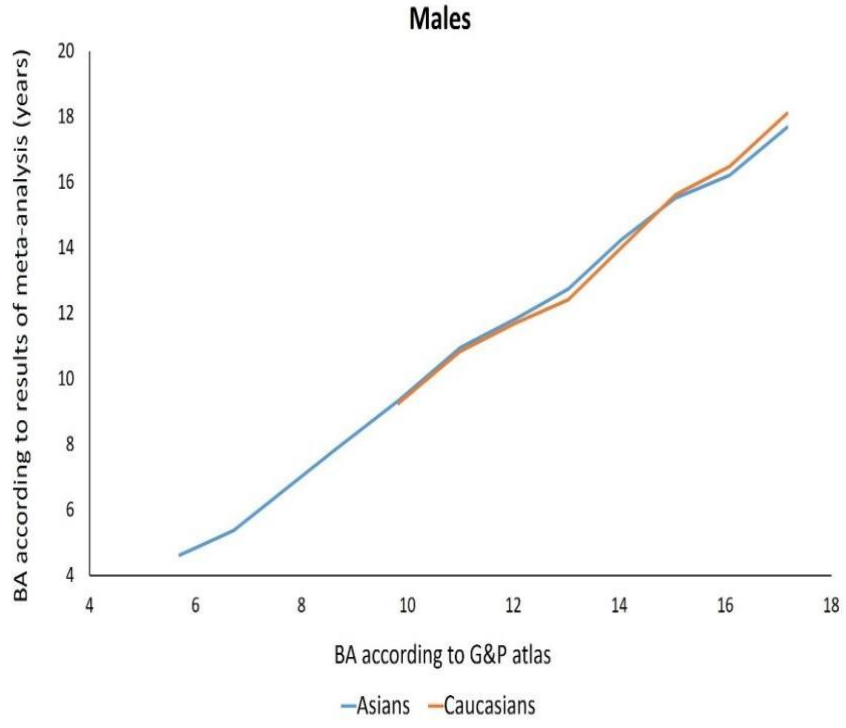


Figure 3.17: G&P bone age after adjustment based on meta-analysis (males)

3.5 Discussion

Bone age assessment is a frequently employed and (in the clinical setting) useful diagnostic technique. Its utility in assessing the age of immigrants and asylum seekers is less secure. Figures from the European Commission estimated that in 2016 about 95,000 unaccompanied minors migrated to Europe, of which more than half were Asians. Although there are no exact figures, many of these immigrants were without valid documents to prove their age. Being unable to prove age, or incorrectly assessing a child as an adult, can restrict the child from having access to their rights such as healthcare and education granted by the law in European countries (Feijen, 2008). Hence, it is important that reliable age estimation methods are used.

Concerned with the reliability of the G&P atlas for different ethnic populations, we considered it important to ascertain its applicability to healthy children. Additionally, Bias in studies can result in poor reproducibly and /or lead to distorted results and wrong conclusions, However, in this systematic review, results of the four studies with high risk of bias (Hansman and Maresh, 1961; Lewis, Lavy and Harrison, 2002; Schmidt *et al.*, 2007b; Rai, 2014) had little impact on (the statistical significance of) our results. The is because the population of these studies contributed less than 5% to the total included population in which only two studies (Hansman and Maresh, 1961; Schmidt *et al.*, 2007b) were included in the meta-analysis, which reduced their impact on sample size and results. A funnel plot shows the absence of a large study with high power as most of the studies scattered toward the bottom, however, minimal risk of publication bias was observed among the studies with three studies switched from the funnel plot (Figure 3.10–page

70) (Mansourvar *et al.*, 2014; Mughal, Hassan and Ahmed, 2014; Öztürk *et al.*, 2016).

The G&P atlas appears to be applicable to Caucasians, although some recent studies (included in the meta-analysis) have reported that bone age is advanced compared to chronological age in girls up to 13 years old and in boys aged 10 years and above, possibly highlighting the fact that children nowadays are maturing faster than when the atlas was established (Calfee *et al.*, 2010; Hackman and Black, 2013). Calfee et al. assessed the bone age of predominately Caucasian American adolescents (where the G&P atlas was developed). Their skeletal maturation exceeded their chronological age indicating advanced bone age. Perhaps this should not be surprising as Himes reported that skeletal maturation increases by about 0.22 to 0.66 years per decade (Himes, 1984).

This systematic review and meta-analysis showed no significant difference between BA and CA in Caucasians, which indicates that the G&P atlas is applicable to this group. This is in line with an earlier meta-analysis conducted by Serinelli et al in which no significant difference between BA and CA were found (Serinelli *et al.*, 2011). Note that Serinelli et al included a smaller number of studies, only reported the overall mean difference between BA and CA and did not account for individual age groups.

Concerning the Asian population, three studies recruited Asians living in America (Ontell *et al.*, 1996; A. Zhang, James W. Sayre, *et al.*, 2009; Mughal, Hassan and Ahmed, 2014) while the remaining 17 studies were all carried out in Asia. It seems that skeletal maturation does not conform to the G&P standard at least for some of those who live in East and South Asia. In

boys, delay in skeletal maturity during early and middle childhood was followed by advancement during adolescence. Our meta-analysis confirms that there are significant differences between BA and CA in Asian males in two ages categories; those aged 6 to 9 years and those aged 17 years. These differences are larger than the standard deviations reported in the G&P atlas for the corresponding age group (±0.77, ±0.84, ±0.90 years at age of 6, 7, 8, 9 years respectively), which may have an impact on patient diagnosis and management. In the clinical context, a healthy Asian boy in early childhood could be misdiagnosed as having delayed bone age when using the G&P atlas. The significant advancement in BA compared to CA in Asian males at age 17 is important because this is a critical age in the forensic/legal context, with the individual judged by adult standards in certain legal instances (Cole, 2015).

The G&P standard also seems to be imprecise for Africans. Our meta-analysis of three papers (Ontell *et al.*, 1996; Mora *et al.*, 2001; Mansourvar *et al.*, 2014) showed significant advancement in bone age of females at all ages ($p<0.01$). Results from meta-regression with covariates supports this difference with BA in Africans being statistically different (table 3.6). Although our meta-analysis did not show significant difference between BA and CA in African males, some studies reported significant advancement ($p<0.01$) in adolescence among African Americans males (Loder, 1993; Ontell *et al.*, 1996; Mansourvar *et al.*, 2014) Concerning those living in Africa, some studies have shown retardation of bone age among males and females (Lewis, Lavy and Harrison, 2002; Garamendi *et al.*, 2005; Dembetembe and Morris, 2012; Govender and Goodier, 2018). It is difficult to attribute these

variations between Africans only to differences in socioeconomic status, as they were not reported across all studies.

In contrast, the G&P standard appears appropriate for the Hispanic population until adolescence. Our meta-analysis shows no significant difference between BA and CA although only three studies were included (Jiménez-Castellanos *et al.*, 1996; Ontell *et al.*, 1996; Mansourvar *et al.*, 2014; Alcina *et al.*, 2018). However, Zhang et al, reported that the G&P significantly overestimated males aged between 10 and 13 years (A. Zhang, James W. Sayre, *et al.*, 2009).

In the current review, a final analysis was performed combining Asians and Hispanics in order to compare our results to those of Serinelli et al, who used the Cavalli-Sforza classification of ethnicity (Cavalli-Sforza *et al.*, 1995), in which Asians and Hispanics are under one ethnic group (Mongoloid). Our meta-analysis of Asians-Hispanics for both females and males showed no significant results (appendix V and VI). This is in contrast to Serinelli's meta-analysis, in which the G&P atlas significantly overestimated chronological age (Serinelli *et al.*, 2011). However, Serinelli et al included only three papers for the Mongoloid population; one related to the Asian population and two to the Hispanic population. One of these latter two studies (Holderbaum *et al.*, 2005), was excluded from the current systematic review because it included unhealthy children.

The major limitation identified in this review is the difficulty in separating ethnicity from socio-economic status. Relatively few studies reported the socioeconomic status of their sample. Children in these studies seemed to follow the same pattern of advancement and delay in bone age as their

peers of the same ethnicity in other studies. When bone age is accelerated, new social and cultural factors rather than economic conditions have been suggested to be the main drive (Büken *et al.*, 2009). However, our results suggest ethnicity should also be considered when assessing bone age. A further limitation of the study is the failure to calculate the mean absolute and root mean square errors, which might have further confirmed the accuracy of the G&P atlas in relation to each population. However, the mean of each variable (BA and CA) was only available for 13 studies (Jiménez-Castellanos *et al.*, 1996; Koc *et al.*, 2001; Chiang *et al.*, 2005; Haiter-Neto *et al.*, 2006; Büken *et al.*, 2007; Griffith, Cheng and Wong, 2007; Cantekin *et al.*, 2012; Patil *et al.*, 2012; Patel *et al.*, 2015; Gungor *et al.*, 2015a; Kim, Lee and Yu, 2015; Mohammed *et al.*, 2015; Öztürk *et al.*, 2016) and for these 13 studies, individual observations were not provided, therefore the mean error could not be calculated.

3.6 Conclusion

This systematic review revealed that the ethnicity/origin of the child can influence the applicability of the G&P standard. The G&P standard is imprecise and should be used with caution in Asian and African populations, particularly when assessing age for forensic/legal purposes. Some caution is also required for Hispanics (particularly males). The G&P atlas can be used with most confidence in Caucasians. There is a complex inter-relationship between the impacts of socioeconomic status and ethnicity on bone age using the G&P atlas, which no study has clearly set out to address. Clinicians should be aware of the limitations of the G&P method presented in this review.

CHAPTER 4

Applicability of Two Commonly Used Bone Age Assessment Methods to 21st Century UK Children

Alshamrani K[1], Offiah A[1,2]

[1] Department of Oncology & Metabolism, University of Sheffield, UK

[2] Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield, UK

Correspondence: Khalaf Alshamrani, Academic Unit of Child Health, Damer Street Building, Sheffield, S10 2TH

kalshamrani1@sheffield.ac.uk

Under review (European Radiology)

4.1 Abstract

Background: Some authors stated that improved socioeconomic status renders the Greulich and Pyle (G&P) and Tanner-Whitehouse (TW3) methods unreliable for bone age assessment. If improved socioeconomic status has indeed advanced skeletal maturation, then BoneXpert would be expected to return significantly advanced bone age compared to chronological age.

Methods: BoneXpert was used to assess bone age on 392 hand trauma radiographs (206 males, 257 left). Paired sample *t* test was performed to assess the difference between mean bone age (BA) and mean chronological age (CA). Socioeconomic status (according to the index of multiple deprivation) was recorded for each child.

Results: Numbers of children living in low, average and high socioeconomic areas were 216 (55%), 74 (19%) and 102 (26%) respectively. However, TW3 underestimated females' age after the age of 3 years with significant differences between BA and CA (-0.43 years ± 1.05 p = <0.001) but not in males (0.01 years ±0.97 p = 0.76). Of the difference in females, 17.8% was accounted for by socioeconomic status.

Conclusion: No significant difference exists between BoneXpert-derived BA and CA when using the G&P atlas. There was a statistically significant underestimation of BoneXpert-derived BA compared to CA in females when using TW3, particularly in those from low and average socioeconomic backgrounds. Improved standard of living has not led to significant advancement in skeletal maturation.

4.2 Introduction

Age estimation is of increasing significance, particularly in forensic and legal contexts. Situations where chronological age is undocumented or is unable to be proven have increased, particularly at geographical borders where conflicts or crises are occurring. The estimation is that approximately 160,000 unaccompanied children entered European countries during 2015 and 2016 (1). Although there is no precise figure available regarding the number of children with a valid documented age, authorities have faced challenges in estimating some of their ages, because many will have lost their documents or may have falsified their age. Hence, it is crucial to have a reliable and appropriate method of determining bone age (D. D. Martin *et al.*, 2011). Bone age assessment also plays an important role in clinical practice, permitting an investigation of whether bone maturity is occurring at an equivalent rate as the chronological ageing (CA) process. In this context, bone age assessment is useful for managing children with skeletal dysplasias and endocrine disorders, as well as planning for orthopaedic procedures (Ritz-Timme *et al.*, 2000).

Numerous approaches have been developed as a means of determining bone age (BA). Among these methods, two techniques are widely utilised based on left hand and wrist radiographs, namely the Greulich and Pyle (G&P) and Tanner and Whitehouse (TW) methods (Greulich and Pyle, 1959; Tanner *et al.*, 2001). The G&P method is based on matching the child's hand radiograph to standard plates provided by the G&P atlas, thus this method compares the hand's general maturational status. The population providing the G&P standard atlas were originally North American Caucasians of good

socioeconomic status in 1938. In contrast to the G&P atlas, the TW method undertakes an assessment and scoring of skeletal maturity for each individual hand and wrist bone. Data provided by the Harpenden Longitudinal Growth Study enabled the TW method's development. In 2001, the TW3 method replaced the TW1 and TW2 methods as a result of documented secular change.

The data that formed the TW3 method was collected from European and American Caucasian children of average socioeconomic status during the 1980s and 1990s. Following the introduction of G&P and TW3 standards, numerous investigations have been undertaken internationally, in order to identify the extent to which these standards are relevant to various populations. This issue is significant, especially in light of the growing volume of studies concluding that certain techniques are inappropriate for particular ethnic groups and as a result of improvements in socioeconomic status (So and Yen, 1990; Schmeling *et al.*, 2006; Ulijaszek, 2006; Büken *et al.*, 2009).

BoneXpert software was developed in 2009, enabling automatic calculation of bone age, according to the G&P and TW3 methods (Thodberg *et al.*, 2009). The software provides standard deviation scores for each hand radiograph, thus assisting the comparison of a child's bone age with healthy children of the same sex and age. There are several advantages in utilising this software tool, including eliminating observer variability and saving rating times.

This study aims to use BoneXpert to test the applicability of the G&P and TW3 methods to United Kingdom (UK) children born in the 21st century, whose standard of living (across all socio-economic categories) is likely to be higher than those of the children used to develop the G&P and TW3 methods and whose bone age is therefore likely to be advanced compared to chronological age (Easterlin, 2000).

4.3 Materials and Methods

4.3.1 Study design

Hand radiographs performed between 2010 and 2016 on children aged between 2 and 15 years presenting following trauma, to the Emergency Department of Sheffield Children's Hospital, United Kingdom, were retrospectively identified from the Picture Archiving and Communication System. Those with a specific request for BA estimation were excluded. Demographic data including sex, ethnicity and CA at the time of the radiograph were recorded. All procedures performed in this study were in accordance with the ethical standards of our institution.

Socioeconomic status of recruited children was documented using the index of multiple deprivation (IMD) (Deprtment of Communities and Local Government, 2015). The IMD measures deprivation based on income, employment, education, health and disability, crime, barriers to housing and service and living environment. The English IMD 2015 data combined with postcode were used to classify the socioeconomic status of the children in which the IMD scores are ranked for each small area within England from 1 to 32,844. IMD scores below 10,894 are deemed to be areas of low socioeconomic status, between 10,895 and 21,788 are average, and above

21,789 are of high socioeconomic status. BoneXpert software (Visiana, Holte, Denmark) was utilised to analyse the hand radiographs. All radiographs were acquired via a computed radiography system and were in DICOM format. The default ethnicity for analysing the radiographs was Caucasian, because the software does not include ethnicity-specific standard deviation scores (SDS). Radiographs were omitted if the software failed to analyse them.

4.3.2 Statistical analysis

Statistical analysis was undertaken via SPSS version 24 for PC (IBM, Armonk, New York). The mean variation for BA and CA was determined for each child by subtracting BA from CA (BA-CA). Therefore, a positive value indicates advanced BA, whereas a negative value indicates delayed BA, compared to CA. The significance of the differences was calculated using a paired sample *t* test. Statistical analysis was undertaken separately for both sexes, in relation to each method (G&P and TW3) and repeated for both sexes for Caucasians only, to investigate the effect of ethnicity on the results. Analysis was also performed to determine the effect of readings from left and right hands. The effect of socioeconomic status was evaluated using the one-way ANOVA test.

Approval was obtained from the Health Research Authority at Yorkshire and Humber. The need for full Research Ethics Committee approval was waived for this retrospective study of hand radiographs.

4.4 Results

In total we identified 401 potentially eligible hand and wrist radiographs of which 9 were omitted due to BoneXpert failing to provide a reading, therefore results are from 392 radiographs, comprising 206 males, 296 Caucasians, 71 Asians, 20 Africans and 5 mixed (Caucasian/Asian). Figure 4.1 and 4.2 (this and following page) illustrates the number of left and right-hand radiographs per age and sex. In regard to socioeconomic status, *216 (*55%), 74 (19%) and 102 (26%) children were of low, average and high socioeconomic status respectively.

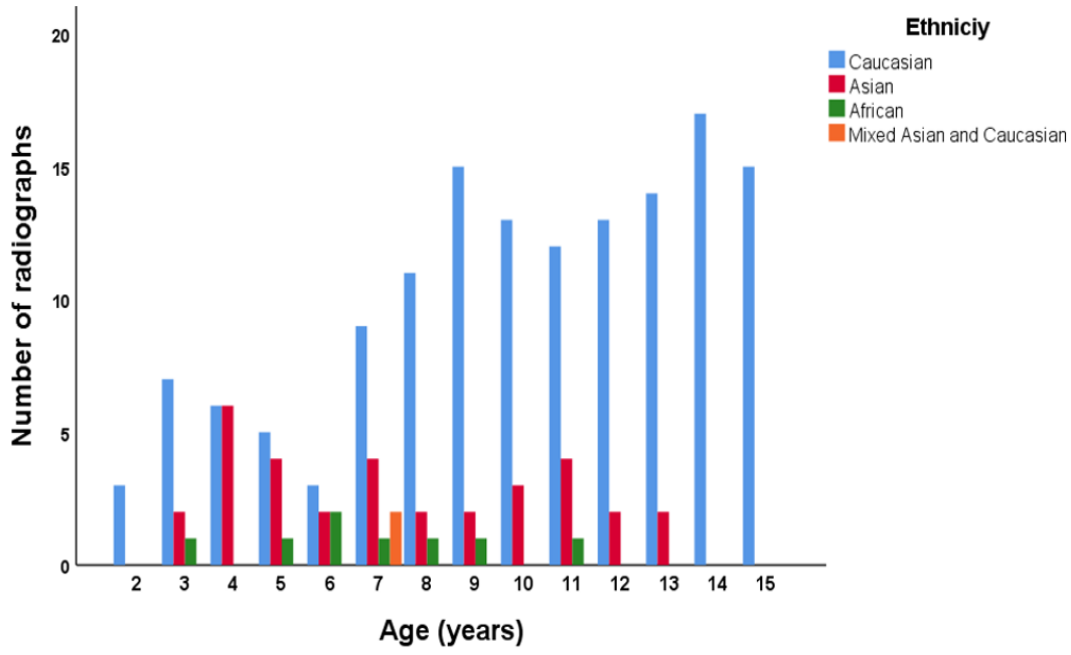


Figure 4.1: Number of hand radiographs by age and ethnic group (females). Radiographs of the Caucasians appear to be the dominant.

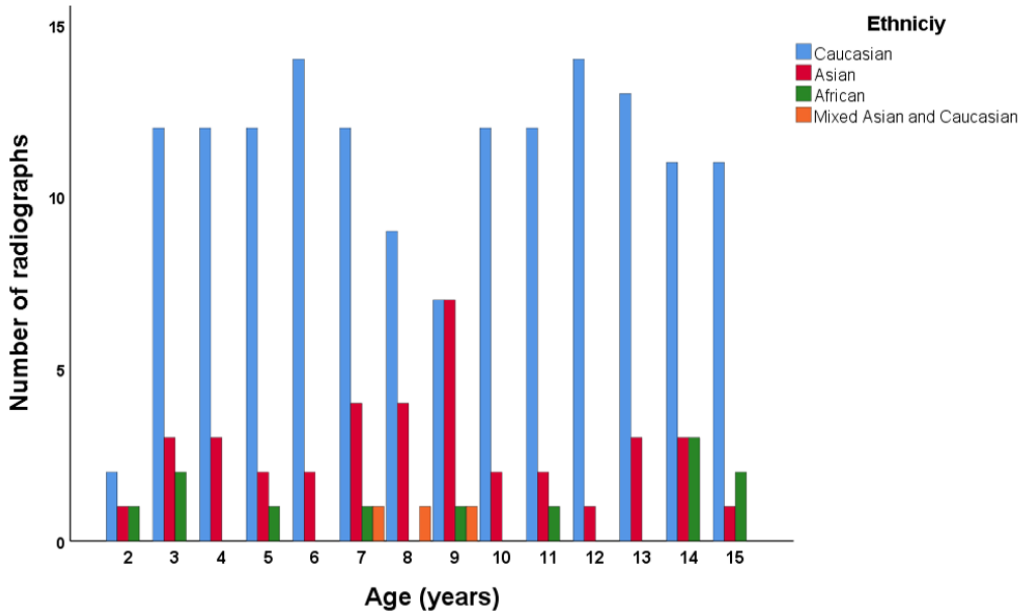


Figure 4.2: Number of hand radiographs by age and ethnic group (males). Radiographs of the Caucasians appear to be the dominant.

4.4.1 G&P atlas

Concerning G&P, mean difference between BA and CA ranged from 33 months underestimation to 36 months overestimation in both females and males. Although differences were not significant, G&P underestimated females' ages by 1 month and overestimated males' ages by 1.6 months (Table 4.1 - following page).

Table 4.1: Mean difference (±SD) in years, between BA and CA in females and males

All Ethnicities	**Sex**	**Mean CA (±SD)**	**Mean BA (±SD)**	**Mean difference BA-CA**	**p value**
G&P BA vs CA	Female	9.96 (±3.7)	9.89 (±3.8)	-0.07 (±1.05)	0.32
	Male	9.32 (±3.9)	9.45 (±4)	0.13 (±1.07)	0.06
TW3 BA vs CA	Female	9.96 (±3.7)	9.53 (±3.5)	-0.43 (± 1.05)	<0.001
	Male	9.32 (±3.9)	9.34 (±3.7)	0.02 (±0.97)	0.76
Caucasians Only	**Sex**	**Mean CA (±SD)**	**Mean BA (±SD)**	**Mean difference BA-CA**	**p value**
G&P BA vs CA	Female	10.57 (±3.6)	10.45 (±3.8)	-0.12 (±1.06)	0.17
	Male	9.44 (±3.8)	9.46 (±4.1)	0.02 (±1.05)	0.79
TW3 BA vs CA	Female	10.57 (±3.6)	10.03(±3.5)	-0.54 (±0.96)	<0.001
	Male	9.44 (±3.8)	9.31 (±3.8)	-0.13 (±0.64)	0.091

A positive value of the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age.

BA was lower than CA in 51% of females and 44% of males, while being equal in 1% of males. With the cohort divided into yearly intervals, G&P overestimated females aged from 2 to 7 years by between 0.8 and 6 months, apart from at 4 years of age. This overestimation was statistically significant ($p < 0.05$) at age 6, in females (Table 4.2).

Table 4.2: Mean difference (±SD) in years, between G&P BA and CA

		All Ethnicities			Caucasians Only		
Males	**Age (years)**	**Mean**	**(±SD)**	**p value**	**Mean**	**(±SD)**	**p value**
	2	0.07	0.43	0.784	0.19	0.09	0.20
	3	-0.08	0.96	0.747	-0.41	0.75	0.08
	4	0.01	0.90	0.962	-0.14	0.95	0.61
	5	0.00	1.10	0.989	-0.11	0.98	0.69
	6	-0.13	0.80	0.530	-0.28	0.70	0.15
	7	0.24	1.05	0.346	0.11	0.97	0.68
	8	0.43	1.29	0.231	0.16	1.27	0.71
	9	0.49	1.23	0.132	0.65	1.46	0.28
	10	0.33	1.00	0.240	0.32	1.09	0.31
	11	0.34	1.13	0.260	0.09	1.09	0.76
	12	-0.13	1.00	0.612	-0.17	1.02	0.52
	13	0.14	1.09	0.620	-0.11	0.99	0.68
	14	0.02	1.06	0.953	0.22	1.05	0.78
	15	0.20	1.52	0.632	0.35	1.56	0.46
Females							
	2	0.11	0.07	0.121	0.10	0.07	0.12
	3	0.35	0.73	0.168	0.56	0.69	0.07
	4	-0.21	0.96	0.468	-0.1	0.75	0.57
	5	0.12	0.95	0.710	0.1	0.78	0.97
	6	0.50	0.39	0.015	0.69	0.34	0.07
	7	0.07	0.76	0.725	-0.29	0.50	0.12
	8	-0.46	1.06	0.130	-0.65	0.83	0.02
	9	-0.01	0.95	0.975	0.04	0.98	0.86
	10	-0.13	1.18	0.659	-0.19	1.24	0.58
	11	-0.47	1.13	0.107	-0.49	1.05	0.12
	12	-0.94	0.99	0.002	-1.06	0.7	0.00
	13	0.12	1.11	0.673	0.1	1.17	0.75
	14	0.49	1.45	0.187	0.48	1.45	0.18
	15	-0.05	0.87	0.822	-0.51	0.86	0.82

A positive value of the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age.

After 7 years of age, G&P consistently underestimated females until 12 years of age by between 0.1 and 11 months, with underestimation being statistically significant ($p < 0.05$) at 12 years of age (Table 4.2). Concerning males, G&P overestimated in all age groups apart from at 3, 6 and 12 years of age, with no statistical difference between BA and CA. ANOVA test showed no statistical difference between low, average and high socioeconomic status groups when using the G&P atlas for either females or males. However, in females, the mean difference between BA and CA tended to be larger in low and average socioeconomic status groups, while in males, the difference tended to be larger within the higher socioeconomic status group. Independent *t* test showed no significant difference between the mean difference of BA and CA when acquired from either the left hand or the right hand for G&P ($p=0.58$ females, $p=0.07$ males). Distribution of the mean difference between CA and BA estimated via G&P for each sex is illustrated in Figures 4.3 and 4.4 (this and following page).

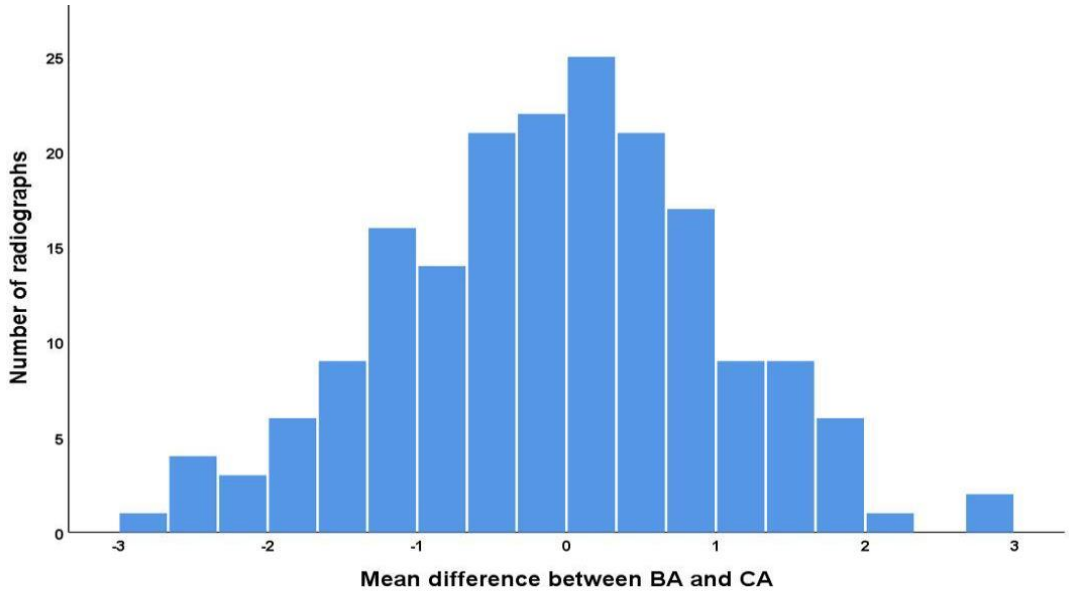


Figure 4.3: Mean difference between G&P-BA and CA (in years, females), showing normal distribution.

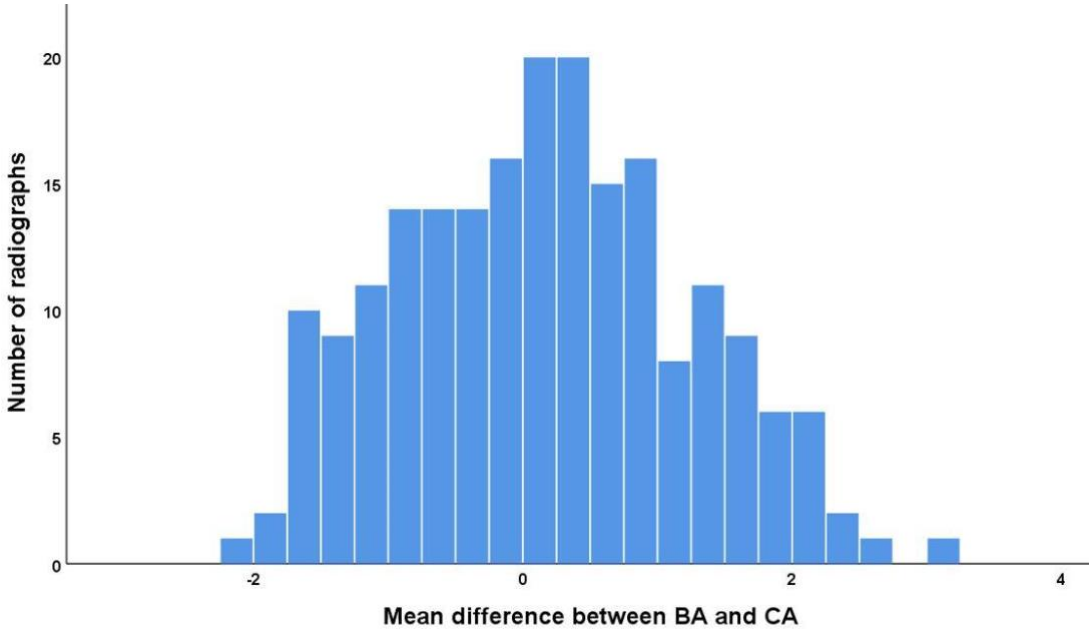


Figure 4.4: Mean difference between G&P-BA and CA (in years, males) showing normal distribution.

4.4.2 TW3 method

Concerning TW3, overall mean difference between BA and CA showed a statistically significant difference in females but not in males. The mean difference between BA and CA ranged from 37 months underestimation to 32 months overestimation in both females and males. BA was lower than CA in 64.5% of females and 49.5% of males, while being equal in 0.5% of males. TW3 underestimated females' ages by between 2 and 15 months (mean 5.2 months, $p < 0.01$) for all chronological age groups above 3 years (Table 4.3–following page).

Table 4.3: Mean difference (±SD) in years, between TW3 BA and CA

		All Ethnicities			Caucasians Only		
	Age (years)	**Mean**	**SD**	**p value**	**Mean**	**SD**	**p value**
Males	2	0.61	0.29	0.02	--	--	--
	3	0.34	0.76	0.08	0.08	0.73	0.74
	4	0.11	0.76	0.59	-0.18	0.81	0.51
	5	0.10	1.02	0.69	-0.6	1.14	0.88
	6	-0.08	0.96	0.75	-0.2	0.95	0.32
	7	0.40	1.02	0.11	0.33	0.80	0.30
	8	0.36	0.98	0.18	0.10	1.06	0.79
	9	0.23	1.00	0.38	0.76	0.75	0.05
	10	-0.07	0.76	0.73	0.05	0.84	0.87
	11	-0.12	1.05	0.67	-0.47	1.04	0.2
	12	-0.50	1.07	0.09	-0.68	1.03	0.05
	13	-0.23	1.08	0.40	-0.9	0.70	<0.01
	14	-0.32	1.03	0.21	-0.33	1.22	0.55
	15	-0.45	1.09	0.14	-0.33	1.21	0.43
Females							
	2	0.34	0.19	0.09	0.33	0.19	0.09
	3	0.44	0.45	0.01	0.73	0.30	0.01
	4	-0.21	0.58	0.23	-0.13	0.50	0.58
	5	-0.26	0.74	0.29	-0.10	0.58	0.73
	6	-0.18	0.46	0.33	0.07	0.56	0.85
	7	-0.29	0.78	0.15	-0.7	0.56	0.01
	8	-0.75	1.15	0.03	-0.61	0.60	0.03
	9	-0.24	0.95	0.30	-0.32	1.13	0.39
	10	-0.38	1.14	0.19	-0.21	1.21	0.62
	11	-0.72	1.03	0.01	-0.76	1.13	0.09
	12	-1.28	0.93	<0.01	-1.69	0.36	<0.01
	13	-0.27	1.28	0.408	-0.47	0.73	0.14
	14	-0.33	1.04	0.21	-0.28	1.01	0.39
	15	-0.88	0.32	<0.01	-0.87	0.19	<0.01

A positive value of the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age.

TW3 significantly underestimated females at 8, 11, 12 and 15 years of age ($p < 0.05$). In males, TW3 underestimated age for those 10 years or above; this was statistically significant in Caucasians at ages 9, 12 and 13 years. Observed differences were larger and significant ($p<0.001$) in females of low and average socioeconomic status (Table 4.4); with 17.8% of the variation between CA and TW3 BA as assessed by BoneXpert being accounted for by socioeconomic status. Distribution of the mean difference between CA and BA estimated via TW3 methods for each sex is illustrated in Figures 4.5 and 4.6 (following page).

Table 4.4: Mean difference (±SD) in years, between G&P, TW3 and CA in three socioeconomic groups

		n	Females	Males
Mean difference between BA and CA (±SD)				
G&P - CA				
All ethnicities	Low	213	- 0.2 (1.1)	0.1 (1.1)
	Average	75	-0.3 (1)	0.14 (0.9)
	High	101	0.06 (1)	0.2 (1)
Caucasians	Low	149	-0.1 (1)	-0.04 (1.1)
	Average	59	-0.3 (1)	0.08 (0.8)
	High	86	-0.02 (1.1)	0.14 (1.1)
TW3 - CA				
All ethnicities	Low	213	-0.5 (0.8)*	-0.01 (1)
	Average	75	-0.6 (0.9)*	-0.02 (0.7)
	High	101	-0.3 (0.8)*	0.06 (0.9)
Caucasians	Low	149	-0.5 (0.9)*	-2.4 (0.9)
	Average	59	-0.6 (0.9)*	-0.7 (0.8)
	High	86	-0.4 (0.9)*	-0.2 (0.9)

**P value < 0.01*

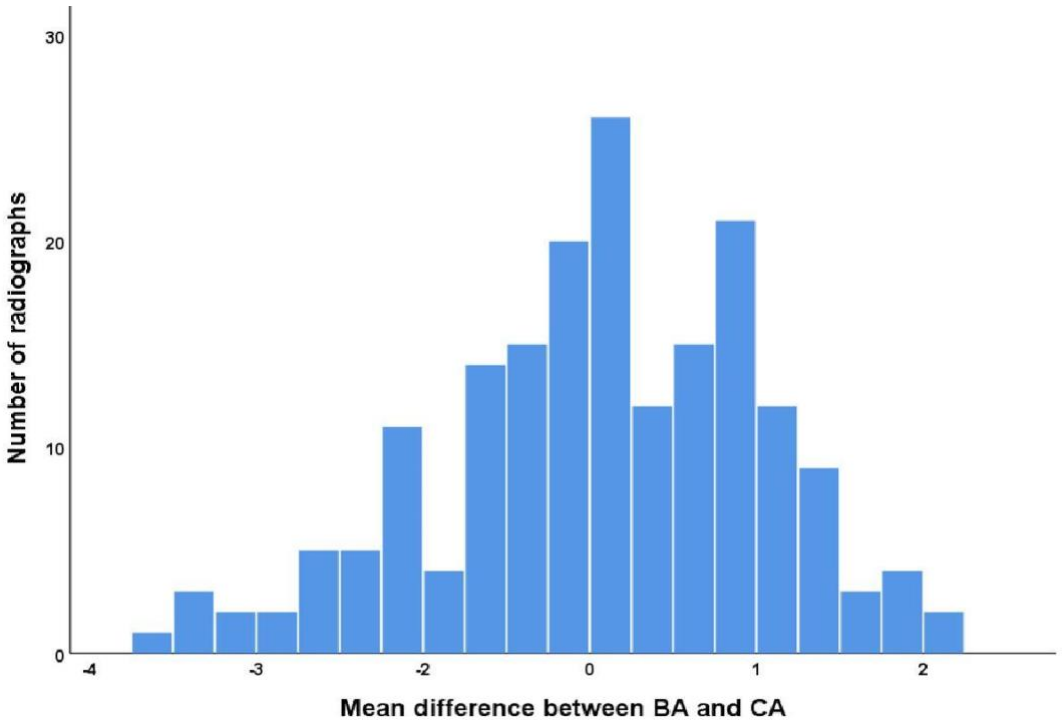


Figure 4.5: Mean difference between TW3 and CA (in years, females)

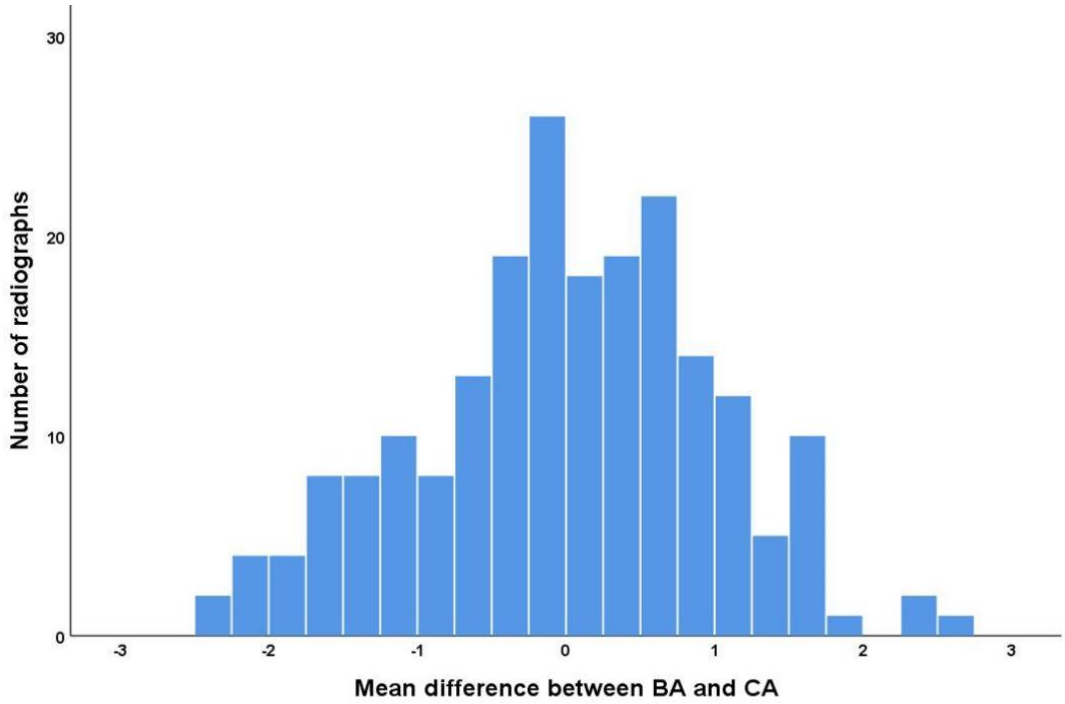


Figure 4.6: Mean difference between TW3 and CA (in years, males)

Subgroup analysis of the Caucasian data showed no statistical difference compared to the results from overall analysis, which included all ethnicities (Tables 4.1, 4.2, 4.3). An independent t test showed no significant difference between the mean difference of BA and CA when acquired from either the left hand or the right hand for TW3 methods (p=0.08 females, p=0.30 males). Mean difference between BA and CA according to body side are illustrated in Table 4.5. Additionally, the findings of this study- especially mean difference between BA and CA - are contrasted with previous studies that focused on the Caucasian population in Table 4.6 (following page)

Table 4.5: Mean difference between BA and CA (in years), according to body side (all ethnicities)

		Females		Males	
	n	Left hand 118	Right hand 68	Left hand 139	Right hand 67
G&P	Mean difference (SD)	0.03 (1.06)	-0.2 (1.02)	0.1 (1.08)	0.21 (1.03)
TW3	Mean difference (SD)	-0.32 (0.94)	-0.6 (0.99)	-0.04 (1.00)	0.09 (0.95)

**No significant difference was observer, p value ranged from (0.07 to 0.58)*

Table 4.6: Mean difference between BA and CA in studies that assessed the reliability of the G&P atlas in Caucasian Children

Study	Origin/ ethnicity	Age (years)	N	Mean (years)	BA-CA
		G&P			
Loder et al, 1993 [22]	White	0-18	M= 203 F= 177	M= -0.1 F= 0.07	
Ontell et al, 1996 [23]	White	3-18	M= 208 F= 130	M= -0.29 F= 0.14	
Buken et al, 2009 [9]	Turkish	11-16	M = 169 F = 164	M = -0.02 F = -0.65	
Zhang et al, 2009 [13]	White	0-18	M = 164 F = 163	M = 0.01 F = -0.15	
Calfee et al, 2010 [24]	Caucasian	12-18	M= 62 F= 76	M= 0.98 F= 0.66	
Santoro et al, 2012 [14]	Italian	7-15	M = 243 F = 261	M = -0.1 F = 0.40	
Suri et al, 2012 [19]	White	9-18	M = 311 F = 261	M = 0.50 F = 0.50	
Paxton et al, 2013 [15]	Australian	0-18	M = 276 F = 130	M = -0.12 F = -0.30	
Hackman & Black 2013 [20]	Scottish	1-20	M = 249 F = 157	M = -0.13 F = -0.16	
Mansourvar et al, 2014 [16]	White	10-16	M = 46	M = 0.04	
Gungor et al, 2015 [25]	Turkish	10-18	M = 259 F = 276	M = 0.64 F = -0.98	
Zabet et al, 2015 [21]	French	10-19	M = 100 F = 90	M = -0.19 F = -0.53	
Maggio et al, 2016 [17]	Western Australian	0-25	M = 180 F = 180	M = 0.24 F = -0.14	
		TW3			
Buken et al, 2009 [9]	Turkish	11-16	M = 169 F = 164	M = -0.18 F = -0.21	
Schmidt et al, 2008 [18]	Germany	1-18	M = 48 F = 40	M = 0.61 F = 0.23	

A positive value of the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age.

4.5 Discussion

Several variables may affect the applicability of BA methods. One is socioeconomic status, in which high socioeconomic status is more likely to accelerate skeletal maturation rate (Schmeling *et al.*, 2006). In light of improved standards of living in the 21st century, with potential increasing rate or changing pattern of skeletal maturation, the reliability of bone age estimation techniques has been debated. We sought to analyse the reliability of the G&P and TW3 methods within the UK context.

Breaking the cohort into yearly intervals showed statistical significance for varying age groups in females and males, when using the G&P atlas. These differences (overestimation at age of 6 and underestimation at age of 12, in females) were still significant when only data from Caucasian children was analysed. In spite of these sub-group differences, there was no statistical difference between overall mean BA and overall mean CA in either males or females. To convey a comprehensive picture, we contrasted our findings - especially mean difference between BA and CA - with previous studies that focused on the Caucasian population (Table 4.6). Some of these studies have concluded that Caucasian children mature skeletally at approximately the same rate as the G&P standard in males across all age groups (A. Zhang, James W Sayre, *et al.*, 2009; Büken *et al.*, 2009; Santoro *et al.*, 2012; Paxton, Lamont and Stillwell, 2013; Mansourvar *et al.*, 2014; Maggio *et al.*, 2016). However, other authors recommend that the G&P atlas be used with reservation due to mean BA being retarded in some age groups compared to the reference population (Schmidt *et al.*, 2008; Hackman and Black, 2013; Suri *et al.*, 2013; Zabet *et al.*, 2014a). Common findings among

these studies of the G&P atlas include underestimation of males aged below 13 years and overestimation during adolescence. G&P was applicable to females during adolescence while overestimation was reported before the age of 12 years. Others have recommended that a new standard altogether is required for precise bone age estimation, given the significant advancement of BA due to secular changes in skeletal maturation, which is thought to be due to improved standard of living. For example, Calfee et al reported that G&P overestimated males and females between 12 and 15 years old, for whom BA exceeded CA by at least 2 years (Calfee *et al.*, 2010). All of these studies used the subjective assessment of experienced raters; our results using an objective software programme indicate that overall, G&P currently remains applicable.

In contrast to the G&P atlas, we found that TW3 significantly underestimated females' ages after 3 years of age. The mean difference between BA and CA was statistically significant in females, especially at the ages of 8, 11, 12 (Figure 4.7–following page) and 15 years, for all ethnicities and for Caucasians alone. However, the TW3 did not show any statistical significant difference (under/overestimation) in the other age groups (at the age of 2, 3, 4, 5, 6, 7, 9, 10, 13 and 14 years). Therefore, it should be mentioned that less than half of the females age groups should a statistical different to the TW3. In Caucasian males, the mean BA was significantly lower than CA at age of 9,12 and 13 years. Furthermore, no statistical significant difference between TW3 and CA was observed on the other age group.

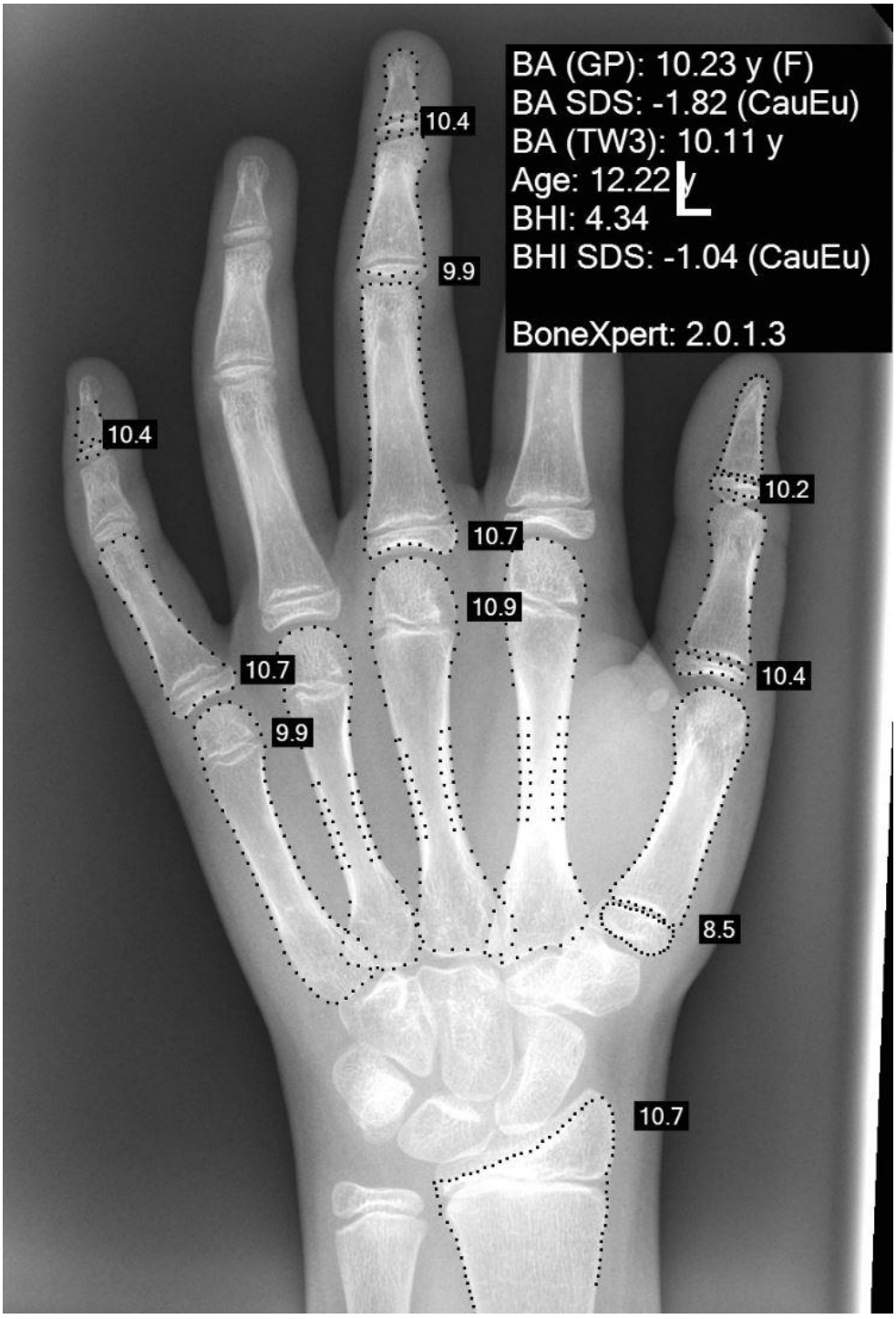


Figure 4.7: BoneXpert reading of the left-hand radiograph of a 12-year-old female. BA (GP); Greulich and Pyle bone age. SDS; standard deviation score. CauEu; Caucasian, European. TW3; Tanner and Whitehouse 3. BHI; bone health index.

A large number of children included in this study (55%) were of low socioeconomic status according to IMD. Socioeconomic status explained 17.8% of the difference between bone age (TW3 method) and chronological age. Although there have been improvements in standard of living over the past decade (Easterlin, 2000), (expected to advance bone age), our results show delayed BA in girls when using the TW3 method. In line with our results, other studies have shown delayed BA compared to CA in females after the age of 10 years (Schmidt *et al.*, 2008; Büken *et al.*, 2009; Pinchi *et al.*, 2014). These results potentially support recent views of some researchers, who argue that the improved secular trend has eased or stopped (Cole, 2000, 2003). As a result of an improving secular trend in standard of living, the TW3 method was established in 2001 such that the TW3 BA is about a year ahead of the previous (TW2) method, especially after the age of 10 or 11 years (Tanner *et al.*, 2001). Our results suggest that a return to TW2 may be necessary.

Several authors argue that socioeconomic status is the predominant reason behind the difference in skeletal maturational rates among populations (Ashizawa *et al.*, 2005; Schmeling *et al.*, 2006). Schmeling et al found that bone age was retarded among 27 studies that reported the socioeconomic status of their participants (Schmeling *et al.*, 2006). This retardation was due to the high socioeconomic status of the children recruited to develop the G&P atlas compared to the children within these studies, such that even the secular trend of increasing standard of living was not sufficient to eliminate any differences in socioeconomic status of the various cohorts. Conversely, Schmidt et al attributed an overestimation of 0.2 years in males aged 12 to

15, to their relatively high socioeconomic status which lead to acceleration of skeletal maturation (Schmidt *et al.*, 2007a).

In spite of the likely effects of socioeconomic status, the impact of ethnicity cannot be neglected. Studies on two different ethnic groups residing in the same region have shown that bone age assessment methods may reveal different results (A. Zhang, James W Sayre, *et al.*, 2009; Gungor *et al.*, 2015b). Ontell et al showed that the G&P atlas is applicable to Caucasian girls at all ages but not to boys before the age of 13, while in Asians in the same region, the G&P atlas is applicable to girls at all ages but only to boys between 7 and 13.3 years. Zhang et al concluded that Asian children mature sooner than do Caucasian children, especially between 10 and 13 years of age in girls and between 11 and 15 years of age in boys. It has been shown that young Asian adults reach the end of maturity prior to the age observed through the TW3 method (Cole, 2003; Pinchi *et al.*, 2014). Research focusing on South African individuals, found that TW3 underestimated bone age for boys but not girls (Cole *et al.*, 2015). We demonstrated no significant difference between all ethnic groups compared to Caucasians alone (the latter formed 76% of the study population).

Measuring BA according to a subjective technique has a greater likelihood of introducing rating variations across analysts, due to varying degrees of expertise. However, this disadvantage has been overcome through the introduction of BoneXpert which is an automated bone age analysis software tool that in addition to elimination observer variability, has the advantage of saving significant time. Our observed 5-month persistent discrepancy between chronological age and TW3 bone age as measured by BoneXpert

in females appears to be a disadvantage not of the software, but of the reference standard (TW3) on which the software depends.

The limitations of this study include: 1) the fact that we did not review hospital notes to ascertain full health in the children (although radiology and ED notes were scrutinised). 2) The exclusion of certain age groups, namely those under 2 years old in females, under 2.5 years in males and individuals of both sexes aged 15 years or older. In order to save time and eliminate subjectivity, this pragmatic study was performed using BoneXpert; however, this software tool is unable to read images from younger age groups due to limited ossification or non-ossification of epiphyses, while its dependability is questionable when used on older age groups (Thodberg *et al.*, 2017).

3) Height and weight of recruited children was not recorded; it may be that body mass index affects the rate of skeletal maturation and the prevalence of overweight and obese children is well documented to be rising (Ng *et al.*, 2014). 4) We do not know the precise socioeconomic status of the reference children. Progress in medicine, education, industry and economic growth have all contributed to higher socioeconomic status which in turn is expected to have had a positive impact on children's skeletal maturation. Our results, showing retardation of BA appear counterintuitive, but may not be if the socioeconomic status of the TW3 reference children was on average higher than that of the children we recruited.

4.6 Conclusion

Our results indicate that 1) No significant difference exists between left and right hand BoneXpert-derived BA 2) No significant difference exists between BoneXpert-derived BA and CA when using the G&P atlas, therefore, this method can be utilised for the modern population in the UK 3) TW3 consistently underestimates the age of females by an average of 5 months, which should be considered by users during application of the method.

CHAPTER 5

Applicability of Two Bone Age Assessment Methods to Children from Saudi Arabia

Alshamrani K[1], Hewitt A[2], Offiah A[1,2]

[1] Department of Oncology & Metabolism, University of Sheffield, UK

[2] Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield, UK

Correspondence: Khalaf Alshamrani, Academic Unit of Child Health, Damer Street, Sheffield, S10 2TH

Kalshamrani1@sheffield.ac.uk

Under Review (Clinical Radiology)

5.1 Abstract

Background: The Greulich & Pyle (G&P) and Tanner & Whitehouse (TW) methods are frequently used to determine bone age. The question to be raised is, “Are these standards applicable to children of different ethnicity to those on which they are based?”

Methods: Bone age was assessed using the G&P and TW3 methods, firstly by independent manual rating of 2 observers, followed by a single observer using the BoneXpert software programme. In total, 420 hand trauma radiographs for Saudi Arabians (220 males, 329 left, age range 1 to 18 years) performed in the period January 2012 – September 2016 were assessed. Paired sample *t* test was used to compare the difference between mean bone age (BA) and mean chronological age (CA) and to compare the difference between manual and BoneXpert ratings. Statistical analysis was undertaken using SPSS v.25.

Results: We found a statistically significant difference between BA and CA in males when using G&P (mean difference -0.36 ± 1 years, $p < 0.01$) and TW3 (mean difference -0.22 ±0.9 years, $p=0.03$) methods but not in females for either G&P (mean difference 0.13 ± 1.2 years) or TW3 (mean difference 0.08 ± 1.1 years). In males, BoneXpert results conformed to the manual ratings for TW3 but not for G&P, for which the mean difference between manual and BoneXpert ratings was -0.27 ±0.5 years ($p<0.01$).

Conclusions: Our results indicate that manual and BoneXpert-derived G&P and TW3 bone age assessment can be applied with no modification to Saudi Arabian females. However, only TW3 BoneXpert-derived BA can be applied without caution to Saudi Arabian males.

5.2 Introduction

The determination of bone age is a routine diagnostic procedure usually required to identify growth disorders in children and plan for therapeutic procedures. It is important to assess bone age using a reliable method, one of which is the assessment of bone age from a left hand radiograph (A. Schmeling *et al.*, 2008). Two approaches are widely used to assess bone age from a left-hand radiograph, namely the Greulich and Pyle (G&P) and the Tanner and Whitehouse (TW3) methods (Greulich and Pyle, 1959; Tanner *et al.*, 2001). The data that were used to establish the G&P atlas and the TW3 standard came from healthy children of North American and western European origin and was collected around 4 and 9 decades ago. In addition to potential secular change, ethnicity and socioeconomic status are factors that have an impact on children's bone age. Therefore, one question to be raised when using these standards is, "Are they relevant to a current population of different ethnicity and/or socioeconomic status to the children used to develop the standards?"

The G&P and TW3 methods were initially (and still most commonly) based on a subjective approach that is likely to suffer from variations in rating between assessors due to different levels of competence, with their reliability partially dependent on the skill of the assessor. To eliminate observer variation and reduce rating time, BoneXpert software was introduced in 2009. This is an automated software programme that calculates bone age according to the G&P and TW3 methods (Thodberg *et al.*, 2009). However, although the software has been validated in Caucasian (van Rijn, Lequin and Thodberg, 2009; Thodberg and Sävendahl, 2010), African-American

(Thodberg and Sävendahl, 2010), Hispanic and Asian-Chinese (Thodberg and Sävendahl, 2010; Kim, Lee and Yu, 2015), studies on other indigenous populations are limited. Therefore, this study will assess the applicability of the G&P and TW3 to children from Saudi Arabia using both subjective (manual) rating and BoneXpert software.

5.3 Material and Methods

After the ethical approval was determined, hand radiographs performed on children aged between 1 and 18 years old presenting to the Emergency Department of King Fahad Hospital, Saudi Arabia, between January 1st, 2012 and September 30th 2016 following trauma were retrospectively identified from the Picture Archiving and Communication System. All radiographs were acquired via a computerised radiography system and were in DICOM format. Studies with a specific request for BA estimation were excluded. Emergency Department notes were scrutinised and any child with an underlying disorder was excluded. Demographic data including sex and age at the time of the radiograph were recorded. Only radiographs of Saudi Arabians were included and were confirmed using the national ID included within the health ID (Khan, 2010; Ministry of Health, 2015).

5.3.1 Manual rating

Observers 1 and 2 independently assessed bone age from all radiographs without knowledge of chronological age using the G&P method. When the patient's bone age was assessed to lie between two adjacent standards, the intermediate value was assigned as the bone age. Observers 1 and 3 assessed the radiographs using the RUS (radius, ulna and short bone) method. The time interval between Observer 1's G&P and TW3 reads was at

least three months. To determine intra-observer reliability, a random sample of 43 radiographs (22 males) were assessed by each observer 1 month following their initial reads.

The maximum potential TW3 bone age score is 1000, which corresponds to an adult standard, while the minimum potential score is 42, which corresponds to 2 years of age. In this study, radiographs that were assigned as adult or did not achieve the minimum score were excluded. Additionally, for both G&P and TW3 reads, radiographs were excluded when bone age could not be assigned as a result of poor positioning or artefact.

5.3.2 BoneXpert rating

All radiographs were exported into an external hard drive and a standalone version of BoneXpert (Visiana, Holte, Denmark, v2.5.1.1) was used to determine bone age (G&P and TW3). Age was limited to 15 years in females and 17 years in males because the software does not provide a precise G&P reading above these ages. The default ethnicity for analysing the radiographs was Caucasian, as the software does not include ethnicity-specific standard deviation scores (SDS).

5.3.3 Statistical analysis

Statistical analysis was undertaken using SPSS version 24 for PC (IBM, Armonk, New York). Inter-observer reliability was assessed using interclass correlation coefficient. The mean variation for BA and CA was determined for each child by subtracting BA from CA (BA-CA). Paired sample *t* test was used to test the significance of the differences between BA and CA for each method and to test the significance of the differences between manual and BoneXpert ratings for each method.

5.4 Results

5.4.1 G&P atlas

Concerning manual G&P ratings, 420 radiographs (220 males) were assessed by each observer. The inter-class correlation coefficient (ICC) showed a high correlation between the two observers with coefficients of 0.984 for females and 0.991 for males. No significant intra-observer difference was identified (p=0.772). In this regard, readings from the first observer were used when comparing the BA to CA using the G&P atlas. BA was lower than CA in 48% of females and 61% of males, while being equal in 1% of males. The mean difference between BA and CA ranged from 37 months underestimation to 36 months overestimation in both females and males. On average, G&P underestimated males by 0.31 years/4 months ($p < 0.01$) and overestimated females by 0.1 years/1 month ($p = 0.089$) (Table 5.1).

Table 5.1: Mean difference (±SD) in years, between BA and CA in females and males

Manual rating	**Sex**	**No**	**Mean CA (±SD)**	**Mean BA (±SD)**	**Mean difference BA-CA**	**p value**
G&P BA vs CA	Female	200	10.21 (± 4.4)	10.34 (±4.8)	0.13 (± 1.2)	0.089
	Male	220	10.48 (± 4.8)	10.12 (±5.2)	-0.36 (± 1.0)	<0.01
TW3 BA vs CA	Female	164	8.80 (±3.6)	8.88 (± 3.8)	0.08 (± 1.1)	0.413
	Male	189	9.59 (± 4.4)	9.37 (±4.7)	-0.22 (± 0.9)	0.03
BoneXpert rating	**Sex**	**No**	**Mean CA (±SD)**	**Mean BA (±SD)**	**Mean difference BA-CA**	**p value**
G&P BA vs CA	Female	98	9.02 (± 3.7)	9.18 (± 4.0)	0.16 (± 1.0)	0.06
	Male	114	9.89 (± 3.9)	9.68 (± 4.0)	-0.21 (± 0.8)	0.03
TW3 BA vs CA	Female	96	8.45 (±3.38)	8.58 (±3.6)	0.13 (± .9)	0.22
	Male	111	9.85 (± 3.9)	9.73 (± 3.9)	-0.12 (± 0.9)	0.09

A positive value for the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age.

With the cohort divided into yearly intervals, G&P overestimated females aged from 1 to 5 years by between 0.5 and 6 months, apart from at 3 years of age. After 5 years of age, G&P consistently underestimated females by between 3 and 8 months until 9 years of age, with underestimation being statistically significant ($p < 0.05$) at 6 years of age (Table 5.2).

Table 5.2: Mean difference (±SD) in years, between G&P BA (manual and BoneXpert) and CA in females

		Manual Rating			BoneXpert Rating			
Age (years)	**No**	**Mean difference**	**(±SD)**	**p value**	**No**	**Mean difference**	**(±SD)**	**p value**
1	4	0.04	0.43	0.86		-	-	-
2	4	0.48	0.65	0.24	3	0.68	0.46	0.16
3	9	-0.41	0.87	0.20	5	0.12	0.44	0.25
4	11	0.03	0.75	0.89	6	0.38	0.62	0.43
5	12	0.21	0.71	0.11	8	0.42	0.79	0.20
6	13	-0.68	1.02	0.03	7	0.32	1.19	0.21
7	14	-0.25	1.10	0.47	6	-0.02	0.96	0.91
8	14	-0.36	0.95	0.18	9	-0.38	0.82	0.08
9	17	-0.41	1.40	0.23	11	-0.29	1.47	0.52
10	13	0.22	1.63	0.65	5	0.47	0.92	0.38
11	15	0.71	1.48	0.08	10	0.35	1.02	0.36
12	14	1.10	1.20	0.00	9	0.89	1.26	0.08
13	16	0.83	1.47	0.04	6	0.98	1.16	0.03
14	11	0.46	1.37	0.29	6	0.41	1.24	0.37
15	12	0.56	1.50	0.22	7	0.02	1.01	0.96
16	8	0.18	1.32	0.72	-	-	-	-
17	8	0.01	0.73	0.97	-	-	-	-
18	5	-0.12	0.34	0.13	-	-	-	-

A positive value for the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age

The G&P atlas then overestimated females by between 1 and 13 months with overestimation being statistically significant (p <0.05) at 12 and 13 years of age. G&P underestimated males from 1 to 13 years by between 2 and 13 months, apart from at 4 years. This underestimation was statistically significant (p<0.05) at the ages of 7, 8, 9 and 10 years (Table 5.3). After the age of 13 years, G&P overestimated males, but this did not reach statistical significance.

Table 5.3: Mean difference (±SD) in years, between G&P BA (manual and BoneXpert) and CA in males

		Manual Rating				BoneXpert Rating		
Age (years)	**No**	**Mean**	**(±SD)**	**p value**	**No**	**Mean**	**(±SD)**	**p value**
1	5	-0.30	0.66	0.37		-	-	-
2	7	-0.20	0.63	0.40	3	0.29	0.59	0.61
3	14	-0.26	0.85	0.28	7	0.04	0.62	0.83
4	11	0.33	0.53	0.07	6	0.41	0.58	0.14
5	13	-0.35	0.59	0.06	8	0.25	0.58	0.24
6	10	-0.21	0.65	0.39	6	0.11	0.63	0.69
7	15	-0.72	1.00	0.01	10	-0.31	0.88	0.18
8	12	-1.12	1.20	0.01	8	-0.97	1.06	0.01
9	14	-1.03	1.09	<0.00	9	-0.97	1.13	<0.01
10	12	-0.84	1.16	0.02	6	-0.72	1.07	0.09
11	15	-0.43	0.92	0.08	7	-0.17	1.03	0.48
12	14	-0.57	1.05	0.11	8	-0.36	0.91	0.30
13	13	-0.38	0.98	0.13	8	0.07	1.11	0.72
14	12	0.33	1.28	0.44	6	0.26	1.05	0.48
15	16	0.51	1.08	0.11	12	0.17	1.16	0.53
16	15	0.56	1.13	0.10	7	0.40	0.71	0.04
17	13	0.22	0.85	0.35	3	-0.24	0.64	0.34
18	9	0.07	0.77	0.78	-	-	-	-

A positive value for the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age.

In regard to BoneXpert-derived G&P BA, the software was not able to analyse 208 (50%) of the radiographs, thus only 212 radiographs (114 males) were included in the final analysis. BoneXpert overestimated G&P BA in females by 2 months ($p = 0.06$) and underestimated G&P BA in males by 2.5 months ($p < 0.05$). Mean difference between BA and CA ranged from 32 months underestimation to 30 months overestimation in both females and males.

With the cohort divided into yearly intervals, G&P BA derived by BoneXpert followed a similar pattern of under/overestimation as the manual rating in females, however, no statistical significance was found, apart from at the age of 13 where the software significantly overestimated females ($p<0.05$) (Table 5.2). In males, in contrast to manual rating BoneXpert overestimated males aged between 2 and 6 years by between 1 and 4 months.

BoneXpert underestimated G&P BA in males aged between 7 and 12 years, with underestimation being statistically significant ($p<0.01$) at ages 8 and 9 years (Table 5.3). The G&P manual rating was lower than BoneXpert derived G&P by an average of 0.27 years/3 months in males ($p < 0.01$) and 0.1 years/1 month ($p = 0.184$) in females. Bland Altman plots comparing manual and BoneXpert ratings in females and males using G&P are illustrated in Figures 5.1 and 5.2 (following page)

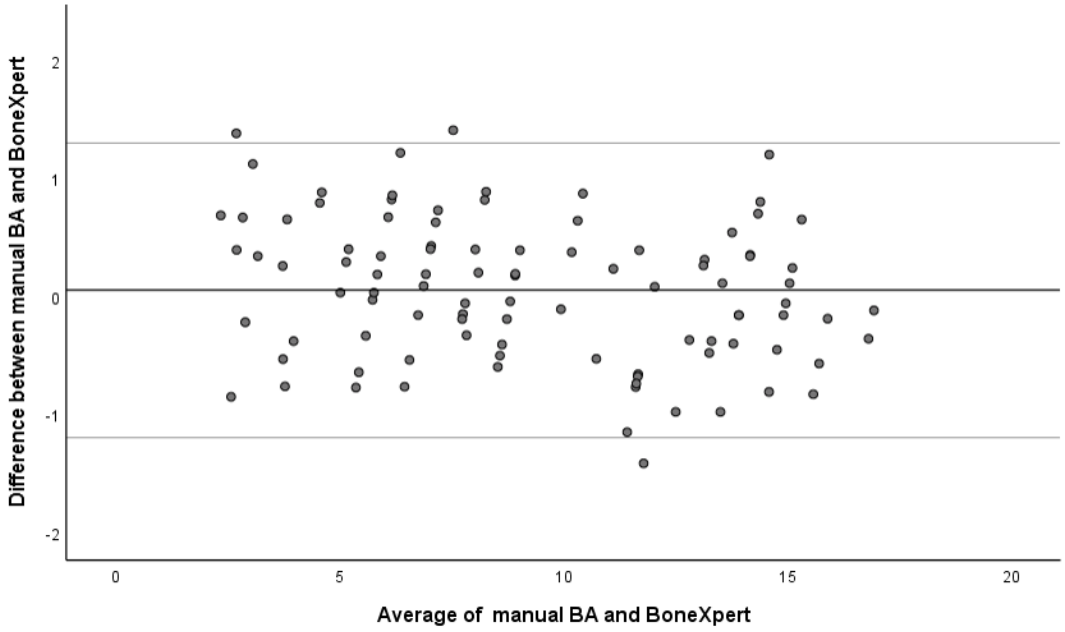


Figure 5.1: Bland Altman plot comparing manual and BoneXpert ratings in females using the G&P method with no systematic bias

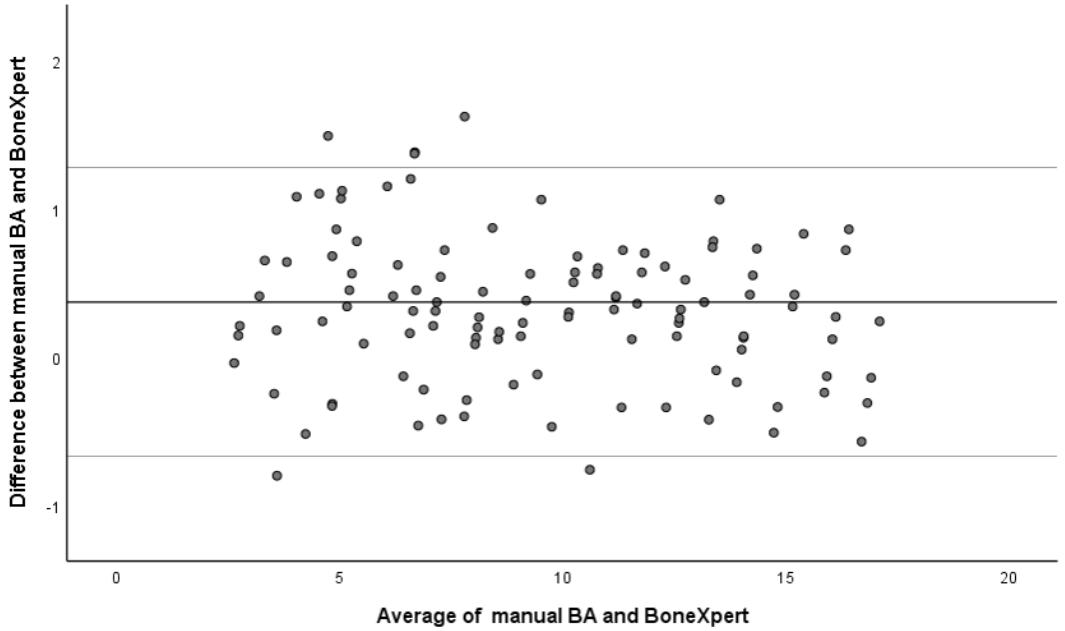


Figure 5.2: Bland Altman plot comparing manual and BoneXpert ratings in females using the G&P method with no systematic bias

5.4.2 TW3 method

Concerning manual TW3 ratings, 67 radiographs were excluded from analysis for the following reasons; (a) 43 radiographs achieved the maximum score (26 females), (b) 14 radiographs did not reach the minimum score (6 females), (c) 11 radiographs were poorly positioned, such that bone age could not be determined. In total, 353 radiographs were included in the final analysis (Tables 5.1). The intra-class correlation coefficient indicated a high correlation between the two observers (0.97 for females and 0.96 for males). As there is no significant intra-observer difference (p=0.351), readings from the first observer was used when comparing BA to CA.

BA was lower than CA in 44% of females and 56% of males, while being equal in 1% of females. The mean difference between BA and CA ranged from 30 months underestimation to 28 months overestimation in both females and males. On average, TW3 underestimated males by 0.22 years/2.5 months ($p < 0.01$) and overestimated females by 0.1 years/1 month ($p = 0.413$) (Table 5.1).

With the cohort divided into yearly intervals, TW3 overestimated females aged from 1 to 13 years by between 0.5 and 7 months, apart from at 6,7 and 8 years, with overestimation being statistically significant ($p <0.05$) at 11 and 12 years of age (Table 5.4 – following page).

In contrast, TW3 underestimated males aged 5 to 11 years, with underestimation being statistically significant ($p <0.05$) at 8 and 9 years. After the age of 11 years, TW3 overestimated males by between 1 to 6 months, with overestimation being statistically significant ($p <0.05$) at 13 years (Table 5.5–page 117).

Table 5.4: Mean difference (±SD) in years, between TW3 BA (manual and BoneXpert) and CA in females

	Manual Rating				BoneXpert Rating			
Age (years)	**No**	**Mean**	**(±SD)**	**p value**	**No**	**Mean**	**(±SD)**	**p value**
2	4	0.66	0.32	0.03	2	0.21	0.21	0.04
3	9	0.28	0.48	0.12	5	0.19	0.34	0.20
4	11	0.35	0.66	0.11	6	0.30	0.78	0.44
5	12	0.08	0.51	0.59	8	-0.19	0.64	0.53
6	13	-0.35	0.73	0.08	7	-0.12	0.88	0.70
7	12	-0.21	0.75	0.37	6	-0.15	0.98	0.73
8	14	-0.26	0.90	0.31	9	-0.63	0.76	0.04
9	15	0.14	1.11	0.60	11	-0.27	1.16	0.45
10	13	0.22	1.27	0.56	5	0.82	1.02	0.06
11	15	0.59	0.87	0.02	10	0.53	1.18	0.24
12	14	0.68	0.97	0.00	9	0.81	0.96	0.05
13	14	0.16	1.16	0.09	6	0.80	0.91	0.03
14	11	-0.07	0.38	0.08	6	0.28	0.71	0.12
15	7	-0.53	0.34	0.02	6	-0.1	0.35	0.15

A positive value for the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age.

Table 5.5: Mean difference (±SD) in years, between TW3 BA (manual and BoneXpert) and CA in males

	Manual Rating				BoneXpert			
Age (years)	**No**	**Mean**	**(±SD)**	**p value**	**No**	**Mean**	**(±SD)**	**p value**
2	9	0.44	0.68	0.14	3	0.82	0.31	0.17
3	11	0.05	0.47	0.72	7	0.48	0.48	0.04
4	12	0.02	0.56	0.89	6	0.62	0.80	0.12
5	10	-0.11	0.50	0.43	8	0.16	0.46	0.32
6	13	-0.33	0.46	0.09	6	0.04	0.50	0.87
7	12	-0.23	0.72	0.22	10	-0.26	0.63	0.23
8	14	-0.84	1.00	0.01	8	-0.44	0.88	0.20
9	12	-0.58	0.92	0.03	9	-0.68	0.78	0.03
10	15	-0.43	0.96	0.13	6	-0.59	0.73	0.08
11	14	-0.17	1.13	0.58	7	-0.21	0.90	0.46
12	13	0.06	1.04	0.84	8	-0.27	1.36	0.59
13	12	0.58	1.09	0.05	8	0.47	1.30	0.25
14	16	0.46	1.11	0.23	6	0.73	1.08	0.16
15	14	0.22	0.68	0.22	12	0.12	0.65	0.59
16	9	-0.16	0.36	0.03	7	-0.21	0.19	0.07
17	3	-0.85	0.25	0.00	-	-	-	-

A positive value for the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age.

Concerning BoneXpert, 5 additional radiographs (2 females) were excluded as the radiographs achieved the maximum score according to the BoneXpert-derived TW3 BA. BoneXpert overestimated TW3 BA in females by an average of 1 month, while underestimating males by 2 months. Mean difference between BA and CA ranged from 28 months underestimation to 30 months overestimation in both males and females.

Breaking the cohort into yearly intervals showed that similar to manual ratings, the software overestimated TW3 BA in females aged between 10 and 13 years, being statistically significant at age of 8 years (Table 5.4). In males, BoneXpert underestimated TW3 BA in males aged between 7 and 12 years, being statistically significant at the age of 9 years (Table 5.5). Mean BA using the manual TW3 method was lower than TW3 derived by BoneXpert by 1 month, with no significant difference between the two methods in both males and females. BoneXpert and manually-derived TW3 are compared as Bland Altman plots in Figures 5.3 and 5.4 (following page). Furthermore, the findings of this study- especially mean difference between BA and CA - are contrasted with previous studies that focused on the Asians population in table 5.6 (page 120).

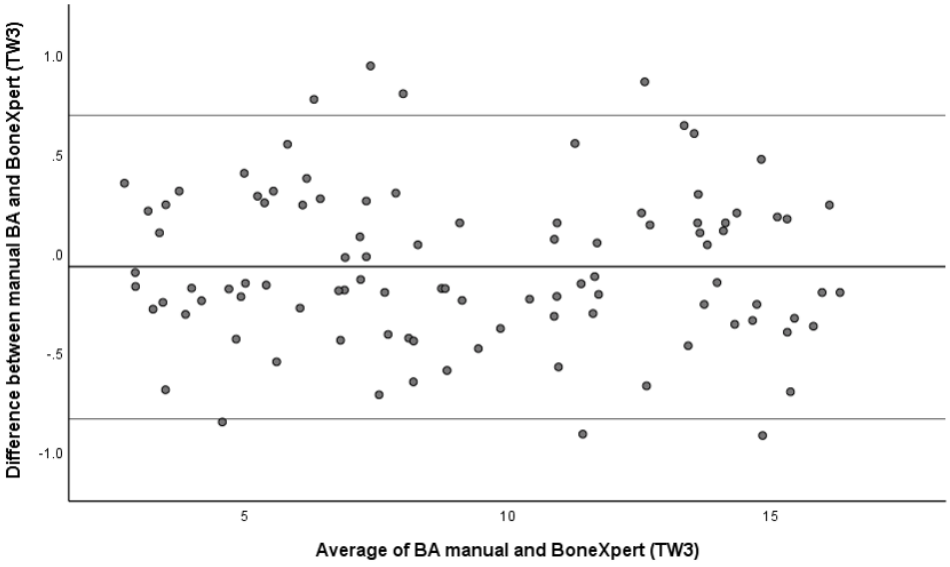


Figure 5.3: Bland Altman plot comparing manual and BoneXpert ratings in females using the TW3 method with no systematic bias.

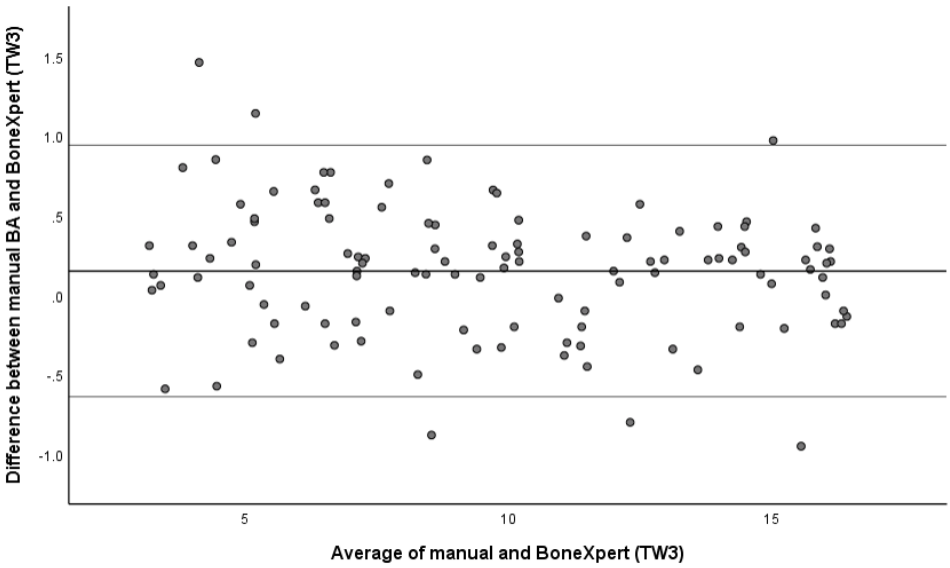


Figure 5. 4: Bland Altman plot comparing manual and BoneXpert ratings in males using the TW3 method with no systematic bias.

Table 5.6: Mean difference between BA and CA in studies that assessed the reliability of the G&P atlas in Asian children

Study	Origin/ ethnicity	Age (years)	N	Mean BA-CA (years)
		G&P		
So & Yen 1990	Chinese	11.9-12.3	F=117	F= 0.6
So & Yen 1991	Chinese	11.9-12.3	F=117	F= 0.6
Ontell et al, 1996	Asian	1-18	M=63 F=30	M= -0.03 F= 0.27
Krailassiri et al, 2002	Thai	7-19	M=139 F=222	M= -0.8 F= 0.8
Chiang et al, 2005	Taiwan	7-19	M=230 F=140	M= 0.82 F= -0.3
Al-Hadlaq et al, 2007	Saudi Arabian	7-15	M=115	M= -0.71
Griffith et al, 2007	Chinese	0-18	M=650 F=366	M= 0.25 F= 0.15
Zhang et al, 2009	Asian	0-18	M=165 F=166	M= 0.41 F= 0.24
Zafar et al, 2010	Pakistan	0-18	M=535 F=354	M= 0.1 F=- 0.19
Moradi et al, 2012	Iran	6-18	M=303 F=122	M= 0.37 F=- 0.04
Soudack et al, 2012	Iseral	0-18	M=375 F=304	M= 0.16 F=-0.04
Patil et al, 2012	India	1-19	M=194 F=181	M= 0.69 F= 0.64
Awais et al, 2014	Pakistani	0-18	M=136 F=147	M= -1.3 F= 0.06
Mansourvar et al, 2014	Asian American	1-8	M=48	M= 0.87
Mughal et al, 2014	Pakistan	4.5-9.5	M=139 F=81	M= -1.3 F= 0.55
Rai et al, 2014	India	5-15	M=75 F=75	M= -0.07 F= -0.33

Kim et al, 2015	Korean	7-12	M=135 F=77	M= -0.48 F= -0.02
Mohammed et al, 2015	South India	9-20	M=330 F=330	M= -0.23 F= 0.02
Benjvongkulchia et al 2018	Thai	8-20	M=172 F=193	M= 0.42 F= 0.90
		TW3 method		
Ashizawa et al, 2005	Beijing	6-16	M=631 F=642	M= 0.07 F= 0.11
Griffith et al, 2007	Hong Kong	0-18	M=645 F=329	M= 0.22 F= 0.3
Kim et al, 2015	Korean	7-12	M=135 F=77	M= 0.41 F= 0.12
Benjvongkulchia et al 2018	Thai	8-20	M=172 F=193	M= -0.12 F= 0.40

A positive value for the mean difference between BA and CA indicates advanced while a negative value indicates delayed bone age compared to chronological age, M = males, F = females, NR = not reported

5.5 Discussion

Using a reliable method to determine bone age is crucial for clinical and legal purposes. Hence, we sought to analyse the applicability of G&P and TW3 bone age standards to Saudi Arabian children, who are of different ethnicity to the population used to generate these two standards. We also sought to compare manual rating to BoneXpert, which software programme has not previously been used in the Saudi Arabian ethnic group.

In relation to G&P, underestimation by an average of 4 months and 2.5 months was observed in males using manual rating and BoneXpert, respectively. In females, both manual rating and BoneXpert, overestimated their age by 1 month and 2 months respectively. These findings are in line with the study by Alhadlaq et al. who found that the bone age of children

from Saudi Arabia aged 9 to 15 tended to be lower than chronological age by 8 months (Al-Hadlaq *et al.*, 2007). In other Asian populations, a large number of studies have shown that the G&P atlas is not applicable due to the large differences between bone age and chronological age (Al-Hadlaq *et al.*, 2007; Moradi, Sirous and Morovatti, 2012; Soudack *et al.*, 2012; Awais *et al.*, 2014; Mansourvar *et al.*, 2014; Mughal, Hassan and Ahmed, 2014; Rai, 2014; Kim, Lee and Yu, 2015). Generally, the G&P atlas seems to underestimate boys within these studies during early and mid-childhood and overestimate boys during adolescence.

Similar to the G&P atlas, the TW3 method underestimated females and males in younger age groups, and overestimated females and males after the age of 9 and 12 years, respectively. Although, there was no significant different between BA and CA when using the TW3 method in females, the TW3 underestimated BA in males by an average of 2.5 months. These finding were also recently observed in the Thai population (Benjavongkulchai and Pittayapat, 2018). Other studies on Asians showed that young adults are reaching the end of maturity prior to the age observed through the TW3 method (Griffith, Cheng and Wong, 2007; Kim, Lee and Yu, 2015). The mean difference between BA and CA observed in similar research focused on Asian populations is summarised in Table 5.6.

One of the main factors that has an impact on skeletal maturation rate is ethnicity (Mora *et al.*, 2001; A. Zhang, James W. Sayre, *et al.*, 2009; Zafar *et al.*, 2010; Cole *et al.*, 2015). This impact has been shown by studies that sought to test the applicability of the methods on two different ethnic groups

residing in the same region (Loder, 1993; Ontell *et al.*, 1996; A. Zhang, James W. Sayre, *et al.*, 2009). One of these studies showed that the G&P atlas was only applicable to Asian children between 7 and 13.5 years (Ontell *et al.*, 1996). Additionally, it seems that Asian children mature sooner than Caucasian children, especially between the age of 10-13 years, and 11-15 years in girls and boys, respectively (A. Zhang, James W. Sayre, *et al.*, 2009).

Socioeconomic status is another factor that may affect skeletal maturation. Bone age is usually delayed in children of low and advanced in those of high socioeconomic status (Schmeling *et al.*, 2006). Some authors suggest that the inapplicability of the bone age standards is more likely to be due to differences in socioeconomic status than ethnicity. For example, Asians-Japanese children living in Japan were skeletally delayed between the age of 5 and 18 years in comparison to the Caucasian children who lived in Cleveland (US) at all age groups (Greulich-Pyle, 1957). However, Greulich argued that this was not due to ethnicity, but due to less favourable environmental conditions, which can be interpreted as low socioeconomic status.

Although BoneXpert agreed with the manual rating in the overall over/underestimation pattern, there was a statistically significant difference between the two methods in males but not in females. This may be due to the method by which BoneXpert calculates G&P bone age; the software does not include the carpal bones in its assessment. In our study, male radiographs in the younger age groups appeared to show less maturity in the carpal compared to the other bones of the hand (Figure 5.5). This has also

been highlighted in other populations, in which carpal maturation pattern has influenced bone age assessment results (Acheson, Vicinus and Fowler, 1966; Krailassiri, Anuwongnukroh and Dechkunakorn, 2002; Al-Hadlaq *et al.*, 2007). However, the value of the carpal bones in bone age assessment has been questioned due to the poor correlation between carpal bone development and chronological age. Johnston and Jahina concluded that the accuracy of bone age assessment increased when the carpal bones were illuminated (Johnston and Jahina, 1965). Therefore, the BoneXpert-derived BA results in the current study are more reliable than the manual results for which all hand and carpal bones were assessed.

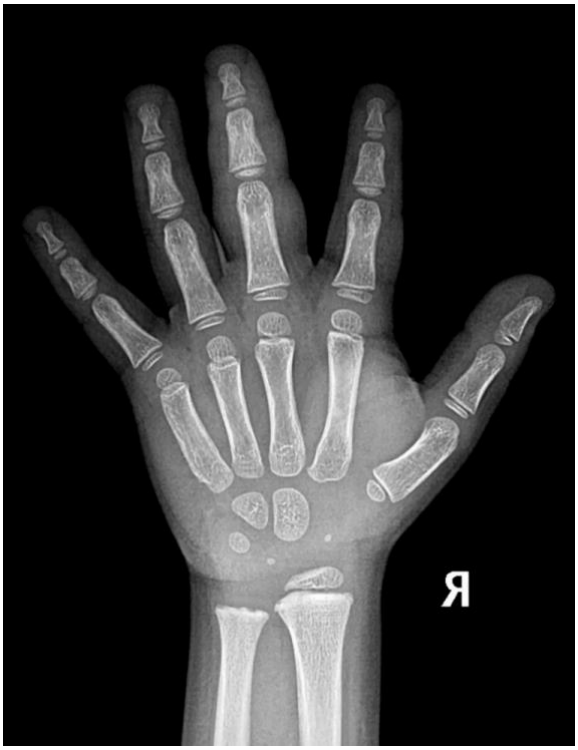


Figure 5.5: DP L hand radiograph of a male, chronological age 5 years and 7 months, showing less maturity in the carpal area compared to the other bones of the hand

BoneXpert could not assess approximately half of all radiographs, mainly because the images were post-processed using a sharpening algorithm, which gave them excessively sharp borders, rendering them unreadable by the software. The relatively small number of radiographs included in each age group for Bonexpert analysis compared to manual rating, may have contributed to the differences between BoneXpert and manually-derived BA.

The limitations of this study include 1) socioeconomic status was not reported due to insufficient information; 2) hospital notes were not reviewed to ascertain full health in the children (although radiology and ED notes were scrutinised) 3) both left and right hand radiographs were used; traditionally BA has been assessed from left hand radiographs, however, it has been shown that there is no significant difference in G&P or TW3 BA between left and right hands (Thodberg *et al.*, 2010) and so this should not have affected our results and 4) only certain age groups were included in BoneXpert analysis, namely between 2 and 15 years old in females and between 2.5 and 17 years in males. This was unavoidable because the software tool is unable to read images from younger age groups due to limited ossification or non-ossification of epiphyses, while its dependability is questionable when used in older age groups. Having said that, due to recruitment method (children attending an Emergency Department with hand trauma) and the high rejection rate of the software within as a result of insufficient image quality, some of the age groups included in the BoneXpert analysis had fewer than 5 radiographs (Tables 5.2 to 5.5), and the results of this study in these age groups should be treated with caution.

5.6 Conclusion

Our results indicate that the G&P and TW3 manual and BoneXpert methods can be applied to Saudi Arabian females. However, significant differences between BA and CA were apparent in Saudi Arabian males for manual and BoneXpert-derived G&P and TW3 BA but not for BoneXpert-TW3 BA.

CHAPTER 6

Bone Age Determination using Dual-Energy X-ray Absorptiometry

Alshamrani K[1], Elzene K [2], Offiah A [1,2]

[1] Department of Oncology & Metabolism, University of Sheffield, UK

[2] Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield, UK

Correspondence: Khalaf Alshamrani, Academic Unit of Child Health, Damer street Building, Sheffield, S10 2TH

Kalshamrani1@sheffield.ac.uk

Under review (Skeletal Radiology)

6.1 Abstract

Objective: To assess whether hand-wrist dual-energy x-ray absorptiometry (DXA) can replace radiographs for bone age assessment using the Greulich & Pyle (G&P) and/or Tanner & Whitehouse (TW3) methods.

Methods: Purposive sampling was used to include a total of 20 patients identified from an Endocrine Clinic; two males and two females from each of 5 age groups (<5; 5 to 7; 8 to 10; 11 to 13; 14 to 16 years). Bone age as determined from DXA and radiographs performed on the same day were compared for each child. Two observers independently assessed all radiographs and DXA scans on two occasions. For each observer, there was a minimum interval of two weeks between the two reads. Interclass correlation coefficient and Bland Altman plots were used to evaluate agreement between the observers and correlation between the two imaging modalities.

Results: The mean chronological age was 9.04 (SD± 3.8) and 9.8 (SD± 3.2) years for girls and boys respectively. Inter-observer agreement for bone age determination was 0.987 for radiographs and 0.980 for DXA using the G&P technique. For Observer 1, intra-observer agreement for radiographs and DXA was 0.993 and 0.983 respectively, and 0.995and 0.994 respectively for Observer 2. Poor DXA image quality did not allow bone age determination using the TW3 method.

Conclusion: Bone age can be determined from left hand/wrist DXA scans using G&P. However, limited DXA image quality prohibits its use for bone age assessment using the TW3 method.

6.2 Introduction

Bone age assessment from left hand radiographs is a frequently employed and useful diagnostic technique. Children with certain endocrine disorders or suspected skeletal dysplasia are usually considered for bone age assessment. Additionally, when planning for orthopaedic surgery or monitoring the response of the skeleton to certain treatments such as hydrocortisone, an assessment of bone age is required. The most commonly used techniques to determine bone age are the Greulich & Pyle (G&P) and Tanner & Whitehouse (TW3) methods (Greulich and Pyle, 1959; Tanner *et al.*, 2001).

There has been an improvement in dual energy x-ray absorptiometry (DXA) scanners during the last decade. This includes higher scan resolution compared to older generation machines. Some studies have suggested that these scans can now be used instead of conventional radiography for diagnostic purposes, including bone age assessment from the left hand and wrist, with the advantage of lower radiation dose. The effective dose produced from a DXA hand-wrist scan has been reported to be 0.1 µSv compared to 1 µSv from a hand-wrist radiograph (Mettler *et al.*, 2008). Although the radiation dose from hand-wrist radiographs is relatively low, no radiation exposure is without risk (Hall, 2009). Furthermore, children who suffer from chronic diseases require regular bone age monitoring, which means repeating hand-wrist radiographs many times during their childhood. Radiation dose should therefore be reduced where possible, particularly when the procedure involves children (Hall, 2009). Finally, several studies have stressed the need for updating the G&P standard or establishing a

local standard that serves children from particular regions (Ontell *et al.*, 1996; Zafar *et al.*, 2010; Hawley *et al.*, 2012b). If bone age can be determined from DXA, healthy children recruited to develop such standards would be exposed to much lower radiation dose.

A previous study suggested that DXA could replace radiographs for bone age assessment (Pludowski, Lebiedowski and Lorenc, 2004). However questionable statistical tests were applied, and a Polish reference method rather than G&P was used. These limitations were addressed in another recent study, with results supporting the use of hand and wrist DXA scans for bone age assessment using G&P, although the authors stated that further validation was required (Heppe *et al.*, 2012). No previous studies have assessed the feasibility of determining bone age from DXA scans using the TW3 method. Furthermore, BoneXpert software was developed in 2009, enabling automatic calculation of bone age, according to the G&P and TW3 methods. The software performance with regard to images taken using modalities other than conventional radiography has not been evaluated, in which this study will evaluate.

6.3 Materials and Methods

6.3.1 Study design

20 patients were recruited from Endocrine Clinic at Sheffield Children's NHS Trust. We recruited only children who were having a left hand radiograph for clinical purposes and who, along with their parents/legal guardians provided full informed assent/consent. Purposive sampling was used to include a total

of 20 patients; two males and two females from each of the following 5 age groups: <5; 5 to 7; 8 to 10; 11 to 13 and 14 to 16 years.

Ethical approval was granted from the local Research Ethics Committee (Yorkshire and Humber). Once informed consent had been obtained, a radiographer performed the left hand radiograph as follows: the left hand and forearm were positioned flat, hand and wrist and at least 1 inch of the distal forearm was included in the radiation field with the axis of the middle finger in direct axis with the forearm. Exposure factors varied slightly according to the patient's age; tube voltage 40-42 kV; 1.6 mAs; FFD 100 cm. Immediately after the radiograph, each patient had a left hand DXA scan (iDXA; General Electric, formerly Lunar Corp., Madison, WI).

All left hand radiographs and DXA scans were anonymised such that corresponding patient's images were not identifiable. Two observers, independently determined bone age from radiographs and DXA scans. Each observer independently assessed all patients' bone age on two different occasions with at least two weeks' interval. On both occasions the images were interpreted in random and varied order. For the G&P method, when the patient's bone age was thought to lie between two adjacent standards the intermediate value was assigned as the bone age. For the purpose of assessing bone age using the BoneXpert, hand-wrist DXA scans were extracted from the scanner in DICOM format. The images were then uploaded to a standalone version (Visiana, Holte, Denmark, v2.5.1.1).

6.3.2 Image Quality Assessment

Adequacy of hand positioning was rated by one of the researchers, a radiographer, using a system developed by Cockill et al (2014) (11). Inclusion of all anatomical structures (bones of the hand, wrist, radius and ulnar), thumb position and finger positions were scored on a 3-point scale (1 = poor, 2 = moderate, 3 = good) generating a possible total score of 9. Poor hand positioning included any image that scored 4 or less. Adequate positioning was a score between 5-7 and good positioning was a score of 8-9. In addition, the overall image quality was assessed by the same radiographer using the system described by Piraino et al (Piraino *et al.*, 1999).

6.3.3 Statistical Analysis:

The concordance between 1) independent readings of the two observers for radiographs and for DXA (i.e. comparing modalities) was evaluated using the interclass correlation coefficient (ICC). Bland Altman plots were also used to evaluate the correlation between the two imaging modalities (Martin Bland and Altman, 1986). The mean bone age obtained by both methods was plotted against the line of equality to assess the agreement between the two methods. Paired *t* tests were used to calculate significant difference between DXA and radiographs in terms of hand positioning and image quality. Statistical analyses were performed using the Statistical Package for the Social Sciences, version 21 for Windows (SPSS, Chicago, IL).

6.4 Results

The chronological age of our sample ranged from 3 to 16 years, with a mean of 9.4 years (± 3.8) in girls and 10.3 years (± 3.2) in boys. The overall mean chronological age was 9.8 years (± 3.5). Time taken to position for and obtain the radiographs and DXA scans ranged from one to two minutes and from three to five minutes respectively.

Results of inter and intra-observer reliability for bone age determination are presented in Table 6.1. G&P bone age assessment results in boys and girls from radiographs and DXA for two observers are presented in Table 6.1 (following page).

Table 6.2: Inter-/intra-observer reliability for bone age determination using the G&P atlas

	Intraclass Correlation Coefficient: Mean (95% CI)		
Modality	Interobserver Reliability	Intraobserver Reliability	
	Observers 1 and 2	Observer 1	Observer 2
Radiographs	0.994 (0.985-0.998)	0.993 (0.983-0.997)	0.995 (0.988-0.998)
DXA	0.987 (0.967-0.995)	0.983 (0.958-0.993)	0.994 (0.985-0.998)

Table 6.1: G&P bone age assessment from radiographs and DXA for 2 observers

Study ID	Chronological Age (years, months) (G&P SD, months)	Bone Age (years, months)			
		Radiographs		DXA	
		Observer 1	Observer 2	Observer 1	Observer 2
1	3y11m (6.5m)	5y9m	5y9m	5y9m	5y
2	4y6m (7.8m)	5y	5y	4y6m	5y
3	4y11m (8m)	5y9m	6y10m	5y9m	6y10m
4	6y7m (9.3m)	5y	5y	4y6m	5y
5	6y8m (9.3m)	5y	5y	5y	5y
6	7y4m (8.3m)	7y10m	8y10m	7y10m	7y10m
7	7y7m (8.3)	4y10m	6y10m	5y9m	6y10m
8	8y (8.8m)	10y	10y	8y10m	10y
9	8y6m (10.8m)	8y	8y	7y	7y
10	8y8m (8.8m)	8y10m	8y10m	10y	10y
11	8y8m (10.8m)	8y	8y	10y	8y
12	10y4m (11.4m)	7y	7y	7y	8y
13	10y6m (10.8m)	13y	12y	13y6m	14y
14	11y5m (12 m)	11y	11y6m	12y6m	11y6m
15	11y11m (10.5m)	6y	7y	6y	7y
16	12y11m (14m)	13y	13y6m	13y6m	14y
17	13y2m (14.6m)	13y	13y6m	13y	13y
18	14y9m (12m)	14y6m	15y	15y	15y
19	15y (14.2m)	13y6m	14y	14y	15y
20	15y3m (11.2m)	13y6m	13y6m	13y6m	15y

Paired samples T test showed a significant difference between DXA scans and radiographs (p< 0.001) for both the overall image quality and hand positioning (Table 6.3). In total, 14 hand radiographs showed good positioning compared to 10 DXA scans. Poor positioning was seen in two hand-wrist DXA scans. The mean rating for hand positioning was 7.95 (±0.68) and 6.7 (±1.12) for radiographs and DXA scans retrospectively, while the overall mean image quality was 3.87 (±0.45) for radiographs and 1.21 (±0.24) for DXA scans.

Table 6.3: Assessment of hand positioning and image quality for DXA scans and radiographs

	Radiographs	DXA	P value	95% CI (Lower, Upper)
Mean positioning rating (SD)	7.95 (0.68)	6.7 (1.12)	< .01	(0.52, 1.78)
Mean image quality rating (SD)	3.87 (0.45)	1.21 (0.24)	< .001	(2.39, 2.93)

Image quality and hand positioning is vital when using the TW3 method, as it requires a specific comparison of each bone to the standard. However, the level of image quality required by TW3 was not be achieved in this study by the DXA scans. Therefore, we could not determine bone age from DXA scans using the TW3 method. Additionally, after extracting the hand-wrist DXA scans in DICOM format, images were then uploaded to the BoneXpert software. All of the hand-wrist hand DXA images were rejected by the software for the reason of poor image quality.

In regard to the G&P method, differences between bone age determined from radiographs and DXA were normally distributed. Figures 6.1 and 6.2 (this and the following page) show separate plots for Observers 1 and 2, with limits of agreement (mean ± 1.96). Bone age assessed from radiographs and DXA were also plotted against the line of equality (Figure 6.3–following page). Only a small difference was observed between DXA scans and radiographs (all paired assessments lie close to the line of equity).

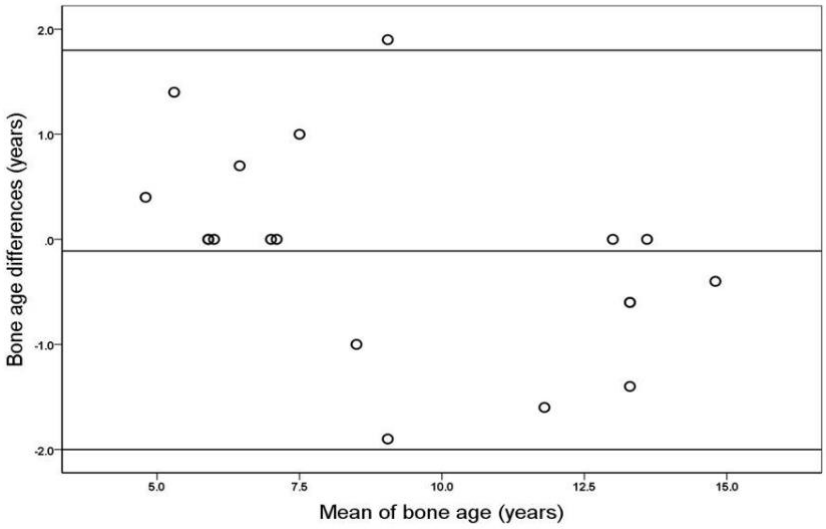


Figure 6.1: Bland Altman plot of the variation between radiographs and DXA for Observer 1 showing the absence of a systematic bias

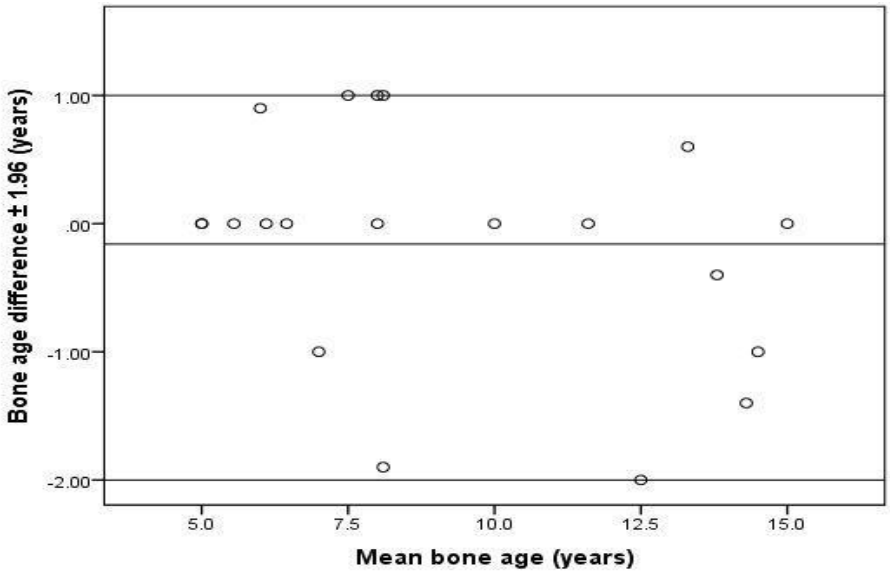


Figure 6.2: Bland Altman plot of the variation between radiographs and DXA for Observer 2 showing the absence of systematic bias

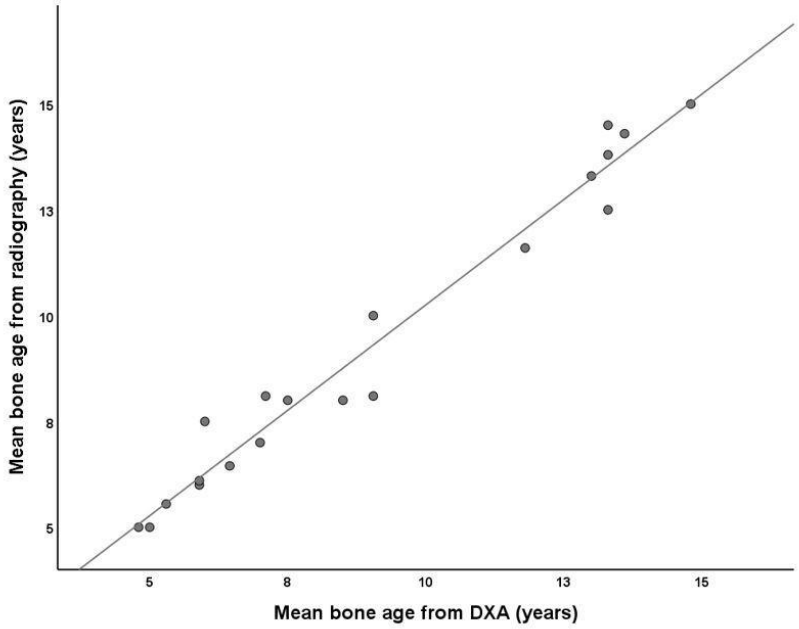
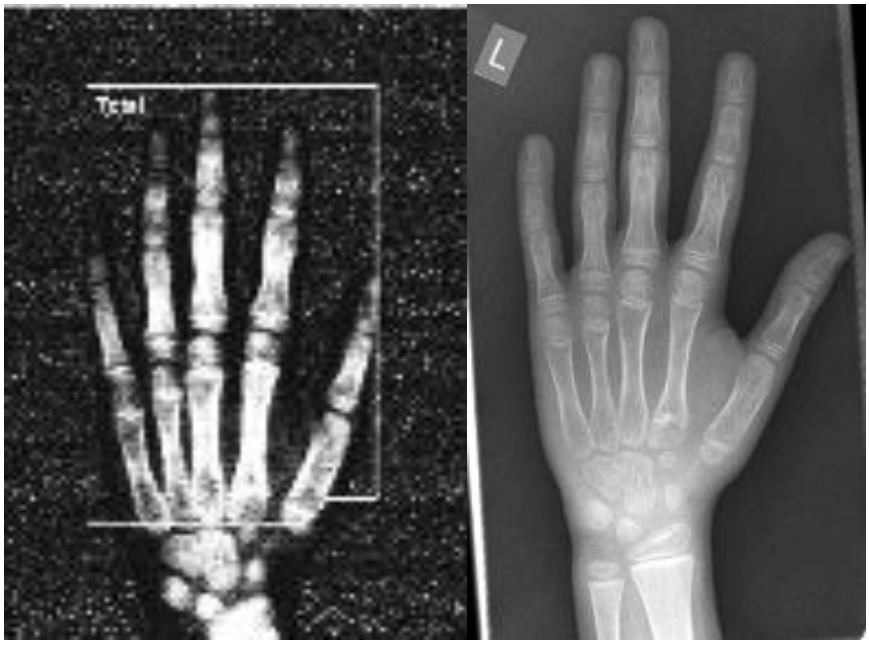


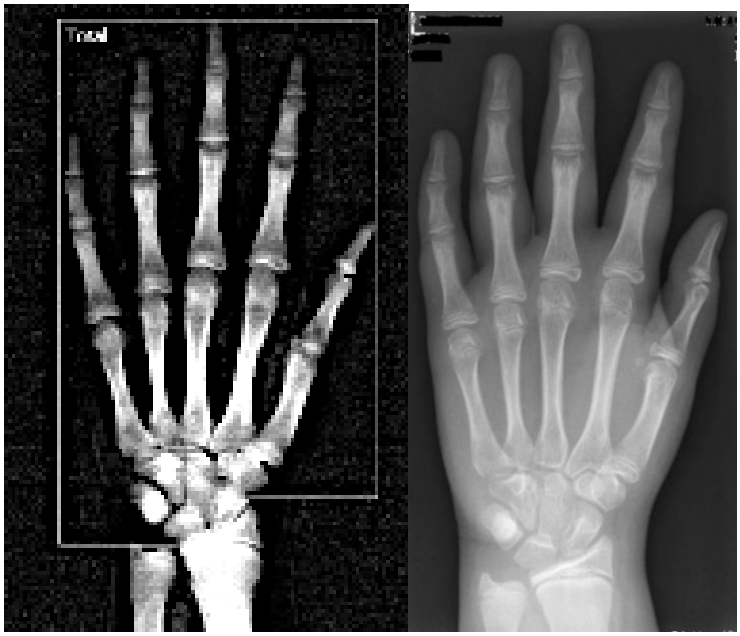
Figure 6.3: Plot of mean bone age for DXA and radiographs against line of equality, which shows high correlation between mean bone age acquired by the two methods.

Figures 6.4 and 6.5 are examples of poor and good quality DXA scans; their corresponding radiographs are also illustrated for comparison.



a b

Figure 6.4: Example of poor quality DXA image with corresponding radiograph for comparison (male; 8 years 8 months old) (a) left hand DXA, (b) left hand radiograph



a b

Figure 6.5: Example of good quality DXA image with corresponding radiograph (female; 12 years 11 months). (a) left hand DXA, (b) left hand radiograph

6.5 Discussion

Bone age assessment is a useful technique for managing children with certain endocrine and hereditary disorders and for planning timing of therapeutic procedures. Two methods are usually used for bone age assessment from left hand radiographs, namely Greulich and Pyle (G&P) and Tanner and Whitehouse (TW3). The aim of this study was to evaluate the feasibility of bone age determination from left hand DXA scans using both methods.

In regard to the G&P method, high inter and intra-observer correlation for both DXA and radiographs was demonstrated and the mean bone age assessed from radiographs and DXA showed high correlation when plotted against the line of equality. Our results suggest that there is no significant difference between bone age acquired by DXA and radiographs using G&P. This agrees with the study by Heppe et al who found high correlation between bone age assessment using DXA scans and radiographs.

All bone age values lay on the range of ± 1 year compared to radiographs, except for two cases where the differences were within ± 2 years. This is likely to be related to the poor quality of the DXA scans as both were of poor quality, one of which is shown in Figure 6.4. A significant difference between DXA scans and radiographs ($p < 0.001$) was observed in terms of hand positioning. The ulna was excluded from the scan field in 15% of the left hand DXA scans. This is likely to be operator dependent, given that different radiographers performed the scans.

Additionally, the duration of the scan which is, on average, 30 seconds compared to less than 1 second for radiographs, may also have led to the

child moving his/her hand, resulting in exclusion of the ulna. Visibility of soft tissue and fat planes was poor in all DXA scans compared to radiographs, but visibility of these structures is not required for bone age estimation. The mean score for visibility of cortical edges and individual trabecula was lower for DXA scans than radiographs. In addition to lower resolution, this might be due to lower radiation used for DXA scans compared to radiographs. The lower quality of DXA did not detract from the ability to assess bone age using the G&P technique.

The accuracy of G&P method of examining a whole radiograph with that of standards in an atlas has been questionable. This is because not all bones/epiphyses mature at exactly the same rate, therefore it is more desirable to examine bones individually to obtain a result that is more likely to be closer to the individual's actual level of skeletal maturity. In contrast, the TW3 method allows the user to examine each relevant bone and/or epiphysis for specified size and shape changes. Although the GP method is quicker to perform for the novice, once an assessor becomes familiar with the TW3 method, it is possible to perform an assessment in less than 3 minutes. Those radiographs that indicate full skeletal maturity in every bone can be assessed in less than 30 seconds. The time taken to assess perforem TW3 assessment is mainly dependent upon the assessor's experience with some studies stated that the TW3 assessment can take up to 7 minutes (De Sanctis *et al.*, 2014).

Within the current study, the assessor found that determining bone age using DXA took more time on average than when using radiographs due to

the reduced spatial resolution of the DXA images. Determining bone age from the DXA images often required more time to view the different maturity markers, whereas the maturity markers on the radiographs were for the most part easier to assess due to the higher resolution.

The TW3 method is a single bone method that requires high quality images in order to assess the epiphyseal plates of the hand/wrist. Furthermore, the method divides the ossification centres within the hand into two groups, the RUS (radius, ulnar and short bones) which involves the radius, ulnar, metacarpals and phalanges and the carpals, which include all the carpal bones, except the pisiform. The RUS technique is widely used which required at least 13 bones of the hand to be clearly visualised on the image (Tanner *et al.*, 2001). However, the images obtained by DXA on the current study showed a lower overall quality compered to radiographs. Although all of the DXA images included the RUS bones (except one image), the assessor could not determine the specific stage of the bones described by the TW3 method.

Due to inadequate DXA image quality, bone age could not be determined from DXA scans using the TW3 method. As an aside, the low quality of DXA also prohibits the use of BoneXpert software for the automatic determination of bone age (14).The main limitation of this study is the small sample size; however, this was always intended as a feasibility study, with the knowledge that should results be encouraging, then a further validation study would be required.

6.6 Conclusion

The clinical role for DXA scans is expanding and this feasibility study gives evidence that DXA scans may potentially be used for more than just measuring bone mineral density. Our results indicate that before widespread use for bone age assessment by the TW3 method, image quality requires further improvement. On the other hand, it is potentially feasible to assess bone age from DXA using G&P, with the advantage of lower radiation dose and thus allowing development of local normal standards from healthy children.

CHAPTER 7

Estimating bone mass in children: can bone health index replace dual energy x-ray absorptiometry?

Alshamrani K[1], Messina F[2], Bishop NJ[1,3], Offiah AC[1,3]

[1] Department of Oncology & Metabolism, University of Sheffield, UK

[2] Leeds Institute of Clinical Trials Research, University of Leeds, UK

[3] Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield, UK

Correspondence: Khalaf Alshamrani, Academic Unit of Child Health, Damer Street, Sheffield, S10 2TH

Kalshamrani1@sheffield.ac.uk

Available online (Pediatric Radiology) DOI: 10.1007/s00247-018-4309-3

7.1 Abstract

Background: Children are commonly treated with bisphosphonates, which have been shown to increase metacarpal cortical width. BoneXpert is an automated software tool that computes bone health index from hand radiographs by measuring cortical thickness, width and length of the three middle metacarpal bones. Bone health index is potentially a low cost, relatively cheap tool that may have the benefit of predicting fracture risk in children.

Objective: To compare bone health index with bone mineral density as measured from dual energy x-ray absorptiometry scans in patients with and without bisphosphonate treatment.

Methods: We documented absolute values and z-scores for whole body less head and lumbar spine bone mineral density then correlated these with bone health index, which have been acquired on the same day, in different patient groups, depending on their ethnicity and diagnosis.

Results: 293 Caucasian patients (mean age 11.5 ± 3.7 years) were included. Bone health index showed moderate to strong correlation with absolute values for whole body (r=0.52) and lumber spine (r=0.70) in those not treated with bisphosphonates and moderate correlation absolute values for whole body (r=0.54) and lumber spine (r=0.51) for those treated with bisphosphonates. Bone health index showed weak correlation z-scores, ranging from r = 0.11 to r = 0.35 in both groups.

Conclusion: The lack of a strong correlation between dual energy x-ray absorptiometry and bone health index suggests that they may be assessing different parameters.

7.2 Introduction

Assessment of bone mineral density and bone quality is essential to diagnose patients with diseases affecting the skeleton. In children, the reference standard for assessment of bone mineral density is dual energy x-ray absorptiometry. Dual energy x-ray absorptiometry is a valuable tool in patient management, where bone mineral density is assessed at appropriate intervals to monitor response to therapy in patients with low bone mass (Bishop *et al.*, 2008). Bisphosphonates are commonly used in such patients (e.g. those with osteogenesis imperfecta) and have been shown to increase cortical width (Glorieux *et al.*, 1998). However, dual energy x-ray absorptiometry values are influenced by bone size, therefore, bone mineral density is usually underestimated in children with small bones and overestimated in children with large bones; this is because the depth of the bone is not accounted for (Adams, 2013). Additionally, dual energy x-ray absorptiometry cannot predict fracture risk in children, rather it forms part of a comprehensive skeletal health assessment to monitor patients with low bone mineral density.

Over the last three decades, quantitative bone imaging techniques have been improved and tools for analysing images have been developed. One of these methods is radiogrammetry, where the middle phalangeal width and cortical thickness are measured and results presented as the “cortical index” (Barnett and Nordin, 1960). BoneXpert software was developed specifically for children and automatically calculates bone age and bone mass (Thodberg *et al.*, 2010). The software measures the cortical thickness, width and length of the three middle metacarpals and results are expressed as

the, “bone health index”. The software also provides a standard deviation score, which enables comparison with healthy Caucasian children. A small number of studies suggest a potential role for the use of bone health index in assessing bone health in children (Nusman *et al.*, 2015; Schündeln *et al.*, 2016; Neelis *et al.*, 2017). However, there are limitations to these studies, including small participant numbers (Nusman *et al.*, 2015; Schündeln *et al.*, 2016) and an extended interval of up to 8 months between dual energy x-ray absorptiometry and radiographs (Neelis *et al.*, 2017). Patients on bisphosphonate therapy were not included in any of these previous studies, yet this group may benefit the most, given that bisphosphonates increase cortical thickness; the very parameter on which the bone health index is based.

The aim of this study was to compare bone mass measured by BoneXpert and expressed as bone health index with bone mineral density dual energy x-ray absorptiometry readings acquired on the same day for different clinical reasons and in a large cohort of children, including those on bisphosphonates.

7.3 Materials and Methods

7.3.1 Study design

We retrospectively identified dual energy x-ray absorptiometry scans and left hand radiographs of patients who attended Sheffield Children’s NHS Foundation Trust Hospital, United Kingdom, between February 2010 and January 2017. The following inclusion criteria were applied; (1) patient aged

above 5 years and under 18 years, (2) dual energy x-ray absorptiometry scans and hand radiographs obtained on the same day.

7.3.2 Hand radiographs and dual energy x-ray absorptiometry scans

BoneXpert software (PACS Server version, Visiana, Holte, Denmark) was used to analyse the hand radiographs. All radiographs were in DICOM format. The software calculated the bone health index based on cortical thickness, width and length of the three middle metacarpals.

For bone health index calculations, “Caucasian” was the default ethnicity at the time of analysis. The data was analysed according to whether patients were or were not on bisphosphonate treatment. Cases were excluded from the study if the BoneXpert software was unable to read the radiograph.

Areal bone mineral density of total body less head and lumber spine L1-L4 were extracted from each patient’s dual energy x-ray absorptiometry scan. These values were adjusted for age and sex based on normative data provided by the manufacturer. Patient’s age, sex, and the indication for dual energy x-ray absorptiometry were extracted.

7.3.3 Statistical analysis

Statistical analysis was performed using SPSS version 25 for PC (IBM, Armonk, New York). The z-scores of bone mineral density of the total body less head and spine were adjusted for bone age to evaluate the impact of this adjustment on correlation with the bone health index standard deviation score. Each z-score adjusted for bone age for those patients treated with bisphosphonates is based on the computed z-score values (i.e. the internally studentised residuals from the regression analysis that includes bone age) from the untreated patients.

The correlation between bone health index and bone mineral density of the total body less head and the spine were assessed separately using Pearson's correlation. Additionally, correlation between bone health index standard deviation score and z-score of bone mineral density of the total body less head and the spine were assessed separately.

The correlation between the adjusted z-scores and bone health index standard deviation score were then determined. The strength of the correlations was interpreted according to Evans, in which the correlation is deemed to be "very weak" when the r value is less than 0.19, "weak" between 0.20 and 0.39, "moderate" between 0.40 and 0.59, "strong" between 0.60 and 0.79, while being "very strong" when the r value is between 0.80 and 1.0 (9).

Finally, we generated Bland Altman plots to graphically illustrate the strength of agreement between the two modalities for the non-bisphosphonate and bisphosphonate groups. All procedures performed in this study were in accordance with the ethical standards of our institution.

7.4 Results

7.4.1 Patient characteristics

Initially, 577 dual energy x-ray absorptiometry/radiograph pairs were identified. Diagnoses included osteogenesis imperfecta (51%), primary osteoporosis (9.5%) and recurrent fracture (5.8%). All diagnoses/indications and patient characteristics are presented in Tables 7.1 (following page) and Table 7.2 (page 150) respectively.

Table 7.1: Diagnosis/Indication for Investigation

Diagnosis/Indication for Investigation	No treatment	bisphosphonate	Current/past bisphosphonate treatment
Acute back pain		4	
Bone marrow transplant		7	
Calcinosis cutis		6	
Cerebral palsy		9	
Crohn's disease		5	3
Cystic fibrosis		9	4
Fanconi anemia		3	
Growth delay		13	
Hypocalcemia		6	
Hypophosphatasia		4	
Juvenile arthritis		10	8
Malabsorption		4	
Osteogenesis Imperfecta		12	138
Post colectomy		3	
Primary osteoporosis		15	13
Recurrent Fracture		11	6
Total		121	172

Table 7.2: Mean and standard deviation of dual energy x-ray absorptiometry and bone health index measurements

	Bisphosphonate group Mean (±SD)	Non- Bisphosphonate group Mean (±SD)
Number	172	121
Age (years)	12.06 (3.5)	10.87 (3.98)
Bone Age* (years)	11.50 (3.7)	9.86 (4.25)
BMD-spine	0.82 (0.18)	0.83 (0.23)
Z-score of BMD-spine	-0.77 (1.45)	-0.26 (1.63)
Adjusted z-score of Bone mineral density-spine	0.00 (1.0)	0.46 (1.28)
BMD-total body	0.86 (0.16)	0.77 (0.19)
Z-score of BMD-total body	-0.62 (1.38)	-0.43 (1.36)
Adjusted Z-score of BMD-total body	0.00 (1)	-0.42 (1.05)
Bone health index	4.39 (0.61)	4.22 (0.68)
Bone health index standard deviation scores	-1.20 (1.23)	-1.32 (1.45)

BoneXpert could not interpret 31 (5.6%) radiographs for a number of reasons including abnormal bone shape, cortical inconsistencies or inconsistencies in length and the image being too sharp. A total of 189 DXA/radiograph pairs were excluded as these pairs were acquired for follow-up which would bias statistical analyses. No dual energy x-ray absorptiometry/radiograph pairs were identified for Africans in comparison to a total number 32 dual energy x-ray absorptiometry/radiograph pairs for Asians.

However, the Asian patients were excluded from the analysis due to the small number of dual energy x-ray absorptiometry/radiograph pairs identified. Therefore, the final analysis included dual energy x-ray absorptiometry and hand radiographs of 293 patients, 172 (59%) of whom had received bisphosphonate treatment.

7.4.2 Dual Energy X-ray Absorptiometry Scans and Bone Health Index

As an overall analysis, bone health index correlated moderately with the absolute values of bone mineral density the total body and the spine ($p<0.01$) (Table 7.3– following page). The data were then divided into two groups depending on whether or not patients had received bisphosphonate treatment. As seen in Table 7.3, correlation was stronger in the non-bisphosphonate group; bone mineral density of the total body and ($r=0.704$) and the spine ($r=0.524$, $p<0,01$).

Table 7.3: Correlation coefficients between bone health index (BHI) and DXA, and bone health index standard deviation (BHI SDS) scores and z-score reads in bisphosphonate naïve and treated patients

		Overall	p value	Bisphosphonate group	p value	Non-bisphosphonate group	p value
BHI	BMD-spine	0.590	<0.01	0.516	<0.01	0.704	<0.01
	BMD-total body	0.532	<0.01	0.542	<0.01	0.524	<0.01
BHI SDS	Z-score of BMD-spine	0.174	<0.01	0.047	0.26	0.350	<0.01
	Z-score of BMD-total body	0.244	<0.01	0.190	0.20	0.306	<0.01
BHI SDS	Z-score of BMD-spine (adjusted for BA)	0.215	0.03	0.115	0.12	0.350	<0.01
	Z-score of BMD-total body (adjusted for BA)	0.258	<0.01	0.257	<0.01	0.253	<0.01

The bone health index standard deviation score showed weak correlation with z-score of the total body less head and the spine (adjusted only for age and sex) in both groups (Table 7.3). The z-score of bone mineral density of the total body less head and the spine were then adjusted for bone age.

The relationship of bone mineral density of the spine “adjusted for age and sex alone” and “adjusted for age, sex and bone age” showed similar slopes in both groups with Pearson correlation of 0.735 (r^2 = 53.9%) (Figure 7.1–following page).

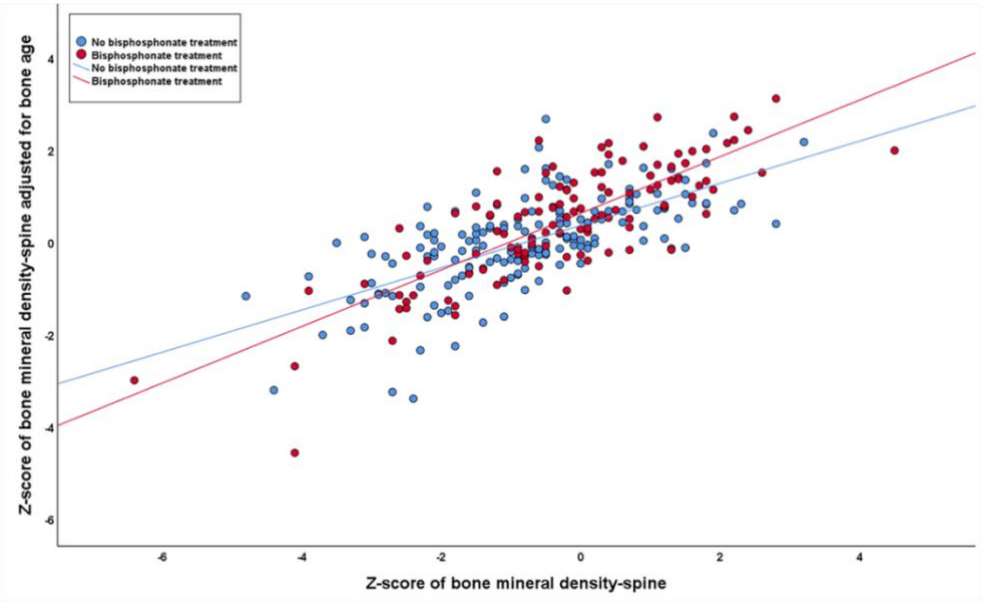


*Figure 7.1: Relationship of z-score of bone mineral density of the **spine** adjusted for age and sex alone, and z-score adjusted for age, sex and bone age showing similar slop.*

Additionally, the relationship of bone mineral density of the total body less head "adjusted for age and sex alone" and "adjusted for age, sex and bone age" showed similar slopes in both groups, with Pearson correlation of 0.459 (r^2 = 20.8%) (Figure 7.2).

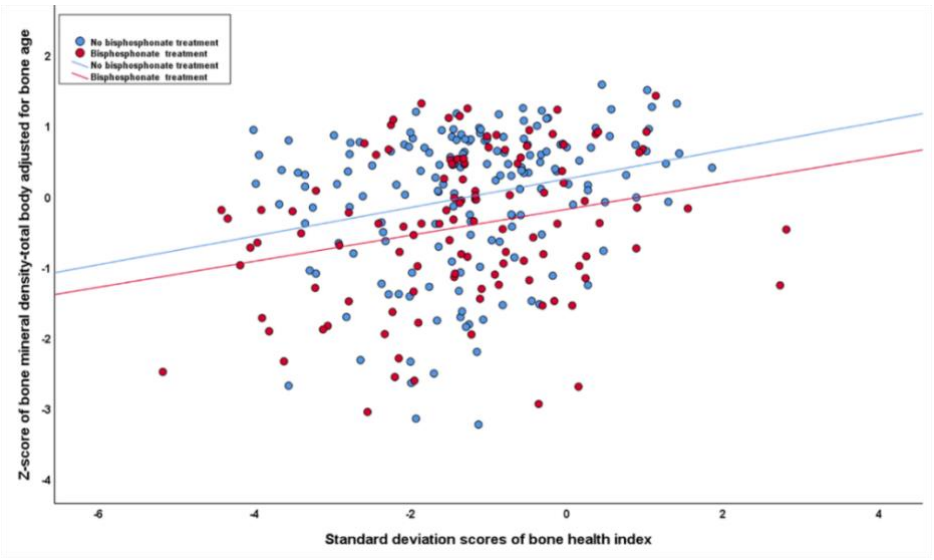


*Figure 7.2: Relationship of z-score of bone mineral density of the **total-body** adjusted for age and sex alone, and z-score adjusted for age, sex and bone age showing similar slop.*

The bone health index standard deviation score showed weak correlation with the z-score of bone mineral density of the total body less head and the spine (adjusted for age, sex, bone age) (Table 7.3). Bland Altman plots showed sup-optimal agreement between z-score of bone mineral density of the total body less head and the spine (adjusted for age, sex, bone age) and bone health index standard deviation scores (Figure 7.3, 7.4 – this and following page).

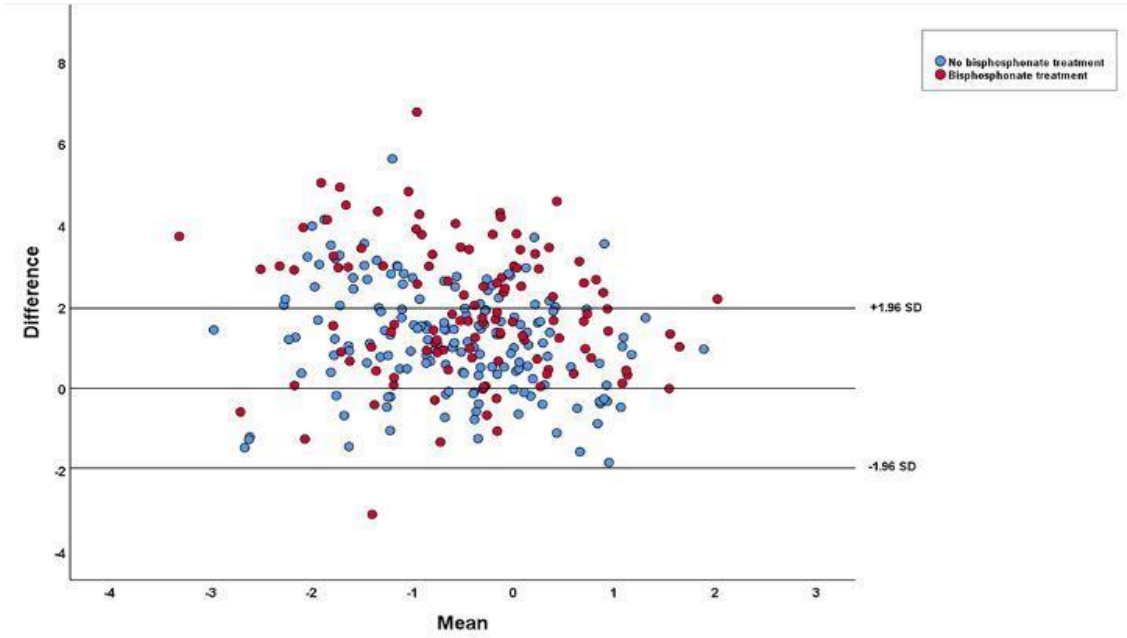


Figure 7.3: Bland Altman plot for the difference in bone mineral density ***spine*** *adjusted for bone age, and Bone Health Index z-score, versus the mean of the two estimates showing sup-optimal agreement between the two methods.*

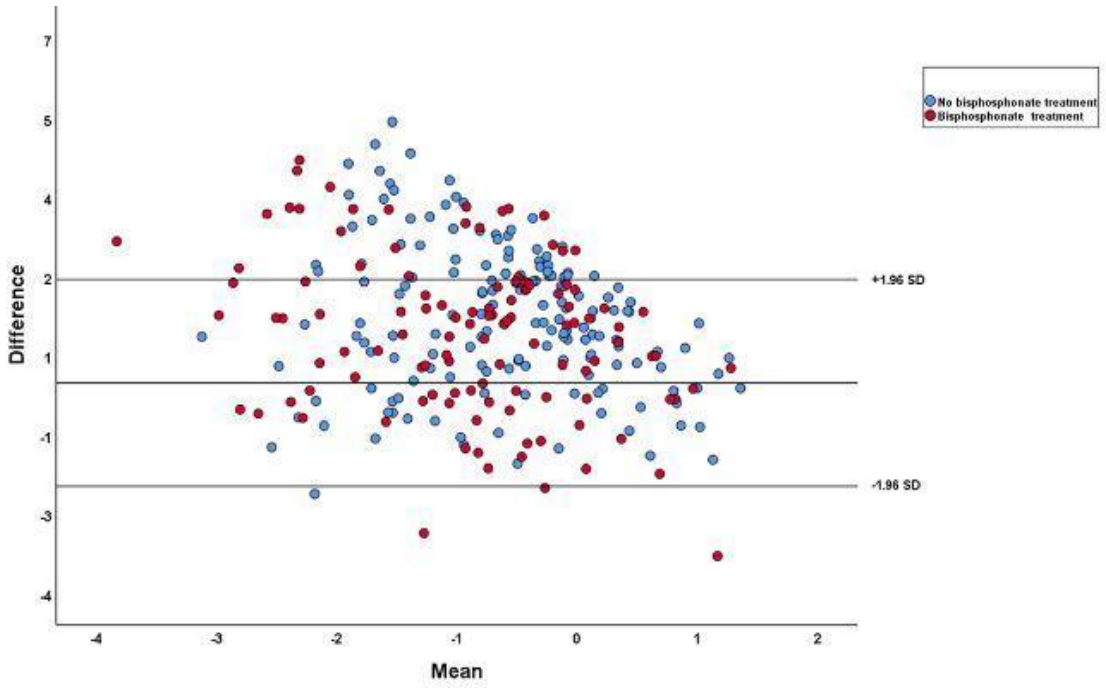


Figure 7.4: Bland Altman plot for the difference in bone mineral density ***total-body*** *adjusted for bone age, and Bone Health Index z-score, versus the mean of the two estimates showing sup-optimal agreement between the two methods.*

7.5 Discussion

This study compares bone mineral density measured by dual energy x-ray absorptiometry with bone mass calculated by BoneXpert in a cohort of Caucasian children. BoneXpert was able to provide a reading in the majority of cases. For bisphosphonate naïve children, there was strong correlation between bone health index and dual energy x-ray absorptiometry absolute values. Previous studies have shown similar correlation ranging from r=0.58 to r=0.85, although ethnicity of patients was not mentioned (Nusman *et al.*, 2015; Schündeln *et al.*, 2016; Neelis *et al.*, 2017).

BoneXpert also provides a bone health index standard deviation score based on data collected from healthy Caucasian children. The bone health index standard deviation score provides a measure of the extent to which a

patient's bone mass is deviated from that of healthy Caucasian children of the same bone age and sex. We found a weak correlation between bone health index standard deviation score and z-scores of dual energy x-ray absorptiometry, even after adjusting the z-scores for bone age. The reasons for this are uncertain but might include differences in other parameters of reference and study populations. Bland Altman plots showed systematic bias in which differences are higher than means when means are lower, and the differences do not reach zero until the average value reaches or exceeds 2 standard deviations. However, this is more likely to be due to the fact that the data adjusted for bone age are based on the computed z-score values from patients who had no bisphosphonate treatment.

Bone health index of patients who had not been on bisphosphonate treatment showed a strong correlation, which might suggest that bone health index is a useful tool to monitor children's bone health in this group of patients. In the bisphosphonate group, bone health index showed moderate correlation with absolute dual energy x-ray absorptiometry measures. Approximately 79% of the bisphosphonate group were patients with osteogenesis imperfect. The metacarpals of those group of patients have smaller bone thickness (external size) and thinner cortices than normal (Arundel and Bishop, 2010). During treatment with bisphosphonates, cortical thickness increases (Marini *et al.*, 2017). This is likely to offer BoneXpert an advantage in this particular group of patients, as the bone health index measured by BoneXpert is dependent on cortical structure, while dual energy x-ray absorptiometry depends on both cortical and trabecular structures. The weaker correlation between bone health index and dual

energy x-ray absorptiometry in this group of patients may be because bone health index more closely reflects the “true” state of the children’s bones than does dual energy x-ray absorptiometry and merits studies to assess its role in predicting fracture risk in children.

There are some limitations to this study. Firstly, dual energy x-ray absorptiometry z-scores were adjusted for sex, age and ethnicity but not for height and weight. Adjusting for height and weight is expected to explain some of the variance, because dual energy x-ray absorptiometry reads may be affected by height and weight. Additionally, dual energy x-ray absorptiometry z-scores adjusted for bone age are based on the computed z-score values from patients who had no bisphosphonate treatment. Ideally, these scores should be based on the z-scores of a “healthy” population, which were not available to the authors.

7.6 Conclusion

Given the limitations of dual energy x-ray absorptiometry (dependence on body size, inability to predict fracture risk, length of time to obtain the scan compared to a hand radiograph), the lack of a strong correlation between dual energy x-ray absorptiometry and bone health index suggests that they may be assessing different parameters. The role of bone health index in assessing bone health in children warrants further study before it can be used as an adjunct to or replacement for dual energy x-ray absorptiometry. Future studies could investigate the clinical use of the bone health index values measured by BoneXpert to predict fracture risk.

CHAPTER 8

Overall discussion

8.1 Overall discussion and conclusions

Bone age assessment plays an important role in the clinical context and it is therefore important to use an accurate and reliable method. One of the aims of this thesis, was to systematically summarise and report the findings of studies in relation to the applicability of the G&P atlas to children and adolescents across different populations. This was followed by a meta-analysis where appropriate for accurate and robust results.

The systematic review revealed a large number of studies that have been conducted to assess the applicability of G&P atlas, which indicates the wide used of the G&P compared to other methods such as TW3, possibly due to its relative simplicity. A recent survey showed that the majority of paediatricians use the G&P atlas to determine bone age in children and adolescents. The applicability of this method has been assessed in different populations in which ethnicity and/or socioeconomic status are different to the original reference standard. These two factors are well known to influence the skeletal maturity of children.

The systematic review showed that the majority of the included studies failed to report the socioeconomic status of the included children. This makes it difficult to distinguish between the effects of socioeconomic status and the effects of ethnicity. The ethnicity/origin was the frequent factor reported within the studies, hence, the papers were grouped according to ethnicity. Meta-analysis of overall mean difference between BA and CA (across all age groups), showed that BA is statistically different to CA only in African females. However, analysing each age group separately for each ethnicity, showed statistically significant differences in Asians at the ages of 6 to 9,

and 17 years. This was confirmed by the current PhD study which assessed the applicability of the G&P to Saudi children and adolescents and which showed that BA was delayed in males younger than 13 years old when using the G&P atlas. These findings support the view that ethnicity has a major influence on BA. The TW3 method can be applied to females but not males from Saudi Arabia as the method tends to underestimate males at the CA of 8 and 9 years. The influence of ethnicity on the maturational rate has been documented in other Asian populations in which Asian children seem to mature sooner than Caucasian children.

In line with the systematic review, the BoneXpert-derived BA according to the G&P atlas appeared applicable to the modern population in the UK who are dominantly Caucasians. Although there has been improvement in socioeconomic conditions, this improvement has not affected the applicability of the G&P atlas to modern UK children and adolescents. In contrast to the G&P atlas, although the BoneXpert-derived BA-TW3 method appeared reliable to use in male populations from the UK, this was not the case for females. The BoneXpert-derived TW3 BA consistently underestimated the age of females by an average of 5 months, which should be considered by users during application of the method.

The complexity, the longer time and the variability among assessors associated with bone age assessment methods have led to the development of the BoneXpert software. The software is clinically acceptable and has been validated to use in certain ethnic groups. In this thesis, the software was used to analyse hand radiographs from the UK for the advantages of saving time and eliminating observer variability. However, the hand

radiographs from Saudi Arabia were subjectively rated by 3 observers as well as using the software, as the software has not previously been used in this particular population. The software seems to conform to the manual rating, although the manual rating showed a significant difference when using TW3 BA while the difference was not statistically significant when using BoneXpert-derived TW3 BA.

One of the other features the software offers beside bone age assessment is the calculation of bone mass from the hand radiographs. The software measures the cortical thickness, width and length of the three middle metacarpals and results are expressed as the, “bone health index” or “BHI”. This can be an advantage to children when monitoring of bone mass is required (e.g. response to therapy). Correlation between BHI and DXA, which is the reference standard currently used in paediatrics, was weak. These two methods are probably assessing different parameters, therefore BHI might be a “true” reflection of the state of the children’s bones, especially in those on treatment that increases cortical thickness (e.g. bisphosphonates).

Additionally, in this thesis, BoneXpert performance with regard to images taken using modalities other than conventional radiography was evaluated. The radiation dose of left hand DXA is much lower in comparison to that of a left hand radiograph, therefore DXA scans are appropriate for children in which the determination of bone age is required numerous times during their childhood. However, the low quality of DXA prohibits the use of BoneXpert software for the automatic determination of bone age. This has also limited the possibility of acquiring bone age using TW3 manually, as the TW3

method requires high quality images in order to assess the epiphyseal plates. Nevertheless, this lower quality of DXA did not detract from the ability to assess bone age using the manual G&P technique.

8.2 Limitations

In this thesis, there are several limitations. Beginning with the systematic review (Chapter Three, page 46), it was difficult to separate ethnicity from socio-economic status when it came to analysing the applicability of the G&P method to different populations. A large number of the included studies (34 studies) failed to report the socioeconomic status of their sample. Socioeconomic status is one of the factors that should be taken into account when determining bone age. With regards to the meta-analysis, a further limitation is the failure to calculate the mean absolute and root mean square errors, which might have further confirmed the accuracy of the G&P atlas in relation to each population. However, the individual observations were not provided within the studies that reported the mean BA and CA, therefore the mean error could not be calculated.

Another limitation is the use of cross-sectional data from the UK and Saudi Arabian populations to determine the applicability of the G&P and TW3 methods. The use of cross-sectional data involves some limitations which cannot be avoided. One is that the data shows what is happening at a particular time rather than what is happening to the individual participants over time. The growth of children is not linear and includes periods of acceleration and slowing down, hence longitudinal data would be more precise. However, the use of longitudinal data means that healthy children

are exposed to more ionising radiation, which is not without risk. Besides this, the use of cross-sectional data minimises time and cost restraints.

Another limitation is that for the studies which assessed the applicability of the G&P and TW3 methods (Chapters 4 and 5, pages 81 and 105 respectively), the hospital notes were not reviewed to ascertain full health in the children (although radiology and ED notes were scrutinised). Additionally, height and weight of recruited children was not recorded; it may be that body mass index affects the rate of skeletal maturation and the prevalence of overweight and obese children is well documented to be rising. Due to the innate inability of BoneXpert to assess images of females under 2 years old, males under 2.5 years in males and individuals of both sexes aged 15 years or older, the full picture in regard to the applicability of this method could not be determined. The study in relation to Saudi Arabians is limited by the lack of information in regard to socioeconomic status. The high rejection rate of the software as a result of insufficient image quality led to a low number of cases in some age groups, and results for these age groups should be interpreted with caution.

The study comparing the BHI to DXA in this thesis has some limitations. Firstly, DXA z-scores were adjusted for sex, age and ethnicity but not for height and weight. Adjusting for height and weight is expected to explain some of the variance, because DXA reads may be affected by height and weight. Additionally, DXA z-scores adjusted for bone age are based on the computed z-score values from patients who had no bisphosphonate treatment. Ideally, these scores should be based on the z-scores of a “healthy” population, which were not available to the authors.

The thesis has also intended to assess the feasibility of bone age determination using DXA hand-wrist scans. The main limitation of this study was the small sample size; however, this was always intended as a feasibility study with a view to a further larger scale validation study should the results be encouraging. In the event, results showed that image quality requires further improvement, before widespread use for bone age assessment by the TW3 method.

8.3 Future work

The findings in this thesis shows that bone age of certain age groups is deviated from the G&P and TW3 standards. It would be useful to produce reference data for each ethnic group. This however, requires collaboration with international investigators, which in turn requires training. The radiographs can be collected from primary care units, community centres, or schools.

The development of automated software that determines bone age and bone mass will lead to further studies. One of the benefits of the BoneXpert software is the availability of reference data. However, these data are for particular ethnic groups and differences exist between ethnicities when it comes to bone age and/or bone mass. The software would benefit from additional data in relation to other populations. The value of the software in monitoring change in bone mass, particularly in those treated with bisphosphonates, should be evaluated in a prospective study. Recording base line data and regular monitoring should produce robust results. Additionally, a further study could investigate the role of the software in predicating fracture risk in children.

References

Acheson, R., Vicinus, J. and Fowler, G. (1966) ‘Studies in the reliabilityof assessing skeltal maturity from x-ray: Part III . Greulich-Pyle atlas and Tanner-Whitehouse method contrasted.’, *Human biology*, 38(3), pp. 204–218. Available at: https://www.jstor.org/stable/pdf/41449263.pdf.

Acheson R M (1954) ‘A method of assessing skeletal maturity from radiographs; a report from the Oxford child health survey.’, *Journal of anatomy*, 88(4), pp. 498–508. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1244660&tool=pmcentrez&rendertype=abstract.

Adams, J. E. (2013) ‘Advances in bone imaging for osteoporosis’, *Nature Reviews Endocrinology*. Nature Publishing Group, 9(1), pp. 28–42. doi: 10.1038/nrendo.2012.217.

Adams, P., Davies, G. T. and Sweetnam, P. M. (1969) ‘Observer error and measurements of the metacarpal.’, *The British journal of radiology*. The British Institute of Radiology, 42(495), pp. 192–197. doi: 10.1259/0007-1285-42-495-192.

Afshari, R. and Bhopal, R. S. (2002) ‘Changing pattern of use of “ethnicity” and “race” in scientific literature [1]’, *International Journal of Epidemiology*. Oxford University Press, p. 1074. doi: 10.1093/ije/31.5.1074.

Al-Hadlaq, A. *et al.* (2007) ‘Comparative study between hand-wrist method and cervical vertebral maturation method for evaluation of skeletal maturity in Saudi boys’, *Pakistan Oral & Dental Journal*, 27(2), pp. 187–192. Available at: https://fac.ksu.edu.sa/sites/default/files/Comparative_study_between_hand-wrist_method_and_cervical_vertebral_maturation_method_for_evaluation_of_skeletal_maturity_in_Saudi_boys-PODJ-2007_0.pdf.

Alcina, M. *et al.* (2018) ‘Reliability of the Greulich and Pyle method for chronological age estimation and age majority prediction in a Spanish sample’, *International Journal of Legal Medicine*, 132(4), pp. 1139–1149.

doi: 10.1007/s00414-017-1760-x.

Andersen, E. (1971) ‘Comparison of Tanner-Whitehouse and Greulich-Pyle methods in a large scale Danish survey’, *American Journal of Physical Anthropology*, 35(3), pp. 373–376. doi: 10.1002/ajpa.1330350312.

Arundel, P. and Bishop, N. (2010) ‘Diagnosing osteogenesis imperfecta’, *Paediatrics and Child Health*. Elsevier Ltd, 20(5), pp. 225–231. doi: 10.1016/j.paed.2009.11.008.

Ashizawa, K. *et al.* (2005) ‘RUS skeletal maturity of children in Beijing.’, *Annals of human biology*, 32(3), pp. 316–325. doi: 10.1080/03014460500087725.

Awais, M. *et al.* (2014) ‘Comparison between Greulich-Pyle and Girdany-Golden Methods for Estimating Skeletal Age of Children in Pakistan’, *Journal of the College of Physicians and Surgeons Pakistan*, 24(12), pp. 889–893. doi: 12.2014/JCPSP.889893.

Barnett, E. and Nordin, B. E. (1960) ‘The radiological diagnosis of osteoporosis: a new approach.’, *Clinical radiology*. W.B. Saunders, 11(3), pp. 166–174. doi: 10.1016/S0009-9260(60)80012-8.

Baumann, U. *et al.* (2009) ‘Reference study on the time frame for ossification of the distal radius and ulnar epiphyses on the hand radiograph’, *Forensic Science International*. Elsevier, 191(1–3), pp. 15–18. doi: 10.1016/j.forsciint.2009.05.023.

Benjavongkulchai, S. and Pittayapat, P. (2018) ‘Age estimation methods using hand and wrist radiographs in a group of contemporary Thais’, *Forensic Science International*. Elsevier Ireland Ltd, 287, p. 218.e1-218.e8. doi: 10.1016/j.forsciint.2018.03.045.

Berst, M. J. *et al.* (2001) ‘Effect of knowledge of chronologic age on the variability of pediatric bone age determined using the Greulich and Pyle standards’, *American Journal of Roentgenology*, 176(2), pp. 507–510. doi: 10.2214/ajr.176.2.1760507.

Bhopal, R. (2004) ‘Glossary of terms relating to ethnicity and race: for reflection and debate.’, *Journal of epidemiology and community health*. BMJ

Publishing Group Ltd, 58(6), pp. 441–445. doi: 10.1136/jech.2003.013466.

Bhopal, R. and Donaldson, L. (1998) 'White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health', *American Journal of Public Health*, 88(9), pp. 1303–1307. doi: 10.2105/AJPH.88.9.1303.

Bishop, N. *et al.* (2008) 'Dual-Energy X-ray Aborptiometry Assessment in Children and Adolescents with Diseases that May Affect the Skeleton: The 2007 ISCD Pediatric Official Positions', *Journal of Clinical Densitometry*, 11(1), pp. 29–42. doi: 10.1016/j.jocd.2007.12.004.

Black, S. (2010) *Age estimation in the living : the practitioners guide*. First edit. Hoboken, New Jersey : John Wiley & Sons Inc.,.

Büken, B. *et al.* (2007) 'Is the assessment of bone age by the Greulich-Pyle method reliable at forensic age estimation for Turkish children?', *Forensic Science International*, 173(2–3), pp. 146–153. doi: 10.1016/j.forsciint.2007.02.023.

Büken, B. *et al.* (2009) 'Comparison of the three age estimation methods: Which is more reliable for Turkish children?', *Forensic Science International*, 183(1–3), p. 103.e1-103.e7. doi: 10.1016/j.forsciint.2008.10.012.

Bull, R. K. *et al.* (1999) 'Bone age assessment: A large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods', *Archives of Disease in Childhood*. BMJ Publishing Group, 81(2), pp. 172–173. doi: 10.1136/adc.81.2.172.

Caldas, M. de P., Ambosano, G. M. B. and Haiter-Neto, F. (2007) 'Use of Cervical Vertebral Dimensions for assessment of children growth', *J Appl Oral Sci*, 15(2), pp. 9–11. doi: S1678-77572007000200014 [pii].

Calfee, R. P. *et al.* (2010) 'Skeletal and chronological ages in American adolescents: current findings in skeletal maturation', *Journal of Children's Orthopaedics*, 4(5), pp. 467–470. doi: 10.1007/s11832-010-0289-z.

Cameron, N. (1997) 'The assessment of maturation', *South African Journal of Science*. Academic Press, pp. 18–22. doi: 10.1016/B978-0-12-383882-7.00020-9.

Cameron, N. (2002) ‘Assessment of Maturation’, in *Human Growth 2and Development*. Amsterdam ; London, pp. 363–382.

Cantekin, K. *et al.* (2012) ‘Bone Age Assessment: The Applicability of the Greulich-Pyle Method in Eastern Turkish Children’, *Journal of Forensic Sciences*, 57(3), pp. 679–682. doi: 10.1111/j.1556-4029.2011.02035.x.

Cavalli-Sforza, L. L. *et al.* (1995) *The History and Geography of Human Genes.*, *The Journal of Asian Studies*. doi: 10.2307/2058750.

Chaumoitre, K. *et al.* (2017) ‘Forensic use of the Greulich and Pyle atlas: prediction intervals and relevance’, *European Radiology*. Springer Berlin Heidelberg, 27(3), pp. 1032–1043. doi: 10.1007/s00330-016-4466-4.

Chiang, K. *et al.* (2005) ‘The reliability of using Greulich-Pyle method to determine children’s bone age in Taiwan’, *Tzu Chi Med J*, 6(17), pp. 15–18. Available at: http://www.tzuchi.com.tw/tzuchi/tcmj/94-6/1-4.pdf.

Cole, T. J. (2000) ‘Secular trends in growth’, *Proceedings of the Nutrition Society*, 59(2), pp. 317–324. doi: 10.1017/S0029665100000355.

Cole, T. J. (2003) ‘The secular trend in human physical growth: A biological view’, *Economics and Human Biology*, 1(2), pp. 161–168. doi: 10.1016/S1570-677X(02)00033-3.

Cole, T. J. *et al.* (2015) ‘Ethnic and sex differences in skeletal maturation among the Birth to Twenty cohort in South Africa’, *Archives of Disease in Childhood*. BMJ Publishing Group, 100(2), pp. 138–143. doi: 10.1136/archdischild-2014-306399.

Cole, T. J. (2015) ‘The evidential value of developmental age imaging for assessing age of majority’, *Annals of human biology*. Taylor & Francis, 42(4), pp. 379–388. doi: 10.3109/03014460.2015.1031826.

Cox, L. A. (1996) ‘Tanner-whitehouse method of assessing skeletal maturity: Problems and common errors’, *Hormone Research in Paediatrics*. Karger Publishers, 45(2), pp. 53–55. doi: 10.1159/000184848.

Crews, D. E. and Bindon, J. R. (1991) ‘Ethnicity as a taxonomic tool in biomedical and biosocial research.’, *Ethnicity & disease*, pp. 42–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1842520 (Accessed: 25

May 2017).

Damilakis, J. *et al.* (2010) ‘Radiation exposure in X-ray-based imaging techniques used in osteoporosis’, *European Radiology*, pp. 2707–2714. doi: 10.1007/s00330-010-1845-0.

Dembetembe, K. A. and Morris, A. G. (2012) ‘Is Greulich-Pyle age estimation applicable for determining maturation in male Africans?’, *South African Journal of Science*, 108(9–10). doi: 10.4102/sajs.v108i9/10.1036.

Demisch, A. and Wartmann, P. (1956) ‘Calcification of the Mandibular Third Molar and Its Relation to Skeletal and Chronological Age in Children’, *Source: Child Development*, 27(4), pp. 459–473. doi: 10.1016/0002-9416(57)90029-5.

Deprtment of Communities and Local Government (2015) ‘English indices of deprivation 2015’, 1(Imd), pp. 1–11. Available at: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 (Accessed: 10 July 2018).

Drèze, J. and Sen, A. (2011) ‘Putting Growth In Its Place’, *Outlook India*, 14 November, pp. 3–10. Available at: https://www.outlookindia.com/magazine/story/putting-growth-in-its-place/278843.

Easterlin, R. A. (2000) ‘The Worldwide Standard of Living Since 1800’, *Journal of Economic Perspectives*, 14(1), pp. 7–26.

European Society of Paediatric Radiology (2018) *Bone age for chronological age determination. Recommendation from the ESPR musculoskeletal task force group*. Vienna.

Eurostat (2017) *Asylum applicants considered to be unaccompanied minors*, *Eurostat Press Office*. Luxembourg. Available at: http://ec.europa.eu/eurostat/documents/2995521/8016696/3-11052017-AP-EN.pdf/30ca2206-0db9-4076-a681-e069a4bc5290 (Accessed: 2 January 2018).

Feijen, L. (2008) ‘The Challenges of Ensuring Protection to Unaccompanied and Separated Children in Composite Flows in Europe’, *Refugee Survey*

Quarterly. Oxford University Press, 27(4), pp. 63–73. doi: 10.1093/rsq/hdn049.

Freitas, D. *et al.* (2012) 'Short-term secular change in height, body mass and Tanner-Whitehouse 3 skeletal maturity of Madeira youth, Portugal', *Annals of Human Biology*, 39(3), pp. 195–205. doi: 10.3109/03014460.2012.674155.

Garamendi, P. M. *et al.* (2005) 'Reliability of the methods applied to assess age minority in living subjects around 18 years old', *Forensic Science International*, 154(1), pp. 3–12. doi: 10.1016/j.forsciint.2004.08.018.

Giuca, M. R. *et al.* (2012) 'Skeletal maturation in obese patients', *American Journal of Orthodontics and Dentofacial Orthopedics*. Mosby, 142(6), pp. 774–779. doi: 10.1016/j.ajodo.2012.07.011.

Glorieux, F. H. *et al.* (1998) 'Cyclic Administration of Pamidronate in Children with Severe Osteogenesis Imperfecta', *New England Journal of Medicine*. Massachusetts Medical Society, 339(14), pp. 947–952. doi: 10.1056/NEJM199810013391402.

De González, A. B. and Darby, S. (2004) 'Risk of cancer from diagnostic X-rays: Estimates for the UK and 14 other countries', *Lancet*. Elsevier, 363(9406), pp. 345–351. doi: 10.1016/S0140-6736(04)15433-0.

Govender, D. and Goodier, M. (2018) 'Bone of contention: The applicability of the Greulich–Pyle method for skeletal age assessment in South Africa', *South African Journal of Radiology*. Cannon Communications, 22(1), p. 6. Available at: https://sajr.org.za/index.php/sajr/article/view/1348/1912 (Accessed: 14 April 2019).

Greulich-Pyle (1957) 'A comparison of the physical growth and development of American born and native Japanese children.', *Am J Phys Antrhopol*, 15(4), pp. 489–515. doi: 10.1115/OMAE2007-29536.

Greulich, W. (1951) 'The Growth and Development Status of Guamanian School Children in 1947', *American Journal of Physical Anthropology*, 9(1), pp. 55–70. Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ajpa.1330090105 (Accessed:

16 October 2018).

Greulich, W. and Pyle, I. (1959) *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, *Stanford University Press*. London: Stanford University Press.

Griffith, J., Cheng, J. and Wong, E. (2007) 'Are Western skeletal age standards applicable to Hong Kong Chinese? – a comparison of Greulich and Pyle and TW3 methods.', *Hong Kong Med J*, 13(3), pp. S28-32. Available at: http://www.hkmj.org/system/files/hkm0706sp3p28.pdf.

Gulati, A. *et al.* (1991) 'Inter-relationship between dental, skeletal and chronological ages in well-nourished and mal-nourished children.', *Journal of the Indian Society of Pedodontics and Preventive Dentistry*, 8(1), pp. 19–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2056342 (Accessed: 12 October 2018).

Gungor, O. E. *et al.* (2015a) 'The reliability of the Greulich and Pyle atlas when applied to a Southern Turkish population', *European Journal of Dentistry*, 9(2), pp. 251–254. doi: 10.4103/1305-7456.156846.

Gungor, O. E. *et al.* (2015b) 'The reliability of the Greulich and Pyle atlas when applied to a Southern Turkish population', *European Journal of Dentistry*, 9(2), pp. 251–254. doi: 10.4103/1305-7456.156846.

Hackman, L. and Black, S. (2013) 'The Reliability of the Greulich and Pyle Atlas When Applied to a Modern Scottish Population', *Journal of Forensic Sciences*, 58(1), pp. 114–119. doi: 10.1111/j.1556-4029.2012.02294.x.

Haiter-Neto, F. *et al.* (2006) 'Skeletal age assessment: A comparison of 3 methods', *American Journal of Orthodontics and Dentofacial Orthopedics*, 130(4), p. 435.e15-435.e20. doi: 10.1016/j.ajodo.2006.03.023.

Hall, E. J. (2009) 'Radiation biology for pediatric radiologists', in *Pediatric Radiology*, pp. 57–64. doi: 10.1007/s00247-008-1027-2.

Hall, E. J. and Brenner, D. J. (2008) 'Cancer risks from diagnostic radiology', *British Journal of Radiology*. British Institute of Radiology, pp. 362–378. doi: 10.1259/bjr/01948454.

Hansman, C. F. and Maresh, M. M. (1961) 'A longitudinal study of skeletal

maturation', *Am J Dis Child*, 101, pp. 305–321. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=13711357.

Harper, R. and Reeves, B. (1999) 'Reporting of precision of estimates for diagnostic accuracy: a review.', *BMJ (Clinical research ed.)*. BMJ Publishing Group, 318(7194), pp. 1322–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10323817 (Accessed: 7 April 2019).

Hawley, N. L. *et al.* (2012a) 'Determinants of relative skeletal maturity in South African children', *Bone*, 50(1), pp. 259–264. doi: 10.1016/j.bone.2011.10.029.

Hawley, N. L. *et al.* (2012b) 'Determinants of relative skeletal maturity in South African children', *Bone*, 50(1), pp. 259–264. doi: 10.1016/j.bone.2011.10.029.

Heppe, D. H. M. *et al.* (2012) 'Bone age assessment by dual-energy X-ray absorptiometry in children: An alternative for X-ray?', *British Journal of Radiology*, 85(1010), pp. 114–120. doi: 10.1259/bjr/23858213.

Himes, J. H. (1984) 'An early hand-wrist atlas and its implications for secular change in bone age.', *Annals of human biology*, 11(1), pp. 71–75. doi: 10.1080/03014468400006911.

Hm Government (1989) *Children Act 1989*, *Time*. Statute Law Database. Available at: http://www.legislation.gov.uk/ukpga/1989/41/contents (Accessed: 15 October 2018).

Hm Government (2018) *Summary of latest statistics - GOV.UK*, *Gov.uk*. Available at: https://www.gov.uk/government/publications/immigration-statistics-year-ending-june-2018/summary-of-latest-statistics (Accessed: 12 October 2018).

Holderbaum, R. M. *et al.* (2005) 'Comparison among dental, skeletal and chronological development in HIV-positive children: a radiographic study', *Brazilian Oral Research*, 19(3), pp. 209–215. doi: 10.1590/S1806-83242005000300010.

Humphrey, L. T. (1998) 'Growth patterns in the modern human skeleton',

American Journal of Physical Anthropology. Wiley-Blackwell, 105(1), pp. 57–72. doi: 10.1002/(SICI)1096-8644(199801)105:1<57::AID-AJPA6>3.0.CO;2-A.

Van Jaarsveld, C. H. M. and Gulliford, M. C. (2015) 'Childhood obesity trends from primary care electronic health records in England between 1994 and 2013: Population-based cohort study', *Archives of Disease in Childhood*. BMJ Publishing Group, 100(3), pp. 214–219. doi: 10.1136/archdischild-2014-307151.

Jiménez-Castellanos, J. *et al.* (1996) 'Skeletal Maturation of Wrist and Hand Ossification Centers in Normal Spanish Boys and Girls: A Study Using the Greulich-Pyle Method', *Acta Anatomica*, 155(3), pp. 206–211. Available at: https://www.karger.com/Article/Pdf/147806.

Johnson, W. *et al.* (2012) 'Patterns of linear growth and skeletal maturation from birth to 18 years of age in overweight young adults', *International Journal of Obesity*. Nature Publishing Group, 36(4), pp. 535–541. doi: 10.1038/ijo.2011.238.

Johnston, Francis E. (2002) 'Social and economic influences on growth and secular trends', in *Human growth and development*. London: Academic Press., pp. 197–211.

Johnston, F. E. (1963) 'Skeletal age and its prediction in Philadelphia children', *Human biology*, 35(2), pp. 192–202. Available at: http://www.jstor.org/stable/41448600 (Accessed: 30 January 2017).

Johnston, F. E. and Jahina, S. B. (1965) 'The Contribution of the Carpal Bones to the Assessment of Skeletal Age', *American Journal of Physical Anthropology*, 23, pp. 349–354. Available at: https://onlinelibrary.wiley.com/doi/epdf/10.1002/ajpa.1330230408.

Khan, B. (2010) 'Biometrics and identity management for homeland security applications in Saudi Arabia', *Biometrics and identity management for homeland security applications in Saudi Arabia*, 4(15), p. 11. Available at: http://www.academicjournals.org/AJBM/PDF/pdf2010/4Nov/Khan et al.pdf.

Kim, J. R., Lee, Y. S. and Yu, J. (2015) '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 Journal of Radiology*, 16(1), pp. 201–205. doi: 10.3348/kjr.2015.16.1.201.

Koc, A. *et al.* (2001) 'Assessment of bone ages: Is the Greulich-Pyle method sufficient for Turkish boys?', *Pediatrics International*, 43(6), pp. 662–665. doi: 10.1046/j.1442-200X.2001.01470.x.

Krailassiri, S., Anuwongnukroh, N. and Dechkunakorn, S. (2002) 'Relationships between Dental Calcification Stages and Skeletal Maturity Indicators in Thai Individuals', *Angle Orthodontist*, 72(2), pp. 155–166. doi: 10.1043/0003-3219(2002)072.

Kullman, L. (1995) 'Acuracy of two dental and one skeletal age estimation methods in Swedish adolescents', *Forensic Sci Int*, 75(95), pp. 225–236. doi: https://doi.org/10.1016/0379-0738(95)01792-55.

Lequin, M. H. *et al.* (2000) 'Normal Values for Tibial Quantitative Ultrasonometry in Caucasian Children and Adolescents (Aged 6 to 19 Years)', *Calcif Tissue Int*, 67, pp. 101–105. doi: 10.1007/s00223001119.

Lewis, C. P., Lavy, C. B. D. and Harrison, W. J. (2002) 'Delay in skeletal maturity in Malawian children', *The Journal of Bone and Joint Surgery*, 84(5), pp. 732–734. doi: 10.1302/0301-620X.84B5.12642.

Loder, R. T. (1993) 'Applicability of the Greulich and Pyle Skeletal Age Standards to Black and White Children of Today', *Archives of Pediatrics & Adolescent Medicine*, 147(12), p. 1329. doi: 10.1001/archpedi.1993.02160360071022.

Maggio, A. *et al.* (2016) 'Assessment of the accuracy of the Greulich and Pyle hand-wrist atlas for age estimation in a contemporary Australian population', *Australian Journal of Forensic Sciences*. Taylor & Francis, 0618(February), pp. 1–11. doi: 10.1080/00450618.2016.1251970.

Mansourvar, M. *et al.* (2014) 'The applicability of Greulich and Pyle atlas to assess skeletal age for four ethnic groups', *Journal of Forensic and Legal Medicine*, 22, pp. 26–29. doi: 10.1016/j.jflm.2013.11.011.

Marini, J. C. *et al.* (2017) 'Osteogenesis imperfecta', *Nature Reviews Disease Primers*. Macmillan Publishers Limited, 3, pp. 1–19. doi: 10.1038/nrdp.2017.52.

Martin Bland, J. and Altman, D. (1986) 'Statistical methods for assessing agreement between two methods of clinical measurement.', *The Lancet*, 327(8476), pp. 307–310. Available at: http://www.sciencedirect.com/science/article/pii/S0140673686908378 (Accessed: 30 January 2017).

Martin, D. D. *et al.* (2011) 'Metacarpal thickness, width, length and medullary diameter in children-reference curves from the First Zürich Longitudinal Study', *Osteoporosis International*, 22(5), pp. 1525–1536. doi: 10.1007/s00198-010-1389-9.

Martin, D. D., Wit, J. M., *et al.* (2011) 'The use of bone age in clinical practice - part 1.', *Hormone research in paediatrics*. Karger Publishers, 76(1), pp. 1–9. doi: 10.1159/000329372.

Martin, D. D., Wit, J. M., *et al.* (2011) 'The Use of Bone Age in Clinical Practice – Part 2', *Hormone Research in Paediatrics*, 76(1), pp. 10–16. doi: 10.1159/000329374.

Menjívara, C. and Krista M. Perreirab (2017) 'Undocumented and unaccompanied: children of migration in the European Union and the United States', *Journal of Ethnic and Migration Studies*, 34(8), pp. 1378–1386. doi: 10.1080/1369183X.2017.1404255.

Mettler, F. A. *et al.* (2008) 'Effective Doses in Radiology and Diagnostic Nuclear Medicine: A Catalog', *Radiology*. Radiological Society of North America, 248(1), pp. 254–263. doi: 10.1148/radiol.2481071451.

Ministry of Health (2015) *National eHealth Strategy and Change Management Office (SCMO)*. Available at: https://www.moh.gov.sa/en/Ministry/ehealthstd/Documents/eHealth Standards Files/Admin - Patient Identification 1/UC0001 Saudi eHealth Patient Identification Interoperability Use Case v1.0.pdf (Accessed: 31 October 2018).

Mohammed, R. *et al.* (2015) 'Is Greulich and Pyle standards of skeletal maturation applicable for age estimation in South Indian Andhra children?', *Journal of Pharmacy and Bioallied Sciences*, 7(3), p. 218. doi: 10.4103/0975-7406.160031.

Mora, S. *et al.* (2001) 'Skeletal Age Determinations in Children of European and African Descent: Applicability of the Greulich and Pyle Standards', *Pediatric Research*, 50(5), pp. 624–628. doi: 10.1203/00006450-200111000-00015.

Moradi, M., Sirous, M. and Morovatti, P. (2012) 'The reliability of skeletal age determination in an Iranian sample using Greulich and Pyle method', *Forensic Science International*, 223(1–3), p. 372.e1-372.e4. doi: 10.1016/j.forsciint.2012.08.030.

Moseley, C. F. (1977) 'A straight-line graph for leg-length discrepancies.', *The Journal of bone and joint surgery. American volume*, 59(2), pp. 174–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/845200.

Mughal, A. M., Hassan, N. and Ahmed, A. (2014) 'The applicability of the Greulich & Pyle Atlas for bone age assessment in primary school-going children of Karachi, Pakistan', *Pakistan Journal of Medical Sciences*, 30(2), pp. 409–412. doi: 10.12669/pjms.302.4296.

Neelis, E. *et al.* (2017) 'Bone health of children with intestinal failure measured by dual energy X-ray absorptiometry and digital X-ray radiogrammetry', *Clinical Nutrition*. doi: 10.1016/j.clnu.2017.02.014.

Ng, M. *et al.* (2014) 'Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the Global Burden of Disease Study 2013', *The Lancet*, 384(9945), pp. 766–781. doi: 10.1016/S0140-6736(14)60460-8.

Nusman, C. M. *et al.* (2015) 'Bone health of patients with juvenile idiopathic arthritis: A comparison between dual-energy X-ray absorptiometry and digital X-ray radiogrammetry', *European Journal of Radiology*. Elsevier Ireland Ltd, 84(10), pp. 1999–2003. doi: 10.1016/j.ejrad.2015.06.015.

Oatway, W. B. *et al.* (2010) *Ionising Radiation Exposure of the UK*

Population: 2010 Review. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/518487/PHE-CRCE-026_-_V1-1.pdf (Accessed: 16 October 2018).

Ontell, F. K. *et al.* (1996) 'Bone age in children of diverse ethnicity.', *American Journal of Roentgenology*, 167(6), pp. 1395–1398. doi: 10.2214/ajr.167.6.8956565.

Öztürk, F. *et al.* (2016) 'Bone age assessment: comparison of children from two different regions with the Greulich–Pyle method In Turkey', *Australian Journal of Forensic Sciences*. Taylor & Francis, 48(6), pp. 694–703. doi: 10.1080/00450618.2015.1119311.

Patel, P. *et al.* (2015) 'Accuracy of two dental and one skeletal age estimation methods in 6-16 year old Gujarati children', *Journal of Forensic Dental Sciences*, 7(1), p. 18. doi: 10.4103/0975-1475.150298.

Pathmanathan, G. and Raghavan, P. (2006) 'Bone Age Based Linear Growth and Weight of Under Privileged North west Indian Children Compared with their Well-Off North west Indian Peers .', 55(2), pp. 34–42. Available at: http://medind.nic.in/jae/t06/i2/jaet06i2p34.pdf (Accessed: 16 October 2018).

Patil, S. T. *et al.* (2012) 'Applicability of Greulich and Pyle skeletal age standards to Indian children', *Forensic Science International*, 216(1–3), p. 200.e1-200.e4. doi: 10.1016/j.forsciint.2011.09.022.

Paxton, M. L., Lamont, A. C. and Stillwell, A. P. (2013) 'The reliability of the Greulich-Pyle method in bone age determination among Australian children: Bone age determination using Greulich-Pyle', *Journal of Medical Imaging and Radiation Oncology*, 57(1), pp. 21–24. doi: 10.1111/j.1754-9485.2012.02462.x.

Pinchi, V. *et al.* (2014) 'Skeletal age estimation for forensic purposes: A comparison of GP, TW2 and TW3 methods on an Italian sample', *Forensic Science International*. Elsevier, 238, pp. 83–90. doi: 10.1016/J.FORSCIINT.2014.02.030.

Piraino, D. W. *et al.* (1999) 'Selenium-based digital radiography versus

conventional film-screen radiography of the hands and feet: A subjective comparison', *American Journal of Roentgenology*, 172(1), pp. 177–184. doi: 10.2214/ajr.172.1.9888764.

Pludowski, P., Lebiedowski, M. and Lorenc, R. S. (2004) 'Evaluation of the possibility to assess bone age on the basis of DXA derived hand scans?preliminary results', *Osteoporosis International*. Springer-Verlag, 15(4), pp. 317–322. doi: 10.1007/s00198-003-1545-6.

Rai, V. (2014) 'Dental and Skeletal Maturity- A Biological Indicator of Chronologic Age', *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH*, 8(9), pp. ZC60-ZC64. doi: 10.7860/JCDR/2014/10079.4862.

Ramsthaler, F. *et al.* (2009) 'How reliable are the risk estimates for X-ray examinations in forensic age estimations? A safety update', *International Journal of Legal Medicine*, pp. 199–204. doi: 10.1007/s00414-009-0322-2.

Van Rijn, R. R. *et al.* (2001) 'Is the Greulich and Pyle atlas still valid for Dutch Caucasian children today?', *Pediatric Radiology*, 31(10), pp. 748–752. doi: 10.1007/s002470100531.

van Rijn, R. R., Lequin, M. H. and Thodberg, H. H. (2009) 'Automatic determination of Greulich and Pyle bone age in healthy Dutch children', *Pediatric Radiology*, 39(6), pp. 591–597. doi: 10.1007/s00247-008-1090-8.

Ritz-Timme, S. *et al.* (2000) 'Age estimation: The state of the art in relation to the specific demands of forensic practise', *International Journal of Legal Medicine*, pp. 129–136. doi: 10.1007/s004140050283.

Roche, A. F., Davila, G. H. and Eyman, S. L. (1971) 'A Comparison Between Greulich-Pyle and Tanner-Whitehouse Assessments of Skeletal Maturity', *Radiology*, 98(2), pp. 273–280. doi: 10.1148/98.2.273.

Romann, M. and Fuchslocher, J. (2016) 'Assessment of skeletal age on the basis of DXA-derived hand scans in elite youth soccer', *Research in Sports Medicine*. Routledge, 24(3), pp. 200–211. doi: 10.1080/15438627.2016.1191490.

Rosholm, A. *et al.* (2001) 'Estimation of bone mineral density by digital x-ray radiogrammetry: Theoretical background and clinical testing', *Osteoporosis*

International, 12(11), pp. 961–969. doi: 10.1007/s001980170026.

Royal College of Paediatrics and Child Health (2018) *Refugee and unaccompanied asylum seeking children and young people*. London.

De Sanctis, V. *et al.* (2014) 'Are the new automated methods for bone age estimation advantageous over the manual approaches?', *Pediatric Endocrinology Reviews*, pp. 201–205. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25581985 (Accessed: 13 March 2018).

Santoro, V. *et al.* (2012) 'Applicability of Greulich and Pyle and Demirijan aging methods to a sample of Italian population', *Forensic Science International*, 221(1–3), p. 153.e1-153.e5. doi: 10.1016/j.forsciint.2012.04.014.

Santos, C. *et al.* (2011) 'Comparative study of Greulich and Pyle Atlas and Maturos 4.0 program for age estimation in a Portuguese sample', *Forensic Science International*, 212(1–3), p. 276.e1-276.e7. doi: 10.1016/j.forsciint.2011.05.032.

Schmeling, A. *et al.* (2000) 'Effects of ethnicity on skeletal maturation: Consequences for forensic age estimations', *International Journal of Legal Medicine*, pp. 253–258. doi: 10.1007/s004149900102.

Schmeling, A. *et al.* (2006) 'The impact of economic progress and modernization in medicine on the ossification of hand and wrist', *International Journal of Legal Medicine*. Springer-Verlag, 120(2), pp. 121–126. doi: 10.1007/s00414-005-0007-4.

Schmeling, A. *et al.* (2008) 'Criteria for age estimation in living individuals', *International Journal of Legal Medicine*, 122(6), pp. 457–460. doi: 10.1007/s00414-008-0254-2.

Schmeling, A. *et al.* (2008) 'Criteria for age estimation in living individuals', *International Journal of Legal Medicine*, 122(6), pp. 457–460. doi: 10.1007/s00414-008-0254-2.

Schmidt, S. *et al.* (2007a) 'Comparative analysis of the applicability of the skeletal age determination methods of Greulich-Pyle and Thiemann-Nitz for forensic age estimation in living subjects', *International Journal of Legal*

Medicine, 121(4), pp. 293–296. doi: 10.1007/s00414-007-0165-7.

Schmidt, S. *et al.* (2007b) 'Comparative analysis of the applicability of the skeletal age determination methods of Greulich–Pyle and Thiemann–Nitz for forensic age estimation in living subjects', *International Journal of Legal Medicine*, 121(4), pp. 293–296. doi: 10.1007/s00414-007-0165-7.

Schmidt, S. *et al.* (2008) 'Applicability of the skeletal age determination method of Tanner and Whitehouse for forensic age diagnostics', *International Journal of Legal Medicine*, 122(4), pp. 309–314. doi: 10.1007/s00414-008-0237-3.

Schündeln, M. M. *et al.* (2016) 'A Piece of the Puzzle: The Bone Health Index of the BoneXpert Software Reflects Cortical Bone Mineral Density in Pediatric and Adolescent Patients', *PloS one*, 11(3), p. e0151936. doi: 10.1371/journal.pone.0151936.

Senior, P. A. and Bhopal, R. (1994) 'Ethnicity as a variable in epidemiological research', *BMJ*. British Medical Journal Publishing Group, 309(6950), p. 327. doi: 10.1136/bmj.309.6950.327.

Serinelli, S. *et al.* (2011) 'Accuracy of three age determination X-ray methods on the left hand-wrist: A systematic review and meta-analysis', *Legal Medicine*, 13(3), pp. 120–133. doi: 10.1016/j.legalmed.2011.01.004.

Shilpa, P. *et al.* (2013) 'Estimation and comparison of dental, skeletal and chronologic age in Bangalore south school going children', *Journal of Indian Society of Pedodontics and Preventive Dentistry*, 31(2), p. 63. doi: 10.4103/0970-4388.115696.

Simpson, L. and Akinwale, B. (2007) 'Quantifying Stability and Change in Ethnic Group', *Journal of Official Statistics*, 23(2), pp. 185–208.

Šimundić, A.-M. (2009) 'Measures of Diagnostic Accuracy: Basic Definitions.', *EJIFCC*. International Federation of Clinical Chemistry and Laboratory Medicine, 19(4), pp. 203–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27683318 (Accessed: 7 April 2019).

So, L. L. and Yen, P. K. (1990) 'Secular trend in skeletal maturation in southern Chinese girls in Hong Kong.', *Zeitschrift fur Morphologie und*

Anthropologie, 78(2), pp. 145–153. Available at: http://www.jstor.org/stable/25757279 (Accessed: 30 January 2017).

Soudack, M. *et al.* (2012) 'Bone age in the 21st century: UIs Greulich and Pyle's atlas accurate for Israeli children?', *Pediatric Radiology*, pp. 343–348. doi: 10.1007/s00247-011-2302-1.

Speiser, P. W. *et al.* (2010) 'Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: An Endocrine Society clinical practice guideline', *Journal of Clinical Endocrinology and Metabolism*. Oxford University Press, pp. 4133–4160. doi: 10.1210/jc.2009-2631.

Suri, S. *et al.* (2013) 'Longitudinal comparison of skeletal age determined by the Greulich and Pyle method and chronologic age in normally growing children, and clinical interpretations for orthodontics', *American Journal of Orthodontics and Dentofacial Orthopedics*, 143(1), pp. 50–60. doi: 10.1016/j.ajodo.2012.08.027.

Tanner, J. M., Whitehouse, R. H., Marshall, W. A., Healy, M. J. R. & Goldstein, H. (1975) *Assessment of Skeletal Maturity and Prediction of Adult Height*. London: Academic Press.

Tanner, J. *et al.* (2001) *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method)*. London: WB Saunders.

Thodberg, H. H. (2009) 'Clinical review: An automated method for determination of bone age.', *The Journal of clinical endocrinology and metabolism*. Oxford University Press, 94(February), pp. 2239–2244. doi: 10.1210/jc.2008-2474.

Thodberg, H. H. *et al.* (2009) 'The BoneXpert Method for Automated Determination of Skeletal Maturity', *IEEE Transactions on Medical Imaging*, 28(1), pp. 52–66. doi: 10.1109/TMI.2008.926067.

Thodberg, H. H. *et al.* (2010) 'A paediatric bone index derived by automated radiogrammetry', *Osteoporosis International*, 21(8), pp. 1391–1400. doi: 10.1007/s00198-009-1085-9.

Thodberg, H. H. *et al.* (2017) 'Automated determination of bone age from hand X-rays at the end of puberty and its applicability for age estimation',

International Journal of Legal Medicine. Springer Berlin Heidelberg, 131(3), pp. 771–780. doi: 10.1007/s00414-016-1471-8.

Thodberg, H. H. and Sävendahl, L. (2010) 'Validation and reference values of automated bone age determination for four ethnicities', *Academic Radiology*, 17(11), pp. 1425–1432. doi: 10.1016/j.acra.2010.06.007.

Uffmann, M. and Schaefer-Prokop, C. (2009) 'Digital radiography: The balance between image quality and required radiation dose', *European Journal of Radiology*. Elsevier, 72(2), pp. 202–208. doi: 10.1016/j.ejrad.2009.05.060.

Ulijaszek, S. J. (2006) 'The International Growth Standard for Children and Adolescents Project: Environmental influences on preadolescent and adolescent growth in weight and height', *Food and Nutrition Bulletin*, pp. 279–294. doi: 10.1177/15648265060274S510.

UNICEF (2018) *levels and trends in child malnutrition*, *WHO / World Bank Group Joint Child Malnutrition Estimates*. Washington DC. Available at: https://data.unicef.org/wp-content/uploads/2018/05/JME-2018-brochure-web.pdf (Accessed: 16 October 2018).

United Nation (2017) *2017 International Migration Report*. New York. doi: ST/ESA/SER.A/404.

Virtama, P. and Mahone, H. (1960) 'Thickness of the cortical layer as an estimate of mineral content of human finger bones.', *The British journal of radiology*, 33(385), pp. 60–62. doi: 10.1259/0007-1285-33-385-60.

Wenzel, A., Droschl, H. and Melsen, B. (1984) 'Skeletal maturity in Austrian children assessed by the GP and the TW-2 methods.', *Annals of human biology*, 11(2), pp. 173–7. doi: 10.1080/03014468400007041.

Y So, L. L. and Lisa So, B. L. (1991) 'Correlation of skeletal maturation with stature and body weight of Southern Chinese girls Correlation of skeletal maturation with stature and body weight of Southern Chinese girls in Hong Kong With 3 tables in the text', *E. Schweizerbart'sche Verlagsbuchhandlung*, 78(3), pp. 307–312. Available at: http://www.jstor.org/stable/25757319.

Zabet, D. *et al.* (2014a) 'Can the Greulich and Pyle method be used on

French contemporary individuals?', *International Journal of Legal Medicine*, 129(1), pp. 171–177. doi: 10.1007/s00414-014-1028-7.

Zabet, D. *et al.* (2014b) 'Can the Greulich and Pyle method be used on French contemporary individuals?', *International Journal of Legal Medicine*, 129(1), pp. 171–177. doi: 10.1007/s00414-014-1028-7.

Zafar, A. M. *et al.* (2010) 'An appraisal of greulich-pyle atlas for skeletal age assessment in Pakistan', *Journal of the Pakistan Medical Association*, 60(7), pp. 552–555. Available at: http://www.jpma.org.pk/PdfDownload/2168.pdf (Accessed: 30 January 2017).

Zaidi, S. *et al.* (2017) 'Expanding access to healthcare in South Asia', *Bmj*, p. j1645. doi: 10.1136/bmj.j1645.

Zhang, A., Sayre, J. W., *et al.* (2009) 'Racial Differences in Growth Patterns of Children Assessed on the Basis of Bone Age', *Radiology*, 250(1), pp. 228–235. doi: 10.1148/radiol.2493080468.

Zhang, A., Sayre, J. W., *et al.* (2009) 'Racial Differences in Growth Patterns of Children Assessed on the Basis of Bone Age', *Radiology*, 250(1), pp. 228–235. doi: 10.1148/radiol.2493080468.

Zhang, S.-Y. *et al.* (2009) 'Standards of TW3 skeletal maturity for Chinese children.', *Annals of human biology*, 35(3), pp. 349–54. doi: 10.1080/03014460801953781.

Appendices

Appendix I: Maximum delay and advancement in BA compared to CA using the G&P atlas on a yearly basis as reported in the literature (Caucasian females)

Age group (years)	Range of differences in mean BA-CA (in years)	References
7	-0.40 to 0.20	(Cantekin et al., 2012, Wenzel et al., 1984)
8	-0.59 to -0.20	(Andersen, 1971, Johnston, 1963)
9	-0.97 to 0.47	(Andersen, 1971, Hackman and Black, 2013)
10	-0.47 to 0.40	(Andersen, 1971, Zabet et al., 2015)
11	-0.63 to 0.58	(Andersen, 1971, Bueken et al., 2009)
12	-0.39 to 0.57	(Andersen, 1971, Bueken et al., 2009)
13	-0.19 to 0.75	(Andersen, 1971, Bueken et al., 2007)
14	-0.25 to 1.40	(Andersen, 1971, Zabet et al., 2015)
15	-0.32 to 1.20	(Andersen, 1971, Zabet et al., 2015)
16	0.95	(Bueken et al., 2007)
17	-0.65 to 0.58	(Hackman and Black, 2013, Johnston, 1963)
18	-0.90	(Hackman and Black, 2013)

*A positive value indicates that the child's bone age (BA) exceeds his/her chronological age (CA) while a negative value indicates a delay in BA compared to CA

Appendix II: Maximum delay and advancement in BA compared to CA using the G&P atlas on a yearly basis as reported in the literature (Caucasian males)

Age group (years)	Range of differences in mean BA-CA (in years)*	References
7	-0.70 to 0.20	(Cantekin et al., 2012, Wenzel et al., 1984)
8	-0.85 to 0.15	(Andersen, 1971, Johnston, 1963)
9	-0.54 to 0.30	(Koc et al., 2001, Johnston, 1963)
10	-0.43 to 0.58	(Andersen, 1971, Johnston, 1963)
11	-0.45 to 0.65	(Andersen, 1971, Johnston, 1963)
12	-0.27 to 0.59	(Andersen, 1971, Johnston, 1963)
13	-0.70 to 0.45	(Wenzel et al., 1984, Johnston, 1963)
14	-0.70 to 0.50	(Wenzel et al., 1984, Zabet et al., 2015)
15	-1.3 to 1.3	(Wenzel et al., 1984, Zabet et al., 2015)
16	-0.66 to 0.98	(Andersen, 1971, Bueken et al., 2007)
17	-0.02 to 0.95	(Bueken et al., 2007, Cantekin et al., 2012)
18	-0.02 to 0.60	(Suri et al., 2013, Bueken et al., 2007)

*A positive value indicates that the child's bone age (BA) exceeds his/her chronological age (CA) while a negative value indicates a delay in BA compared to CA

Appendix III: Maximum delay and advancement in BA compared to CA using the G&P atlas on a yearly basis as reported in the literature (Asian females)

Age group (years)	Average Mean BA-CA (range of differences)*	References
6	-0.07 to -0.42	(Chiang et al 2005, Patil et al 2015)
7	-0.47 to 0.22	(Chiang et al 2005, Griffith et al 2007)
8	-0.84 to 0.11	(Chiang et al 2005, Chiang et al 2005)
9	-0.60 to 0.52	(Patil et al., 2012, Mohammed et al., 2015)
10	-1 to 0.23	(Mohammed et al., 2015, Kim et al., 2015)
11	-0.79 to 0	(Kim et al., 2015, Patel et al., 2015)
12	-0.87 to 0.22	(Patil et al., 2012, Patel et al., 2015)
13	-0.7	(Mohammed et al., 2015)
14	-0.51	(Mohammed et al., 2015)
15	-1.21	(Mohammed et al., 2015)
16	-0.50 to 0.29	(Patel et al., 2015, Mohammed et al., 2015)
17	-0.01 to 0.51	(Griffith et al 2007, Mohammed et al 2015)

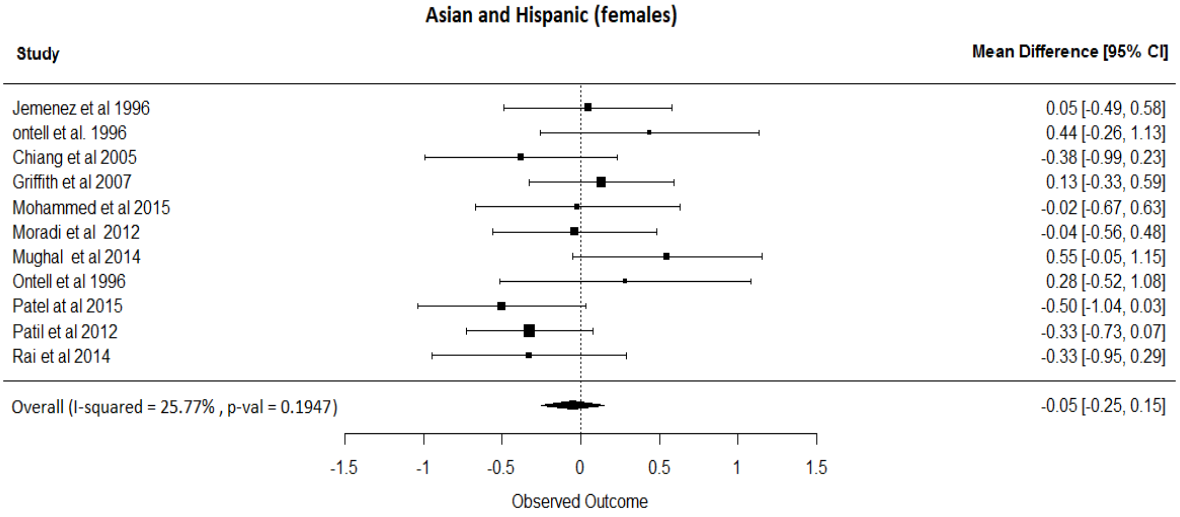
*A positive value indicates that the child's bone age (BA) exceeds his/her chronological age (CA) while a negative value indicates a delay in BA compared to CA

Appendix IV: Maximum delay and advancement in BA compared to CA using the G&P atlas on a yearly basis as reported in the literature (Asian males)

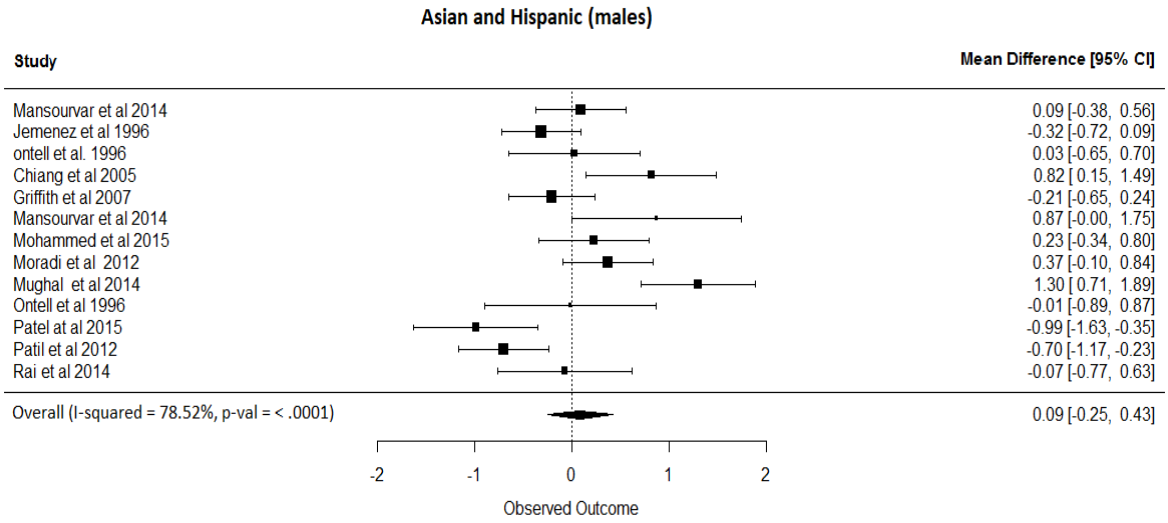
Age group (years)	Average Mean BA-CA (range of differences)*	Reference
6	-1.47 to -1	(Chiang et al 2005, Patel et al 2012)
7	-1.9 to -0.9	(Patil et al 2012, Griffith et al 2007)
8	-2.11 to -0.27	(Patil et al 2012, Patel at al 2015)
9	-1.71 to 0.32	(Patil et al., 2012, Mohammed et al., 2015)
10	-1.11	(Patil et al., 2012)
11	-1.11	(Patil et al., 2012)
12	-1.46 to 0.12	(Patil et al., 2012)
13	-1.39 to 0.45	(Patel et al., 2015)
14	-1.75 to 0.19	(Patel et al., 2015)
15	-1.08 to 0.58	(Patel et al., 2015)
16	-0.68 to 1.21	(Patil et al., 2012, Mohammed et al., 2015)
17	0.22 to 0.82	(Chaing et al, 2005, Griffiths et al., 2007)

*A positive value indicates that the child's bone age (BA) exceeds his/her chronological age (CA) while a negative value indicates a delay in BA compared to CA

Appendix V: Forest plot of Asians-Hispanics for females showing no significant results



Appendix VI: Forest plot of Asians-Hispanics for males showing no significant results



Appendix VII: permission for the Copyright holder (Springer-Nature) to reuse the paper.

SPRINGER NATURE

Title:	Estimating bone mass in children: can bone health index replace dual energy x-ray absorptiometry?
Author:	Khalaf Alshamrani, Fabrizio Messina, Nick Bishop et al
Publication:	Pediatric Radiology
Publisher:	Springer Nature
Date:	Jan 1, 2018

Appendix III: Ethical approval obtained from The Health Research Authority Yorkshire and Humber (United Kingdom)

NHS
Health Research Authority

Dr Amaka Offiah
Academic Unit of Child Health
Sheffield Children's NHS Foundation Trust
Western Bank, Sheffield
S10 2TH

Email: hra.approval@nhs.net

30 August 2016

Dear Dr Offiah

Letter of HRA Approval

Study title:	**Automatic determination of bone age and bone mass in a modern UK paediatric cohort**
IRAS project ID:	**207437**
Sponsor	**Sheffield Children's Hospital NHS Foundation Trust**

I am pleased to confirm that **HRA Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details

Appendix IX: Ethical approval obtained from King Fahad Hospital (Saudi Arabia) to collect left hand radiographs

المملكة العربية السعودية
وزارة الصحة
مستشفى الملك فهد بجدة

الرقم: /١٧/٢/ ح
التاريخ: / / ١٤هـ
المشفوعات:

To Whom it May Concern

We have no objection to the collection of data from King Fahd Hospital at which will be carried out by Mr. Khalaf AlShamrani, who is a PhD student at The University of Sheffield. Khalaf is permitted to collect data from Radiology Department retrospectively and there will be no contact with patients or their relatives.
Please do not hesitate to contact me should you required any further enquiries.

Regards,

Dr. Iyad Feteih MBBCH, FRCPC

Head of Radiology Department
King Fahd Hospital - Jeddah

Dr. Iyad M Feteih
Head of Radiology
1100075

المملكة العربية السعودية - هاتف ٦٤٧٠٠٠٠ فاكس ٦٤٧٠٩٤٢ ص.ب ٨٤٨٨ - جدة ٢١٤٩٦

European Radiology
https://doi.org/10.1007/s00330-018-5792-5

FORENSIC MEDICINE

Is the Greulich and Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis

Khalaf Alshamrani[1,2,3] &Fabrizio Messina[4,5] &Amaka C. Offiah[1,3]

Received: 4 June 2018 / Revised: 6 September 2018 / Accepted: 21 September 2018

Abstract

Objective To determine whether the Greulich and Pyle (G&P) atlas is applicable when applied to populations of different ethnicity.

Methods A systematic review of studies published between 1959 and 15th February 2017 identified from the Embase, MEDLINE and Cochrane databases was undertaken. Quality of the studies was assessed using the National Institute for Health and Care Excellence tool. Meta-analysis used mean differences and standard deviations as summary statistics for the difference between bone age (BA) and chronological age (CA).

Results A total of 49 studies were included of which 27 (55%) were related to Caucasian populations. Of the 49 eligible studies, 35 were appropriate for further meta-analysis. In African females, meta-analysis showed a significant mean difference between BA and CA of 0.37 years (95% CI 0.04, 0.69). In Asian males, meta-analysis showed significant differences between BA and CA of -1.08, -1.35, -1.07, -0.80 and 0.50 years for chronological ages of 6, 7, 8, 9 and 17 years, respectively. Meta-analysis showed no significant differences between BA and CA in African males, Asian females, Caucasians and Hispanics.

Conclusions The G&P standard is imprecise and should be used with caution when applied to Asian male and African female populations, particularly when aiming to determine chronological age for forensic/legal purposes.

Key Points

- In African females, bone age is significantly advanced when compared to the G&P standard.
- In Asian males, bone age is significantly delayed between 6 and 9 years old inclusive and significantly advanced at 17 years old when compared to the G&P standard.
- The G&P atlas should be used with caution when applied to Asian and African populations, particularly when aiming to determine chronological age for forensic/legal purposes.

Keywords Age determination by skeleton · Forensic medicine · X-rays · Meta-analysis

Abbreviations

BA	Bone age
CA	Chronological age
EV	External validity
G&P	Greulich and Pyle
IV	Internal validity

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-018-5792-5) contains supplementary material, which is available to authorized users.

* Khalaf Alshamrani
kalshamrani1@sheffield.ac.uk

[1] Department of Oncology & Metabolism, University of Sheffield, Sheffield, UK

[2] College of Applied Medical Sciences, Najran University, Najran, Saudi Arabia

[3] Academic Unit of Child Health, Damer Street Building, Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield S10 2TH, UK

[4] School of Health and Related Research, University of Sheffield, Sheffield, UK

[5] Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK

Published online: 07 January 2019

NICE The National Institute for Health and Care Excellence
TW Tanner and Whitehouse

Introduction

Determining maturity and understanding growth in a child is critical for medical and psychosocial purposes. Assessing bone age is important to investigate whether the maturity of bones is occurring at the same rate as the chronological ageing process. Furthermore, bone age assessment has a role in forensic and legal investigations when the individual's chronological age is in doubt. For example, in asylum seekers and unaccompanied minors without valid documents to prove their ages [1], it is important to assess bone age using a reliable and suitable method [2]. Incorrectly assessing a child as an adult leaves the child with limited access to education, healthcare and other support provided to children.

There are two approaches widely used to determine bone age from a left hand radiograph: the Greulich and Pyle (G&P) and Tanner and Whitehouse (TW) methods [3, 4]. The population which formed the G&P standard atlas were North American Caucasians of good socioeconomic status. The assessment process is typically based on comparing a hand-wrist radiograph of a child with the age-matched standard radiographs as contained in the atlas. The G&P method depends on comparing the overall maturational status and is known to be straightforward and quick, therefore widely used. In contrast, the TW method depends on assessing and scoring the skeletal maturity of each individual bone of the hand, hence taking a longer time than the G&P method. Since the establishment of the G&P atlas, many studies have been conducted in different parts of the world to determine whether it is applicable to different populations. This question is important, particularly given the increasing legal and illegal influx of immigrants to certain parts of Europe. This systematic review and meta-analysis aims to provide a better understanding of the applicability of the G&P atlas to children and adolescents who are of a different population from the original standard.

Materials and methods

Search strategy

A systematic search of the MEDLINE, Embase and Cochrane databases was conducted. We searched MEDLINE using keywords ((Greulich and Pyle)) OR Greulich Pyle, ((bone age assessment OR bone age determination)) AND left hand and refined the search to include articles in English published between 1st January 1959 and 15th February 2017. No free text was used in this search. For Embase, we used the term (Greulich and Pyle) and refined the search to include articles in English published between 1st January 1959 and 15th February 2017. We also searched the Cochrane library using the keywords (Greulich and Pyle) and the MeSH term (Age Determination by Skeleton). The search was refined to include articles in English published between 1st January 1959 and 15th February 2017. Each study's title and abstract was screened to determine whether it presented data correlating bone age assessed by the G&P with chronological age. The full text was retrieved when the reviewers could not decide on the study's eligibility from the title and abstract alone. The following exclusion criteria were then applied:

1. Health status of participants could not be confirmed from the article or participants with developmental disorders or subjected to nutritional supplementation (these represent unhealthy children expected to show delayed or advanced bone age).
2. Using a modified method of G&P and/or using modalities other than conventional radiography
3. Full text not available within the resources available to the reviewers
4. Full text not in English
5. Review articles
6. When the mean difference between bone age (BA) and chronological age (CA) was not reported or could not be calculated by the reviewers based on the study results presented.

The search was independently carried out by two reviewers (KA and ACO), followed by a consensus meeting to agree the final selection of studies for inclusion in this review.

Quality assessment

Two reviewers KA and ACO independently assessed the quality of included studies using the tool developed by the National Institute for Health and Care Excellence (NICE, Appendix G) [5]. Discrepancies were resolved by discussion. The tool considers five aspects of a study: population, method of participant selection, outcomes, analysis and generalisability of the study. Then, an overall study quality grading is given to each study for internal validity (IV) and a separate grading for external validity (EV) as follows:

- ++ All/most of the checklist criteria have been fulfilled and the conclusions are unlikely to alter.

- + Some of the checklist criteria have been fulfilled; the conclusions are unlikely to alter even when they have not been fulfilled.
- − Few or no checklist criteria have been fulfilled, and the conclusions are likely or very likely to alter.

Data extraction

A single reviewer (K.A.) extracted and recorded the following data from eligible studies:

1. Sample size (males and females)
2. Ethnicity or country of origin
3. Mean difference and standard deviation (SD) between bone age and chronological age (BA-CA)
4. Mean and SD of bone age
5. Mean and SD of chronological age
6. Authors' conclusions
7. Applicability of the standard

Given the review question, studies were divided into four groups based on major ethnic groups: African, Asian, Caucasian and Hispanic. Data for each major ethnic group were summarised and analysed separately. Some studies reported the place/country from which participants were recruited, and in such cases, the study was grouped under the major ethnicity of that country. The mean differences between BA and CA are to be interpreted as follows: a positive value indicates that the child's bone age exceeds the child's chronological age and a negative value indicates delayed bone age compared to chronological age.

Additionally, we defined four categories to reflect the applicability of the G&P standard to the studied population as follows: (a) applicable, (b) not applicable (determined by the authors' use of words identical or similar to "applicable" or "not applicable", respectively, in the study's discussion or conclusion), (c) needs some modification (authors use phrases such as, "can be used with caution" or when the standard was found to be applicable to a certain age group but not others) and (d) not clear (when the study failed to mention whether the standard was applicable, not applicable or needed modification).

Statistical analysis

A combination of random effect meta-analyses by ethnicity (African, Asian, Caucasian and Hispanic) and sex was conducted using R Software [6]. Overall meta-analysis of all ethnicities was also determined. Additionally, meta-regression with covariates analysis (including sex and ethnicity as explanatory variables) was determined. Yearly interval sub-analysis of Asians aged 6 to 17 years and Caucasians aged 10 to 17 years was carried out in males and females. Other ethnicities were excluded from interval sub-analysis as the age groups were not constant between studies.

In total, 50 meta-analyses were performed using mean differences and standard deviations as summary statistics for the difference between bone age and chronological age. When a study examined more than one ethnicity, each ethnicity was treated as a separate study (only for the meta-analysis). Heterogeneity was assessed between 0 (no heterogeneity) and 100% (maximum heterogeneity) using the I-squared statistic. A funnel plot was determined to assess bias or the present of any systematic heterogeneity.

Results

This systematic review identified 907 studies of which 45 were eligible for inclusion (Fig. 1). Four additional studies were identified from the reference lists of the initial 45 extracted papers; therefore, the total number of included studies was 49 [7–55], of which 27 (55%) were related to Caucasian populations. The total number of children in the included studies was 21,081 (11,445 boys), comprising 11,194 Caucasians (5922 boys), 6776 Asians (3731 boys), 1705 Africans (1073 boys) and 1406 Hispanics (781 boys). As summarised in Table 1, there was minimal risk of bias for internal validity alone in one study [33], for external validity alone in five studies [17, 18, 25, 40, 50] and for both internal and external validity in 12 studies [11, 20–22, 27, 30, 35, 42, 43, 45, 51, 54]. There was significant risk of bias for internal validity alone in 0 studies, for external validity alone in two studies [8, 46] and for both internal and external validity in 2 studies [23, 29]. Sources of bias in these four studies requiring that their results be interpreted with caution include:

1. Absent documentation of statistical criteria such as p-values and/or observer reliability [8, 23]
2. Insufficient detail about the source of the study population [29]
3. Non-representative samples [46]

Studies included in this systematic review reported the mean difference between bone age and chronological age in different forms. Twenty-nine studies (60%) [8–14, 18, 19, 22, 24, 26–28, 30, 35, 36, 38, 41, 42, 44, 45, 48–54] presented the mean difference for each year of age for each sex. In such cases, the maximum delay and advancement in bone age was extracted. Twelve studies [15, 17, 20, 25, 31, 33, 34, 37, 39, 40, 46, 47] divided their sample into subgroups, where each subgroup contains up to five age groups, e.g. children aged between 1 and 5 years old. For each subgroup, the overall mean difference for each sex is reported. Eight studies [7, 16, 21, 23, 29, 32, 43, 55] only reported the overall mean

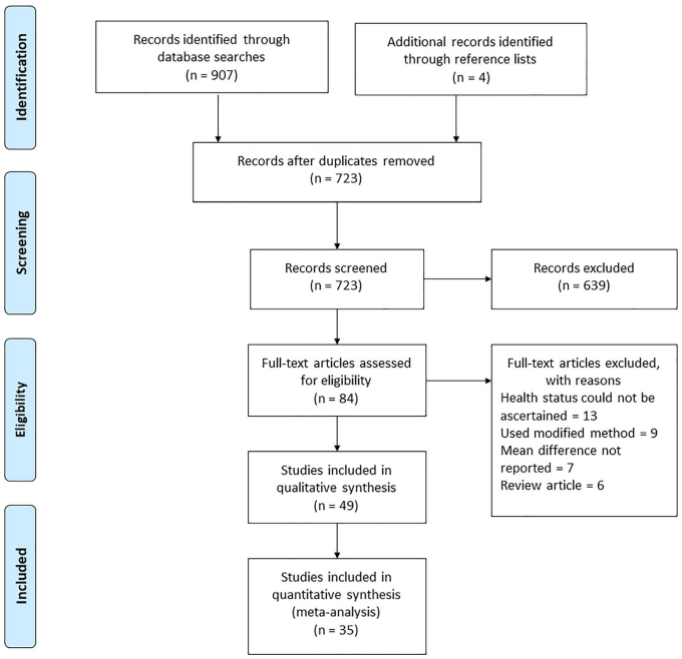


Fig. 1 Flow chart to show article selection process

difference between bone age and chronological age, limiting the applicability of their results to individual age groups. Data relating to ethnicity or country of origin, sample size, mean BA-CA and the authors' conclusions are summarised for each study in Tables 2, 3, 4, and 5.

Meta-analysis based on ethnicity

1. Caucasian females: Fifteen studies were included in the meta-analysis. These 15 studies presented moderate heterogeneity (*I*-squared 76%, Fig. 2) but did not show any statistically significant results, with overall mean difference BA-CA of 0.13 years (95% CI -0.17, 0.43).
2. Caucasian males: Seventeen studies were included in the meta-analysis. These 17 studies presented low heterogeneity (*I*-squared 22%, Fig. 2) and did not show any statistically significant results, with an overall mean difference BA-CA of -0.10 years (95% CI, -0.24, 0.04).
3. African females: Only three studies were included in the meta-analysis. The three studies were homogeneous (*I*-squared 0%, Fig. 3) and showed statistically significant results, with overall mean difference BA-CA of 0.37 years (95% CI 0.04, 0.69).
4. African males: Only five studies were included in the meta-analysis. The five studies presented moderate heterogeneity (*I*-squared 78%, Fig. 3) but did not show any statistically significant results, with overall mean difference BA-CA of 0.62 years (95% CI -0.01, 1.26).
5. Asian females: Only nine studies were included in the meta-analysis. These nine studies presented low to moderate heterogeneity (*I*-squared 27%, Fig. 4) but did not show any statistically significant results, with overall mean difference BA-CA of -0.10 years (95% CI -0.32, 0.12).
6. Asian males: Ten studies were included in the meta-analysis. The studies were highly heterogeneous (*I*-squared 82%, Fig. 4) but did not show any statistically significant results, with overall mean difference BA-CA of 0.15 years (95% CI -0.30, 0.59).
7. Hispanic females: Only two studies were included in the meta-analysis. The two studies presented no heterogeneity (*I*-squared 0%, Fig. 5) and did not show any statistically significant results, with overall mean difference BA-CA of 0.19 years (95% CI -0.23, 0.61).
8. Hispanic males: Only three studies were included in the meta-analysis. The three studies presented low heterogeneity (I-squared 11%, Fig. 5) but did not show any

Table 1 Quality assessment of the included studies (after agreement between the two assessors)

Study [Reference]	IV*	EV**	Study	IV*	EV**
Demish and Wartmann 1956 [7]	+	+	Calfee et al 2010 [32]	+	+
Hansman and Maresh 1961 [8]	+	-	Zafar et al 2010 [33]	++	+
Johnston 1963 [9]	+	+	Santos et al 2011 [34]	+	+
Andersen 1971 [10]	+	+	Cantekin et al 2012 [35]	++	++
Roche et al 1971 [11]	++	++	Dembetembe and Morris 2012 [36]	+	+
Wenzel et al 1984 [12]	+	+	Moradi et al 2012 [37]	+	+
So and Yen 1990 [13]	+	+	Patil et al 2012 [38]	+	+
So 1991 [14]	+	+	Santoro et al 2012 [39]	+	+
Loder et al 1993 [15]	+	+	Soudack et al 2012 [40]	+	++
Kullman 1995 [16]	+	+	Suri et al 2013 [41]	+	+
Ontell et al 1996 [17]	+	++	Hackman and Black 2013 [42]	++	++
Jiménez-Castellanos et al 1996 [18]	+	++	Paxton et al 2013 [43]	++	++
Koc et al 2001 [19]	+	+	Shilpa et al 2013 [44]	+	+
Mora et al 2001 [20]	++	++	Awais et al 2014 [45]	++	++
Van Rijn et al 2001 [21]	++	++	Rai et al 2014 [46]	+	-
Krailassiri et al 2002 [22]	++	++	Mansourvar et al 2014 [47]	+	+
Lewis et al 2002 [23]	-	-	Mughal et al 2014 [48]	+	+
Chiang et al 2005 [24]	+	+	Gungor et al 2015 [49]	+	+
Garamendi et al 2005 [25]	+	++	Kim et al 2015 [50]	+	++
Haiter-Neto et al 2006 [26]	+	+	Mohammed et al 2015 [51]	++	++
Büken et al 2007 [27]	++	++	Öztürk et al 2015 [52]	+	+
Griffith et al 2007 [28]	+	+	Patel et al 2015 [53]	+	+
Schmidt et al 2007 [29]	-	-	Zabet et al 2014 [54]	++	++
Büken et al 2009 [30]	++	++	Maggio et al 2016 [55]	+	+
Zhang et al 2009 [31]	+	+			

*IV: internal validity, **EV: external validity

++ Indicates that the study has been designed or conducted in such a way as to minimise the risk of bias

+ Indicates that either the answer to the checklist question is not clear from the way the study is reported, or that the study may not have addressed all potential sources of bias

- Reserved for those aspects of the study design in which significant sources of bias may persist

statistically significant results, with overall mean difference BA-CA of -0.11 years (95% CI -0.41, 0.19).

In regard to the meta-regression, the coefficient for the Africans showed statistical significance with estimate being ($p > 0.05$) (Supplementary Table 2).

Meta-analyses by yearly interval (see also Supplementary Tables 3 to 6 and Supplementary Figs. 1 to 5)

For Caucasian males, seven studies were included [9, 19, 27, 30, 35, 41, 52]. These studies did not show any statistically significant results. The mean difference BA-CA ranged from -0.32 years (at 13 years old) to 0.44 years (at 17 years old). For Caucasian females, six studies were included [9, 27, 30, 35, 41, 52]. These studies did not show any statistically significant results, with mean difference BA-CA ranging from -0.20 (at 10 years old) to 0.34 (at 14 years old).

For Asians, five studies were included [24, 28, 38, 51, 53]. The studies did not show any statistically significant results in females, with mean BA-CA ranging from -0.27 (at 6 years old) to 0.50 years (at 15 years old). In males, however, the studies showed statistically significant results for the following ages:

- Six years: overall mean difference BA-CA of -1.08 years (95% CI -1.49, -0.67)
- Seven years: overall mean difference BA-CA of -1.35 years (95% CI -1.85, -0.85)
- Eight years: overall mean difference BA-CA of -1.07 years (95% CI -1.97, -0.17)
- Nine years: overall mean difference BA-CA of -0.80 years (95% CI -1.43, -0.18)
- Seventeen years: overall mean difference BA-CA of 0.50 years (95% CI -0.08, 0.93)

Table 2 Summary of studies that assessed the reliability of the G&P atlas in Caucasian children

Study	Origin/ ethnicity	Age (years)	*N*	Mean BA-CA (years)	Authors' conclusion	Applicability
Demish and Wartmann 1956	White	9–15	M = 81 F = 70	M = 0 F = 0.5	There is a high positive correlation between BA and CA.	Applicable
Hansman and Maresh 1961	White	0–18	M = 27 F = 36	M = -0.33 F = 0.75	The mean BA for both sexes is equal to CA during infancy but less than CA toward adolescents.	Applicable
Johnston 1963	White	7–17	M = 388 F = 405	M = 0.40 F = 0.20	Children show significant differences between CA and BA.	Needs some modification
Andersen, 1971	Danish	7–18	M = 477 F = 535	M = 0.49 F = -0.43	BA is lower than CA, indicating that the American children mature earlier than the Danish.	Needs some modification
Roche et al 1971	British	2–13	M = 62 F = 82	M = 0.01 F = 0.07	Children matured skeletally at about the same as the G&P standard.	Applicable
Wenzel et al 1984	Austrian	7–16	M = 459 F = 178	M = -0.2 F = -0.13	Major deviations between BA and CA were at and after puberty.	Not clear
Loder et al 1993	White	0–18	M = 203 F = 177	M = -0.1 F = 0.07	The G&P atlas is applicable to white girls at all ages and white boys in early childhood (less than 4 years old). BA of white boys was delayed during middle and late childhood but advanced during adolescence by 5 years.	Not applicable
Kullman, 1995	Swedish	12–19	M = 38 F = 34	M = -0.4 F = -0.4	It is recommended to assess skeletal development using G&P.	Applicable
Ontell et al 1996	White	3–18	M = 208 F = 130	M = -0.29 F = 0.14	The G&P standard is applicable to white girls at all ages, while in boys it can only be applied in adolescence.	Applicable (for girls but not boys)
Koc et al 2001	Southeast Turkey	7–17	M = 225	M = -0.2	Mean BA was delayed between 7 and 13 years old and then advanced between 14 and 17 years. The atlas can be used with some modification.	Needs some modification
Mora et al 2001	European American	0–19	M = 130 F = 130	M = 0.09 F = -0.14	Prepubertal European American children have significantly delayed BA when compared to African American children. Post-pubertal European-American males have significantly advanced BA when compared with African American males. A new standard is needed for reliable BA assessment.	Not applicable
Van Rijn et al 2001	Dutch	5–20	M = 294 F = 294	M = -0.28 F = -0.14	Significant correlation between BA and CA in boys and girls. The G&P atlas is still applicable to Dutch Caucasian children and adolescents.	Applicable
Buken et al 2007	Turkish	11–19	M = 251 F = 241	M = 0.13 F = 0.54	Mean skeletal ages were significantly advanced for boys and girls between 11 and 17 years old. The cause of this acceleration might be new social and cultural factors rather than economic conditions.	Needs some modification
Schmidt et al 2007	Germany	1–18	M = 303 F = 303	M = -0.49 F = -0.39	The G&P atlas method overestimated the samples' age. This may be due to high acceleration in growth.	Applicable
Buken et al 2009	Turkish	11–16	M = 169 F = 164	M = -0.02 F = -0.65	The G&P atlas is appropriate in girls 11–15 years old and boys 11–16 years old from the Black Sea region of Turkey.	Needs some modification
Zhang et al 2009	White	0–18	M = 164 F = 163	M = 0.01 F = -0.15	BA was relatively close to CA in white children.	Applicable
Calfee et al 2010	Caucasian	12–18	M = 62 F = 76	M = 0.98 F = 0.66	American children between 12 and 18 years demonstrate BA exceeding CA. Females between 12 and 15 years old are most likely to demonstrate a discrepancy of at least 2 years between BA and CA, while males demonstrate this throughout adolescence.	Not clear
Cantekin et al 2012	Eastern Turkish	7–17	M = 342 F = 425	M = -0.13 F = 0.20	The mean differences between BA and CA are low enough to be of no practical significance, and thus, this method can be used in all age groups within the current study.	Needs some modification
Santoro et al 2012	Italian	7–15	M = 243 F = 261	M = -0.1 F = 0.40	The G&P method is accurate, particularly in the age ranges of 7–9 years and 10.4–11.5 years.	Applicable
Suri et al 2012	White	9–18	M = 311 F = 261	M = 0.50 F = 0.50	Wide range of differences between BA and CA at each yearly age group from 9 to 18 years. Overall, the differences in skeletal and chronological age were positively correlated.	Not clear
Hackman and Black 2013	Scottish	1–20	M = 249 F = 157	M = -0.13 F = -0.16	The G&P atlas over-aged females from birth until 13 years of age and underestimated males from birth until 13 years of age after which point it consistently over-aged boys between 13 and 17 years of age.	Needs some modification
Paxton et al 2013	Australian	0–18	M = 276 F = 130	M = -0.12 F = -0.30	The G&P atlas is an accurate means of BA determination in Australian children.	Applicable
	White	10–16	M = 46	M = 0.04	The G&P is reliable in Caucasian males.	Applicable

Table 2 (continued)

Study	Origin/ ethnicity	Age (years)	*N*	Mean BA-CA (years)	Authors' conclusion	Applicability
Mansourvar et al 2014						
Gungor et al 2015	Turkish	10–18	M = 259 F = 276	M = 0.64 F = -0.98	It is appropriate to use the G&P method in southern Turkish children; however, a revision is needed for better results and to minimise errors.	Needs some modification
Zabate et al 2015	French	10–19	M = 100 F = 90	M = -0.19 F = -0.53	The G&P overestimated all males and females except boys who are 12 years and girls who are 11 and 18 years old. G&P can be used on French population but not without caution because of a tendency for this method to overestimate age.	Needs some modification
Ozturk et al 2015	Central Turkey Eastern Turkey	9–17 9–17	M = 186 F = 249 M = 189 F = 225	M = -0.10 F = 0.90 M = -0.90 F = -0.90	The G&P atlas was applicable to Caucasian boys of younger age groups and Caucasian girls of all ages. However, some improvement is needed.	Needs some modification
Maggio et al 2016	Western Australian	0–25	M = 180 F = 180	M = 0.24 F = -0.14	The G&P standard is not suitable for the determination of legal majority.	Not clear

A positive value of the mean difference between BA and CA indicates advanced, while a negative value indicates delayed bone age compared to chronological age

M males, *F* females

Based on the results of the yearly interval meta-analysis, we produced graphs for Asians and Caucasians of both sexes (Fig. 6), which show BA according to our meta-analysis compared to BA as assessed by the G&P atlas.

Table 3 Summary of studies that assessed the reliability of the G&P atlas in African children

Study	Origin/ethnicity	Age (years)	*N*	Mean BA-CA (years)	Authors' conclusion	Applicability
Loder et al 1993	African American	0–18	M = 249 F = 212	M = 0.28 F = 0.51	African girls were skeletally advanced by 0.4 to 0.7 years except during middle childhood. While BA for boys was only advanced during adolescence.	Applicable for boys but not for girls
Ontell et al 1996	African American	3–18	M = 95 F = 65	M = 0.28 F = 0.55	African girls showed significant differences at all ages except middle childhood. G&P is applicable to African boys until adolescence.	Applicable for boys but not for girls
Mora et al 2001	African American	0–19	M = 135 F = 139	M = -0.01 F = 0.11	On average, the BA of 10% of prepubertal African American children was 2 SD above the normative data in the G&P atlas. The atlas is imprecise for African American children born after 1980.	Not applicable
Lewis et al 2002	Malawian	1–28	M = 93 F = 46	M = -1.7 F = -1.5	The atlas is inaccurate for this group of children. Poor nutrition and chronic diseases such as malaria and diarrhoea which are endemic in Malawi are likely to be contributing factors.	Not applicable
Garamendi et al 2005	Moroccan	13–25	M = 144	M = -1.7	G&P has a high error rate and therefore should not be considered as an optimal diagnostic method.	Not applicable
Zhang et al 2009	African American	0–18	M = 179 F = 170	M = -0.02 F = 0.03	BA was relatively close to the CA in African American children.	Applicable
Dembetembe and Morris 2012	South African (black)	13–19 20–21	M = 104 M = 27	M = 0.2 M = 2.1	Skeletal maturity as characterised by complete epiphyseal fusion occurred approximately 2.1 years later than G&P method. G&P is not directly applicable to African males.	Not applicable
Mansourvar et al 2014	African American	8–15	M = 47	M = 1.87	G&P is not reliable for assessment of children between 8 and 15 years.	Needs some modification

A positive value of the mean difference between BA and CA indicates advanced, while a negative value indicates delayed bone age compared to chronological age

M males, *F* females

Table 4 Summary of studies that assessed the reliability of the G&P atlas in Asian children

Study	Origin/ ethnicity	Age (years)	*N*	Mean BA-CA (years)	Authors' conclusion	Applicability
So and Yen 1990	Southern Chinese	11.9–12.3	F = 117	F = 0.6	Earlier skeletal maturation was demonstrated. Such a difference was contributed to by improved socioeconomic, nutritional and sociohygienic conditions.	Not clear
So and Yen 1991	Southern Chinese	11.9–12.3	F = 117	F = 0.6	Earlier skeletal maturation was demonstrated. This is attributed to improved socioeconomic condition.	Not clear
Ontell et al 1996	Asian	3–18	M = 63 F = 30	M = -0.03 F = 0.27	The G&P standard is applicable to Asian girls at all ages, while in boys, it can only be applied from birth to 4 years old and from 7 to 13.3 years old.	Applicable (for girls but not boys.)
Krailassiri et al 2002	Thai	7–19	M = 139 F = 222	M = 0.8 F = 0.8	Although the mean difference in BA and CA was equal in both sexes, males clearly differed from the G&P more frequently than females.	Not applicable
Chiang et al 2005	Taiwan	7–19	M = 230 F = 140	M = 0.82 F = -0.3	There is a discrepancy of more than 1 year between BA and CA in some age groups. We believe that some modification of the GP atlas is necessary	Needs some modification
Griffith et al 2007	Chinese	0–18	M = 650 F = 366	M = 0.25 F = 0.15	Hong Kong children appear to mature more slowly in the first decade but more quickly thereafter.	Needs some modification
Zhang et al 2009	Asian	0–18	M = 165 F = 166	M = 0.41 F = 0.24	Asian children mature sooner than white children, especially between 10 and 13 years in girls and between 11 and 15 years in boys.	Not clear
Zafer et al 2010	Pakistan	0–18	M = 535 F = 354	M = 0.1 F = -0.19	This study suggests against the applicability of G&P in Pakistani children. Authors propose a cautious approach while employing G&P in this population to ensure appropriate clinical and medico-legal decisions.	Not applicable
Moradi et al 2012	Iran	6–18	M = 303 F = 122	M = 0.37 F = -0.04	Considering the possibility of a few months' difference, the G&P atlas can be used for the Iranian population.	Needs some modification
Soudack et al 2012	Israeli	0–18	M = 375 F = 304	M = 0.16 F = -0.04	There was no discrepancy between BA and CA in Israeli girls using G&P. There were discrepancies for boys, but these were small.	Applicable
Patil et al 2012	India	1–19	M = 194 F = 181	M = 0.69 F = 0.64	G&P is not applicable to males, especially for age group 4 to 12 years. G&P is applicable to females except age groups 4–7 years, 9–10 years, 15–16 years. A new standard is needed for Indian children.	Not applicable
Shilpa et al 2013	India (Bangalore)	6–15	M = 124 F = 126	M = 0.18 F = 0.29	The G&P method of skeletal age estimation showed accuracy in only certain age groups in Bangalore South zone children.	Needs some modification
Awais et al 2014	Pakistani	0–18	M = 136 F = 147	M = -1.3 F = 0.06	G&P is reliable for girls in all age groups. However, G&P is not accurate for boys in whom it underestimated BA.	Not applicable
Mansourvar et al 2014	Asian American	1–8	M = 48	M = 0.87	The delay in skeletal maturity was more than 2 years for the 4–6 years' age group. Some improvement is needed to enhance the precision of G&P.	Needs some modification
Mughal et al 2014	Pakistan	4.5–9.5	M = 139 F = 81	M = -1.3 F = 0.55	G&P standard significantly underestimates CA in Pakistani children between the ages of 4.5 and 9.5 years.	Not applicable
Rai et al 2014	India	5–15	M = 75 F = 75	M = -0.07 F = -0.33	G&P atlas underestimates CA in children aged between 5 and 9 years.	Needs some modification
Kim et al 2015	Korean	7–12	M = 135 F = 77	M = -0.48 F = -0.02	G&P is applicable to Korean children aged between 7 and 12 years.	Applicable
Mohammed et al 2015	South India	9–20	M = 330 F = 330	M = -0.23 F = 0.02	Mild underestimation of BA was noted in boys. G&P remains applicable to South Indian children.	Needs some modification
Patel et al 2015	West India	6–16	M = 90 F = 90	M = -0.99 F = -0.40	G&P can be used in West Indian children aged between 6 and 16 years.	Applicable

A positive value of the mean difference between BA and CA indicates advanced, while a negative value indicates delayed bone age compared to chronological age

M males, *F* females

Table 5 Summary of studies that assessed the reliability of the G&P atlas in Hispanic children

Study	Origin/ ethnicity	Age (years)	*N*	Mean Ba-CA (years)	Authors' conclusion	Applicability
Jimenez et al 1996	Spanish	0–14	M = 139 F = 100	M = -0.31 F = 0.04	Boys show a delay of around 3 months with respect to the G&P atlas. Girls show a better fit to the corresponding (female) standard of the atlas.	Not clear
Ontell et al 1996	Hispanic	3–18	M = 105 F = 69	M = 0.28 F = 0.38	The G&P atlas is applicable to boys aged between 4 and 13 years and to girls except during adolescence.	Needs some modification
Haiter-Neto et al 2006	Brazilian	7–15	M = 180 F = 180	M = -0.2 F = 0.1	The means of estimated and chronologic ages were similar in all age ranges. The standards can be used with some modification.	Needs some modification
Zhang et al 2009	Hispanic	0–18	M = 178 F = 182	M = 0.30* F = 0.24*	Hispanic children mature sooner than the G&P atlas, especially between 10 and 13 years of age in girls and between 11 and 15 years of age in boys.	Not clear
Santos et al 2011	Portuguese	12–20	M = 136 F = 94	M = 0.12 F = 0.02	The G&P atlas can be used; however, caution must be taken at the end of the growing period.	Needs some modification
Mansourvar et al 2014	Hispanic	15–18	M = 43	M = 0.37	The G&P method is reliable in Hispanic males.	Applicable

A positive value of the mean difference between BA and CA indicates advanced, while a negative value indicates delayed bone age compared to chronological age

M males, *F* females

*$p < 0.05$

Discussion

Bone age assessment is a frequently employed and (in the clinical setting) useful diagnostic technique. Its utility in assessing the age of immigrants and asylum seekers is less secure. Figures from the European Commission estimated that in 2016, about 95,000 unaccompanied minors migrated to Europe, of which more than half were Asians [1]. Although there are no exact figures, many of these immigrants were without valid documents to prove their age. Being unable to prove age, or incorrectly assessing a child as an adult, can restrict the child from having access to their rights such as healthcare and education [56] granted by the law in European countries. Hence, it is important that reliable age estimation methods are used.

Concerned with the reliability of the G&P atlas for different ethnic populations, we considered it important to ascertain its applicability to healthy children. Additionally, bias in studies can result in poor reproducibly and/or lead to distorted results and wrong conclusions. However, in this systematic review, results of the four studies with high risk of bias [8, 23, 29, 46] had little impact on (the statistical significance of) our results. This is because the population of these studies contributed less than 5% to the total included population in which only two studies [8, 29] were included in the meta-analysis, which reduced their impact on sample size and results. A funnel plot shows the absence of a large study with high power as most of the studies scattered toward the bottom; however, minimal risk of publication bias was observed among the studies with three studies switched from the funnel plot (Supplementary Fig. 6) [47, 48, 52].

The G&P atlas appears to be applicable to Caucasians, although some recent studies (included in the meta-analysis) have reported that bone age is advanced compared to

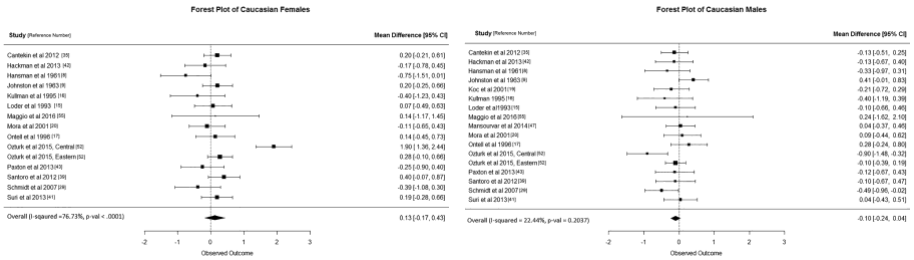


Fig. 2 Forest plot of Caucasians (females and males)

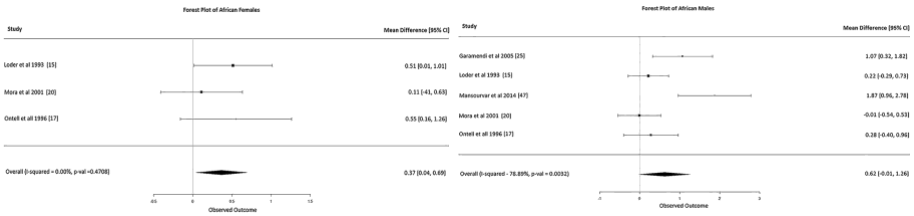


Fig. 3 Forest plot of Africans (females and males)

chronological age in girls up to 13 years old and in boys aged 10 years and above, possibly highlighting the fact that children nowadays are maturing faster than when the atlas was established [32, 42]. Calfee et al [32] assessed the bone age of predominately Caucasian American adolescents (where the G&P atlas was developed). Their skeletal maturation exceeded their chronological age indicating advanced bone age. Perhaps this should not be surprising as Himes [57] reported that skeletal maturation increases by about 0.22 to 0.66 years per decade.

This systematic review and meta-analysis showed no significant difference between BA and CA in Caucasians, which indicates that the G&P atlas is applicable to this group. This is in line with an earlier meta-analysis conducted by Serinelli et al [58] in which no significant difference between BA and CA were found. Note that Serinelli et al included a smaller number of studies; only reported the overall mean difference between BA and CA and did not account for individual age groups.

Concerning the Asian population, three studies recruited Asians living in America [17, 31, 47] while the remaining 17 studies were all carried out in Asia. It seems that skeletal maturation does not conform to the G&P standard at least for some of those who live in East and South Asia. In boys, delay in skeletal maturity during early and middle childhood was followed by advancement during adolescence. Our meta-analysis confirms that there are significant differences between BA and CA in Asian males in two age categories: those aged 6 to 9 years and those aged 17 years. These differences are larger than the standard deviations reported in the G&P atlas for the corresponding age group (± 0.77, ± 0.84, ± 0.90 years at age of 6, 7, 8, and 9 years, respectively), which may have an impact on patient diagnosis and management. In the clinical context, a healthy Asian boy in early childhood could be misdiagnosed as having delayed bone age when using the G&P atlas. The significant advancement in BA compared to CA in Asian males at age 17 is important because this is a critical age in the forensic/legal context, with the individual judged by adult standards in certain legal instances [59].

The G&P standard also seems to be imprecise for Africans. Our meta-analysis of three papers [17, 20, 47] showed significant advancement in bone age of females at all ages ($p < 0.01$). Results from meta-regression with covariates support this difference with BA in Africans being statistically different (Fig. 3). Although our meta-analysis did not show significant difference between BA and CA in African males, some studies reported significant advancement ($p < 0.01$) in adolescence among African American males [15, 17, 47]. Concerning those living in Africa, some studies have shown retardation of bone age among males and females [23, 25, 36]. It is difficult to attribute these variations between Africans only to differences in socioeconomic status, as they were not reported across all studies.

In contrast, the G&P standard appears appropriate for the Hispanic population until adolescence. Our meta-analysis shows no significant difference between BA and CA although

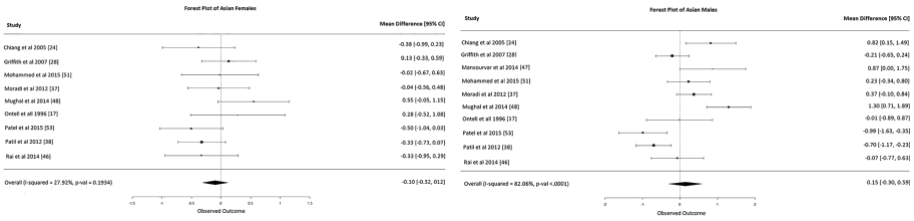


Fig. 4 Forest plot of Asians (females and males)

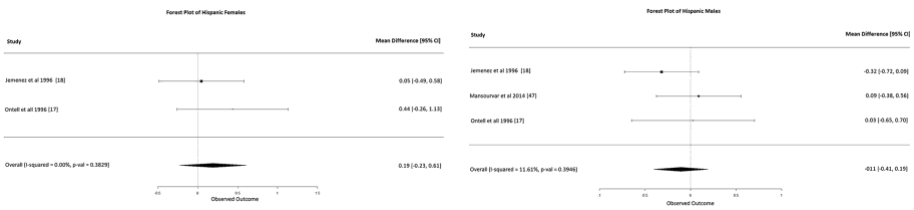


Fig. 5 Forest plot of Hispanics (females and males)

only three studies were included [17, 18, 47]. However, Zhang et al, reported that the G&P significantly overestimated males aged between 10 and 13 years [31].

In the current review, a final analysis was performed combining Asians and Hispanics in order to compare our results to those of Serinelli et al, who used the Cavalli-Sforza classification of ethnicity [60], in which Asians and Hispanics are under one ethnic group (Mongoloid). Our meta-analysis of Asian Hispanics for both females and males showed no significant results (Suppl. Fig.6). This is in contrast to Serinelli's meta-analysis, in which the G&P atlas significantly overestimated chronological age [58]. However, Serinelli et al included only three papers for the Mongoloid population: one related to the Asian population and two to the Hispanic population. One of these latter two studies [61] was excluded from the current systematic review because it included unhealthy children. We therefore believe our results to be more robust.

The major limitation identified in this review is the difficulty in separating ethnicity from socioeconomic status. Relatively few studies reported the socioeconomic status of their sample [9–12, 20, 22, 26, 27, 30, 31, 38, 42, 46, 48, 51]. Children in these studies seemed to follow the same pattern of advancement and delay in bone age as their peers of the same ethnicity in other studies. When bone age is accelerated, new social and cultural factors rather than economic conditions have been suggested to be the main drive [27]. However, our results suggest ethnicity should also be considered when assessing bone age. A further limitation of the study is the failure to calculate the mean absolute and root mean square errors, which might have further confirmed the accuracy of the G&P atlas in relation to each population. However, the mean of each variable (BA and CA) was only available for 13 studies [18, 19, 24, 26–28, 35, 38, 49–53], and for these 13 studies, individual observations were not provided; therefore, the mean error could not be calculated.

Conclusion

This systematic review revealed that the ethnicity/origin of the child can influence the applicability of the G&P standard. The G&P standard is imprecise and should be used with caution in Asian and African populations, particularly when assessing age for forensic/legal purposes. Some caution is also required for Hispanics (particularly males). The G&P atlas can be used with most confidence in Caucasians. There is a complex interrelationship between the impacts of socioeconomic status and ethnicity on bone age using the G&P atlas, which no study has clearly set out to address. Although the graphs in Fig. 6 may be helpful, until new ethnicity-related standards are created,

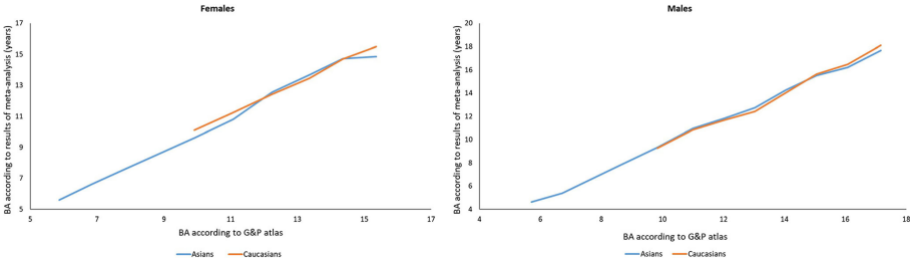


Fig. 6 G&P bone age after adjustment based on meta-analysis (females and males)

clinicians should be aware of the limitations of the G&P method presented in this review.

Funding This study was funded by the Najran University.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Amaka C. Offiah.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry Fabrizio Messina is a part time PhD student at the School of Health and Related Research, University of Sheffield and a Medical Statistician at the Leeds Institute of Clinical Trials Research, University of Leeds.

Informed consent Written informed consent was not required for this systematic review and meta-analysis.

Ethical approval Institutional Review Board approval was not required for this systematic review and meta-analysis.

Methodology
• systematic review
• meta-analysis

References

1. Menjívara C, Perreirab KM (2017) Undocumented and unaccompanied: children of migration in the European Union and the United States. J Ethn Migr Stud 34(8):1378–1386 Available from: http://www.tandfonline.com/doi/pdf/10.1080/1369183X.2017.1404255?needAccess=true
2. Ritz-Timme S, Cattaneo C, Collins MJ et al (2000) Age estimation: the state of the art in relation to the specific demands of forensic practise. Int J Legal Med 113:129–136 Available from: http://www.ncbi.nlm.nih.gov/pubmed/10876982
3. Greulich W, Pyle I (1959) Radiographic atlas of skeletal development of the hand and wrist. Stanford University Press, London
4. Tanner JM, Healy MJR, Goldstein H, Cameron N (2001) Assessment of skeletal maturity and prediction of adult height (TW3 method). WB Saunders, London
5. National Institute for Health and Care Excellence. Methods for the development of NICE public health guidance (third edition) | Guidance and guidelines [Internet]. NICE; 2012 [cited 2017 May 18]. Available from: https://www.nice.org.uk/process/pmg4/chapter/appendix-g-quality-appraisal-checklist-quantitativestudies-reporting-correlations-and
6. R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
7. Demisch A, Wartmann P (1956) Calcification of the mandibular third molar and its relation to skeletal and chronological age in children. Child Dev 27(4):459–473 Available from: http://www.ncbi.nlm.nih.gov/pubmed/13396881
8. Hansman CF, Maresh MM (1961) A longitudinal study of skeletal maturation. Am J Dis Child 101:305–321 Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=13711357
9. Johnston FE (1963) Skeletal age and its prediction in Philadelphia children. Hum Biol 35(2):192–202 Available from: http://www.jstor.org/stable/41448600
10. Andersen E (1971) Comparison of Tanner-Whitehouse and Greulich-Pyle methods in a large scale Danish survey. Am J Phys Anthropol 35(3):373–376 Available from: http://onlinelibrary.wiley.com/doi/10.1002/ajpa.1330350312/full
11. Roche AF, Davila GH, Eyman SL (1971) A comparison between Greulich-Pyle and Tanner-Whitehouse assessments of skeletal maturity. Radiology 98(2):273–280 Available from: http://pubs.rsna.org/doi/10.1148/98.2.273
12. Wenzel A, Droschl H, Melsen B (1984) Skeletal maturity in Austrian children assessed by the GP and the TW-2 methods. Ann Hum Biol 11(2):173–177 Available from: http://www.tandfonline.com/doi/abs/10.1080/03014468400007041
13. So LL, Yen PK (1990) Secular trend in skeletalmaturation in southern Chinese girls in Hong Kong. Z Morphol Anthropol 78(2):145–153 Available from: http://www.jstor.org/stable/25757279
14. So LL (1991) Correlation of skeletal maturation with stature and body weight of Southern Chinese girls in Hong Kong. Z Morphol Anthropol 78(3):307–312 Available from: http://www.jstor.org/stable/25757319
15. Loder RT, Estle DT, Morrison K et al (1993) Applicability of the Greulich and Pyle skeletal age standards to black and white children of today. Am J Dis Child 147(12):1329–1333
16. Kullman L (1995) Accuracy of two dental and one skeletal age estimation methods in Swedish adolescents. Forensic Sci Int 75(95):225–236
17. Ontell FK, Ivanovic M, Ablin DS, Barlow TW (1996) Bone age in children of diverse ethnicity. Am J Roentgenol 167(6):1395–1398
18. Jiménez-Castellanos J, Carmona A, Catalina-Herrera C, Viñuales M (1996) Skeletal maturation of wrist and hand ossification centers in normal Spanish boys and girls: a study using the Greulich-Pyle method. Acta Anat (Basel) 155(3):206–211
19. Koc A, Karaoglanoglu M, Erdogan M, Kosecik M, Cesur Y (2001) Assessment of bone ages: is the Greulich-Pylemethod sufficient for Turkish boys? Pediatr Int 43(6):662–665 Available from: http://onlinelibrary.wiley.com/doi/10.1046/j.1442-200X.2001.01470.x/full
20. Mora S, Boechat MI, Pietka E, Huang HK, Gilsanz V (2001) Skeletal age determinations in children of European and African descent: applicability of the Greulich and Pyle standards. Pediatr Res 50(5):624–628 Available from: http://www.nature.com/doifinder/10.1203/00006450-200111000-00015
21. van Rijn RR, Lequin MH, Robben SG, Hop WC, Van Kuijk C (2001) Is the Greulich and Pyle atlas still valid for Dutch Caucasian children today? Pediatr Radiol 31(10):748–752 Available from: http://www.springerlink.com/index/2MBWW9YW6J35K8HH.pdf
22. Krailassiri S, Anuwongnukroh N, Dechkunakorn S (2002) Relationships between dental calcification stages and skeletal maturity indicators in Thai individuals. Angle Orthod 72(2):155–166
23. Lewis CP, Lavy CB, Harrison WJ (2002) Delay in skeletal maturity in Malawian children. J Bone Joint Surg Br 84(5):732–734 Available from: http://www.bjj.boneandjoint.org.uk/cgi/doi/10.1302/0301-620X.84B5.12642
24. Chiang KH, Chou ASB, Yen PS et al (2005) The reliability of using Greulich-Pyle method to determine children's bone age in Taiwan.

Tzu Chi Med J 6:15–18 Available from: http://www.tzuchi.com.tw/tzuchi/tcmj/94-6/1-4.pdf
25. Garamendi PM, Landa MI, Ballesteros J, Solano MA (2005) Reliability of the methods applied to assess age minority in living subjects around 18 years old: a survey on a Moroccan origin population. Forensic Sci Int 154:3–12
26. Haiter-Neto F, Kurita LM, Menezes AV, Casanova MS (2006) Skeletal age assessment: a comparison of 3 methods. Am J Orthod Dentofacial Orthop 130(4):435.e15–435.e20
27. Büken B, Safak AA, Yazici B, Büken E, Mayda AS (2007) Is the assessment of bone age by the Greulich-Pyle method reliable at forensic age estimation for Turkish children? Forensic Sci Int 173(2–3):146–153 Available from: http://pubs.rsna.org/doi/abs/10.1148/radiol.2481071451
28. Griffith JF, Cheng JCY (2007) Are Western skeletal age standards applicable to Hong Kong Chinese? – a comparison of Greulich and Pyle and TW3 methods. Hong Kong Med J 13(3):S28–S32
29. Schmidt S, Koch B, Schulz R, Reisinger W, Schmeling A (2007) Comparative analysis of the applicability of the skeletal age determination methods of Greulich–Pyle and Thiemann–Nitz for forensic age estimation in living subjects. Int J Legal Med 121(4):293–296 Available from: http://link.springer.com/article/10.1007/s00414-007-0165-7
30. Büken B, Erzengin ÖU, Büken E, Şafak AA, Yazici B, Erkol Z (2009) Comparison of the three age estimation methods: which is more reliable for Turkish children? Forensic Sci Int 183(1–3):103.e1–103.e7 Available from: http://www.sciencedirect.com/science/article/pii/S0379073808003939
31. Zhang A, Sayre JW, Vachon L, Liu BJ, Huang HK (2009) Racial differences in growth patterns of children assessed on the basis of bone age. Radiology 250(1):228–235 Available from: http://pubs.rsna.org/doi/abs/10.1148/radiol.2493080468
32. Calfee RP, Sutter M, Steffen JA, Goldfarb CA (2010) Skeletal and chronological ages in American adolescents: current findings in skeletal maturation. J Child Orthop 4(5):467–470. https://doi.org/10.1007/s11832-010-0289-z
33. Zafar AM, Nadeem N, Husen Y, Ahmad MN (2010) An appraisal of greulich-pyle atlas for skeletal age assessment in Pakistan. J Pak Med Assoc 60(7):552–555 Available from: http://www.jpma.org.pk/PdfDownload/2168.pdf
34. Santos C, Ferreira M, Alves FC, Cunha E (2011) Comparative study of Greulich and Pyle atlas and Maturos 4.0 program for age estimation in a Portuguese sample. Forensic Sci Int 212(1–3):276.e1–276.e7
35. Cantekin K, Celikoglu M, Miloglu O, Dane A, Erdem A (2012) Bone age assessment: the applicability of the Greulich-Pylemethod in Eastern Turkish Children. J Forensic Sci 57(3):679–682. https://doi.org/10.1111/j.1556-4029.2011.02035.x/full Available from: http://onlinelibrary.wiley.com
36. Dembetembe KA, Morris AG (2012) Is Greulich-Pyle age estimation applicable for determining maturation in male Africans? S Afr J Sci 108(9–10) Available from: http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S0038-23532012000500015
37. Moradi M, Sirous M, Morovatti P (2012) The reliability of skeletal age determination in an Iranian sample using Greulich and Pyle method. Forensic Sci Int 223(1–3):372.e1–372.e4
38. Patil ST, Parchand MP, Meshram MM, Kamdi NY (2012) Applicability of Greulich and Pyle skeletal age standards to Indian children. Forensic Sci Int 216(1–3):200.e1–200.e4
39. Santoro V, Roca R, De Donno A et al (2012) Applicability of Greulich and Pyle and Demirijan aging methods to a sample of Italian population. Forensic Sci Int 221(1–3):153.e1–153.e5
40. Soudack M, Ben-Shlush A, Jacobson J, Raviv-Zilka L, Eshed I, Hamiel O (2012) Bone age in the 21st century: is Greulich and Pyle's atlas accurate for Israeli children? Pediatr Radiol 42:343–348
41. Suri S, Prasad C, Tompson B, Lou W (2013) Longitudinal comparison of skeletal age determined by the Greulich and Pyle method and chronologic age in normally growing children, and clinical interpretations for orthodontics. Am J Orthod Dentofacial Orthop 143(1):50–60 Available from: http://www.sciencedirect.com/science/article/pii/S0889540612010293
42. Hackman L, Black S (2013) The reliability of the Greulich and Pyle atlas when applied to a modern Scottish population. J Forensic Sci 58(1):114–119. https://doi.org/10.1111/j.1556-4029.2012.02294.x
43. Paxton ML, Lamont AC, Stillwell AP (2013) The reliability of the Greulich-Pyle method in bone age determination among Australian children: bone age determination using Greulich-Pyle. J Med Imaging Radiat Oncol 57(1):21–24. https://doi.org/10.1111/j.1754-9485.2012.02462.x/full Available from: http://onlinelibrary.wiley.com
44. Shilpa PH, Sunil RS, Sapna K, Kumar NC (2013) Estimation and comparison of dental, skeletal and chronologic age in Bangalore south school going children. J Indian Soc Pedod Prev Dent 31(2):63 Available from: http://www.jisppd.com/text.asp?2013/31/2/63/115696
45. Awais M, Nadeem N, Husen Y, Rehman A, Beg M, Khattak YJ (2014) Comparison between Greulich-Pyle and Girdany-Golden methods for estimating skeletal age of children in Pakistan. J Coll Physicians Surg Pak 24(12):889–893 Available from: https://www.jcpsp.pk/archive/2014/Dec2014/03.pdf
46. Rai V, Saha S, Yadav G, Tripathi AM, Grover K (2014) Dental and skeletal maturity- a biological indicator of chronologic age. J Clin Diagn Res 8(9):ZC60–ZC64
47. Mansourvar M, Ismail MA, Raj RG et al (2014) The applicability of Greulich and Pyle atlas to assess skeletal age for four ethnic groups. J Forensic Leg Med 22:26–29
48. Manzoor Mughal A, Hassan N, Ahmed A (2014) The applicability of the Greulich & Pyle Atlas for bone age assessment in primary schoolgoing children of Karachi, Pakistan. Pak J Med Sci 30(2):409–412
49. Gungor OE, Celikoglu M, Kale B, Gungor AY, Sari Z (2015) The reliability of the Greulich and Pyle atlas when applied to a southern Turkish population. Eur J Dent 9(2):251–254. https://doi.org/10.1111/j.1556-4029.2012.02294.x/full Available from: http://onlinelibrary.wiley.com
50. Kim JR, Lee YS, Yu J (2015) 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 16(1):201–205
51. Mohammed RB, Rao DS, Goud AS, Sailaja S, Thetay AA, Gopalakrishnan M (2015) Is Greulich and Pyle standards of skeletal maturation applicable for age estimation in South Indian Andhra children? J Pharm Bioallied Sci 7(3):218–225 Available from: http://www.ncbi.nlm.nih.gov/pubmed/26229357. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4517325
52. Öztürk F, Karataş OH, Mutaf HI, Babacan H (2016) Bone age assessment: comparison of children fromtwo different regions with the Greulich–Pyle method in Turkey. Aust J Forensic Sci Taylor & Francis 48(6):694–703. https://www.tandfonline.com/doi/full/10.1080/00450618.2015.1119311
53. Patel PS, Chaudhary AR, Dudhia BB, Bhatia PV, Soni NC, Jani YV (2015) Accuracy of two dental and one skeletal age estimation methods in 6–16 year old Gujarati children. J Forensic Dent Sci 7(1):18 Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4330614/
54. Zabet D, Rérolle C, Pucheux J, Telmon N, Saint-Martin P (2014) Can the Greulich and Pylemethod be used on French contemporary individuals? Int J Legal Med 129(1):171–177. https://doi.org/10.1007/s00414-014-1028-7 Available from: http://link.springer.com
55. Maggio A, Flavel A, Hart R, Franklin D (2016) Assessment of the accuracy of the Greulich and Pyle hand-wrist atlas for age

estimation in a contemporary Australian population. Aust J Forensic Sci Taylor & Francis 618(February):1–11. https://www.tandfonline.com/doi/full/10.1080/00450618.2016.1251970
56. Feijen L (2008) The Challenges of ensuring protection to unaccompanied and separated children in composite flows in Europe. Refug Surv Q 27(4):63–73. https://doi.org/10.1093/rsq/hdn049 Available from: https://academic.oup.com/rsq/article-lookup/
57. Himes JH (1984) An early hand-wrist atlas and its implications for secular change in bone age. Ann Hum Biol 11(1):71–75 Available from: http://www.tandfonline.com/doi/abs/10.1080/03014468400006911
58. Serinelli S, Panetta V, Pasqualetti P, Marchetti D (2011) Accuracy of three age determination X-ray methods on the left hand-wrist: a systematic review and meta-analysis. Leg Med (Tokyo) 13:120–133
59. Cole TJ (2015) The evidential value of developmental age imaging for assessing age of majority. Ann Hum Biol 42(4):379–388 Available from: http://www.ncbi.nlm.nih.gov/pubmed/26133364
60. Cavalli-Sforza LL, Chisholm B, Menozzi P, Piazza A (1995) The history and geography of human genes. J Asian Stud 54:490 Available from: http://www.jstor.org/stable/2058750?origin=crossref
61. Holderbaum RM, Veeck EB, Oliveira HW, Silva CL, Fernandes A (2005) Comparison among dental, skeletal and chronological development in HIV-positive children: a radiographic study. Braz Oral Res 19(3):209–215

Pediatric Radiology (2019) 49:372–378
https://doi.org/10.1007/s00247-018-4309-3

ORIGINAL ARTICLE

Estimating bone mass in children: can bone health index replace dual energy x-ray absorptiometry?

Khalaf Alshamrani[1] · Fabrizio Messina[2] · Nick Bishop[1,3] · Amaka C. Offiah[1,3]

Received: 9 March 2018 / Revised: 24 September 2018 / Accepted: 9 November 2018 / Published online: 24 November 2018

Abstract
Background Bisphosphonates have been shown to increase metacarpal cortical width. Bone health index is computed from hand radiographs by measuring cortical thickness, width and length of the three middle metacarpals, and may potentially help predict fracture risk in children.
Objective To compare bone health index with bone mineral density as measured from dual energy X-ray absorptiometry scans in patients with and without bisphosphonate treatment.
Materials and methods Two hundred ninety-three Caucasian patients (mean age: 11.5±3.7 years) were included. We documented absolute values and z-scores for whole-body less head and lumbar spine bone mineral density then correlated these with the bone health index, which were acquired on the same day, in different patient groups, depending on their ethnicity and diagnosis.
Results Bone health index showed moderate to strong correlation with absolute values for whole-body (r=0.52) and lumbar spine (r=0.70) bone mineral density in those not treated with bisphosphonates and moderate correlation absolute values for whole-body (r=0.54) and lumber spine (r=0.51) bone mineral density for those treated with bisphosphonates. There was weak correlation of z-scores, ranging from r=0.11 to r=0.35 in both groups.
Conclusion The lack of a strong correlation between dual energy X-ray absorptiometry and bone health index suggests that they may be assessing different parameters.

Keywords Bisphosphonates · Bone health index · Bone mineral density · Children · Dual energy x-ray absorptiometry

Introduction

Assessment of bone mineral density and bone quality is essential to diagnose patients with diseases affecting the skeleton. In children, the reference standard for assessing bone mineral density is dual energy x-ray absorptiometry. Dual energy x-ray absorptiometry is a valuable tool in patient management, where bone mineral density is assessed at appropriate intervals to monitor response to therapy in patients with low bone mass [1]. Bisphosphonates are commonly used in such patients (e.g., those with osteogenesis imperfecta) and have been shown to increase cortical width [2]. However, dual energy x-ray absorptiometry values are influenced by bone size; therefore, bone mineral density is usually underestimated in children with small bones and overestimated in children with large bones because the depth of the bone is not accounted for [3]. Additionally, dual energy x-ray absorptiometry cannot predict fracture risk in children. Rather, it forms part of a comprehensive skeletal health assessment to monitor patients with low bone mineral density.

During the last three decades, quantitative bone imaging techniques have been improved and tools for analysing images have been developed. One of these methods is radiogrammetry, where the middle phalangeal width and cortical thickness are measured and results are presented as the cortical index [4]. Computer software developed specifically for children automatically calculates bone age and bone mass

✉ Khalaf Alshamrani
Kalshamrani1@sheffield.ac.uk

[1] Department of Oncology and Metabolism, Academic Unit of Child Health, University of Sheffield, Damer Street, Sheffield S10 2TH, UK

[2] Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK

[3] Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield, UK

[5]. The software measures the cortical thickness, width and length of the three middle metacarpals and results are expressed as the bone health index. The software also provides a standard deviation score, which enables comparison with healthy Caucasian children. A small number of studies suggest a potential role for the use of bone health index in assessing bone health in children [6–8]. However, there are limitations to these studies, including small participant numbers [6, 7] and an extended interval of up to 8 months between dual energy x-ray absorptiometry and radiographs [8]. Patients on bisphosphonate therapy were not included in previous studies, yet this group may benefit the most, given that bisphosphonates increase cortical thickness, the very parameter on which the bone health index is based.

The aim of this study is to compare bone health index with bone mineral density dual energy x-ray absorptiometry readings acquired on the same day for different clinical reasons in a large cohort of children, including children on bisphosphonate treatment.

Materials and methods

All procedures performed in this study were in accordance with the ethical standards of our institution. For this study, formal patient consent and research ethics committee approval were not required.

Patient selection

We retrospectively identified dual energy x-ray absorptiometry scans and left-hand radiographs of patients who attended Sheffield Children's NHS Foundation Trust Hospital, United Kingdom, between February 2010 and January 2017. The following inclusion criteria were applied: (1) patients 5–18 years old and (2) dual energy x-ray absorptiometry scans and hand radiographs obtained on the same day.

Hand radiography and dual energy x-ray absorptiometry

BoneXpert software (PACS Server version; Visiana, Holte, Denmark) was used to analyse the hand radiographs [5]. All radiographs were in DICOM format. The software calculated the bone health index based on cortical thickness, width and the length of the three middle metacarpals (Fig. 1).

For bone health index calculations, Caucasian was the default ethnicity at the time of analysis. The data was analysed according to whether patients were or were not on bisphosphonate treatment. Cases were excluded from the study if the BoneXpert software was unable to read the radiograph.

Area bone mineral density of total body less head and lumber spine L1-L4 were extracted from each patient's dual energy x-ray absorptiometry scan. These values were adjusted for age and gender based on normative data provided by the manufacturer. Patient age, gender and the indication for dual energy x-ray absorptiometry were extracted.

Statistical analysis

Statistical analysis was performed using SPSS version 25 for PC (IBM, Armonk, NY). The z-scores of bone mineral density of the total body less head and spine were adjusted for bone age to evaluate the impact of this adjustment on correlation with the bone health index standard deviation score. Each z-score adjusted for bone age for those patients treated with bisphosphonates is based on the computed z-score values (i.e. the internally standardised residuals from the regression analysis that includes bone age) from the untreated patients. The correlation between bone health index and bone mineral density of the total body less head and the spine was assessed separately using Pearson's correlation. Additionally, correlation between bone health index standard deviation score and z-score of bone mineral density of the total body less head and the spine was assessed separately. The correlation between the adjusted z-scores and bone health index standard deviation score were then determined. The strength of the correlations was interpreted according to Evans [9], in which the correlation is deemed to be very weak when the r value is less than 0.19, weak between 0.20 and 0.39, moderate between 0.40 and 0.59, strong between 0.60 and 0.79, and very strong between 0.80 and 1.0. Finally, we generated Bland-Altman plots to graphically illustrate the strength of agreement between the two modalities for the non-bisphosphonate and bisphosphonate groups.

Results

Patient characteristics

Initially, 577 dual energy x-ray absorptiometry/radiograph pairs were identified. Diagnoses included osteogenesis imperfecta (51%), primary osteoporosis (9.5%) and recurrent fracture (5.8%). All diagnoses/indications and patient characteristics are presented in Tables 1 and 2, respectively.

BoneXpert could not interpret 31 (5.6%) radiographs for a number of reasons, including abnormal bone shape, cortical inconsistencies, inconsistencies in length and the image being too sharp. A total of 189 dual energy x-ray absorptiometry/radiograph pairs were excluded as these pairs were acquired for follow-up, which would bias statistical analyses. No dual energy x-ray absorptiometry/radiograph pair was identified for Africans in comparison to a total of 32 dual energy x-ray absorptiometry/radiograph pairs for Asians. However, the Asian patients were excluded from the analysis due to the

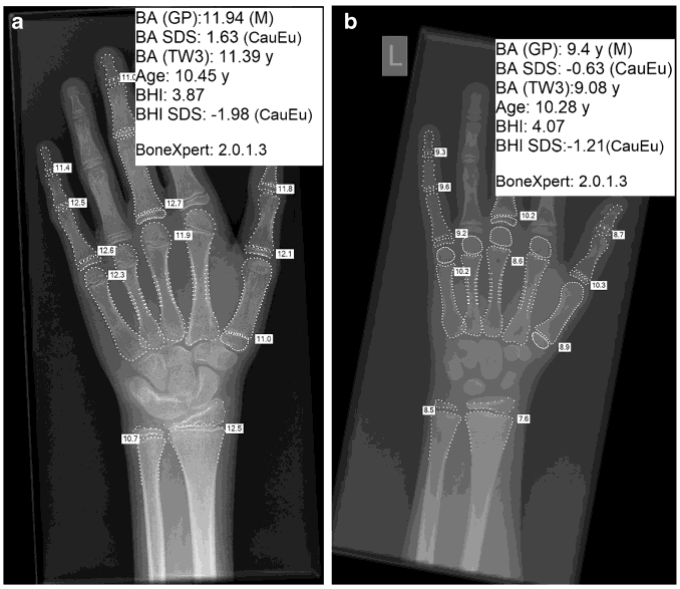


Fig. 1 Left-hand radiographs in two 10-year-old Caucasian boys show relatively similar bone health index: (**a**) bisphosphonate naïve and (**b**) on bisphosphonate treatment. *BA (GP)* bone age using Greulich and Pyle's atlas, *BHI* bone health index, *SDS* standard deviation score, *TW3* Tanner-Whitehouse 3

small number of dual energy x-ray absorptiometry/radiograph pairs identified. Therefore, the final analysis included dual energy x-ray absorptiometry and hand radiographs of 293 patients, 172 (59%) of whom had received bisphosphonate treatment.

Dual energy X-ray absorptiometry and bone health index

As an overall analysis, bone health index correlated moderately with the absolute values of bone mineral density, the total

Table 1 Diagnosis or indication for investigation

	No bisphosphonate treatment	Current/past bisphosphonate treatment
Acute back pain	4	
Bone marrow transplant	7	
Calcinosis cutis	6	
Cerebral palsy	9	
Crohn disease	5	3
Cystic fibrosis	9	4
Fanconi anemia	3	
Growth delay	13	
Hypocalcemia	6	
Hypophosphatasia	4	
Juvenile arthritis	10	8
Malabsorption	4	
Osteogenesis imperfecta	12	138
Post colectomy	3	
Primary osteoporosis	15	13
Recurrent fracture	11	6
Total	121	172

Table 2 Dual energy x-ray absorptiometry and bone health index measurements

	Bisphosphonate group mean (standard deviation)	Non-bisphosphonate group mean (standard deviation)
Number	172	121
Age (years)	12 (3.5)	11.0 (4.0)
Bone age* (years)	12 (3.7)	9.9 (4.3)
Bone mineral density-spine	0.82 (0.18)	0.83 (0.23)
Z-score of bone mineral density-spine	−0.77 (1.5)	−0.26 (1.6)
Adjusted z-score of bone mineral density-spine	0.0 (1.0)	0.5 (1.3)
Bone mineral density-total body	0.86 (0.16)	0.77 (0.19)
Z-score of bone mineral density-total body	−0.62 (1.4)	−0.43 (1.4)
Adjusted Z-score of bone mineral density-total body	0.0 (1.0)	−0.42 (1.1)
Bone health index	4.4 (0.61)	4.2 (0.68)
Bone health index standard deviation score	−1.2 (1.2)	−1.3 (1.5)

*Greulich and Pyle method

body and the spine ($P<0.01$) (Table 3). The data were then divided into two groups depending on whether patients had received bisphosphonate treatment. As seen in Table 3, correlation was stronger in the non-bisphosphonate group as depicted by bone mineral density of the total body ($r=0.70$) and the spine ($r=0.52$, $P<0,01$).

The bone health index standard deviation score showed weak correlation with z-score of the total body less head and the spine (adjusted only for age and gender) in both groups (Table 3). The z-score of bone mineral density of the total body less head and the spine were then adjusted for bone age. The relationship of bone mineral density of the spine adjusted for age and gender alone and adjusted for age, gender and bone age showed similar slopes in both groups with Pearson correlation of 0.74 ($r^2=0.54$) (Fig. 2). Additionally, the relationship of bone mineral density of the total body less head adjusted for age and gender alone and adjusted for age, gender and bone age showed similar slopes in both groups, with Pearson correlation of 0.459 ($r^2=0.21\%$) (Fig. 2). The bone health index standard deviation score showed weak correlation with the z-score of bone mineral density of the total body less head and the spine (adjusted for age, gender and bone age) (Table 3). Bland-Altman plots showed limited agreement between z-score of bone mineral density of the total body less head and the spine (adjusted for age, gender and bone age) and bone health index standard deviation scores (Fig. 3).

Discussion

This study compares bone mineral density measured by dual energy x-ray absorptiometry with bone mass calculated by BoneXpert in a cohort of Caucasian children. BoneXpert was able to provide a reading in the majority of cases. For bisphosphonate naïve children, there was strong correlation

Table 3 Correlation coefficients between bone health index and dual energy x-ray absorptiometry, and bone health index standard deviation scores and z-score reads in bisphosphonate naive and treated patients

		Overall	*P*-value	Bisphosphonate group	*P*-value	Non-bisphosphonate group	*P*-value
Bone health index	Bone mineral density-spine	0.59	<0.01	0.52	<0.01	0.70	<0.01
	Bone mineral density-total body	0.53	<0.01	0.54	<0.01	0.52	<0.01
Bone health index standard deviation scores	Z-score of bone mineral density-spine	0.17	<0.01	0.047	0.26	0.35	<0.01
	Z-score of bone mineral density-total body	0.24	<0.01	0.19	0.20	0.31	<0.01
Bone health index standard deviation scores	Z-score of bone mineral density-spine (adjusted for bone age)	0.22	0.03	0.12	0.12	0.35	<0.01
	Z-score of bone mineral density-total body (adjusted for bone age)	0.26	<0.01	0.26	<0.01	0.25	<0.01

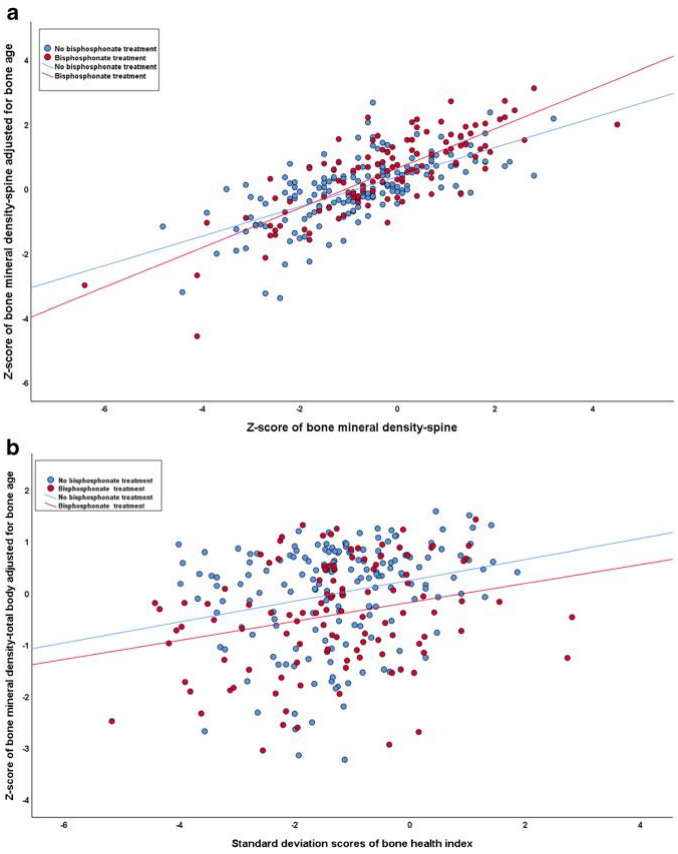


Fig. 2 The relationship of z-score adjusted for age and gender alone, and z-score adjusted for age, gender and bone age show similar slopes: (**a**) bone mineral density of the spine and (**b**) bone mineral density of the whole body

between bone health index and dual energy x-ray absorptiometry absolute values. Previous studies have shown similar correlation ranging from r=0.58 to r=0.85, although ethnicity of patients was not mentioned [6–8].

BoneXpert also provides a bone health index standard deviation score based on data collected from healthy Caucasian children. The bone health index standard deviation score provides a measure of the extent to which a patient's bone mass is deviated from that of healthy Caucasian children of the same bone age and gender. We found a weak correlation between the bone health index standard deviation score and z-scores of dual energy x-ray absorptiometry, even after adjusting the z-scores for bone age. The reasons for this are uncertain but might include differences in other parameters of reference and study populations. Bland-Altman plots showed systematic bias in which differences are higher than means when means are lower, and the differences do not reach zero until the average value reaches or exceeds two standard deviations. However, this is more likely to be due to the fact that the data adjusted for bone age are based on the computed z-score values from patients who had no bisphosphonate treatment.

The bone health index of patients who had not been on bisphosphonate treatment showed a strong correlation, which might suggest that bone health index is a useful tool to monitor children's bone health in this group of patients. In the bisphosphonate group, bone health index showed moderate correlation with absolute dual energy x-ray absorptiometry measures. Approximately 79% of the bisphosphonate group were

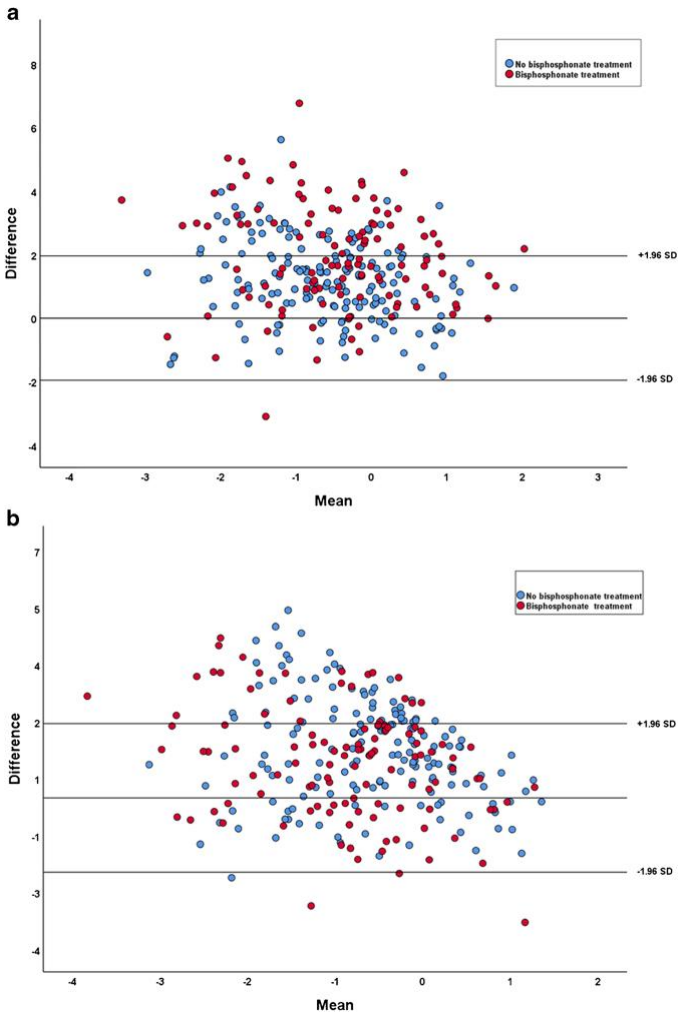


Fig. 3 Bland-Altman plots for the difference in bone mineral density adjusted for bone age, and bone health index z-score, versus the mean of the two estimates. **a** Bone mineral density of the spine. **b** Bone mineral density of the whole body. The plots show limited agreement between z-scores of bone mineral density of the total body less head and of the spine

patients with osteogenesis imperfecta. The metacarpals of that group of patients have smaller bone thickness (external size) and thinner cortices than normal [10]. During treatment with bisphosphonates, cortical thickness increases [11]. This is likely to offer BoneXpert an advantage in this particular group of patients, as the bone health index measured by BoneXpert is dependent on cortical structure, while dual energy x-ray absorptiometry depends on both cortical and trabecular structures. The weaker correlation between bone health index and dual energy x-ray absorptiometry in this group of patients may be because bone health index more closely reflects the true state of the children's bones than dual energy x-ray absorptiometry and merits studies to assess its role in predicting fracture risk in children.

There are some limitations to this study. Firstly, dual energy x-ray absorptiometry z-scores were adjusted for gender, age and ethnicity but not for height and weight. Adjusting for height and weight is expected to explain some of the variance because dual energy x-ray absorptiometry reads may be affected by those parameters. Additionally, dual energy x-ray absorptiometry z-scores adjusted for bone age are based on the computed z-score values from patients who had no bisphosphonate treatment. Ideally, these scores should be based on the z-scores of a healthy population, which were not available to the authors.

Conclusion

The lack of a strong correlation between dual energy x-ray absorptiometry and bone health index suggests that they may be assessing different parameters. The role of bone health index in assessing bone health in children warrants further study before it can be used as an adjunct to or replacement for dual energy x-ray absorptiometry. Future studies need to investigate the clinical use of the bone health index values for predicting fracture risk.

Acknowledgements This study was funded by Najran University.

Compliance with ethical standards

Conflicts of interest None

References

1. Bishop N, Braillon P, Burnham J et al (2008) Dual-energy x-ray aborptiometry assessment in children and adolescents with diseases that may affect the skeleton: the 2007 ISCD pediatric official positions. J Clin Densitom 11:29–42
2. Glorieux FH, Bishop NJ, Plotkin H et al (1998) Cyclic administration of pamidronate in children with severe osteogenesis imperfecta. N Engl J Med 339:947–952
3. Adams JE (2013) Advances in bone imaging for osteoporosis. Nat Rev Endocrinol 9:28–42
4. Barnett E, Nordin BE (1960) The radiological diagnosis of osteoporosis: a new approach. Clin Radiol 11:166–174
5. Thodberg HH, Van Rijn RR, Tanaka T et al (2010) A paediatric bone index derived by automated radiogrammetry. Osteoporos Int 21:1391–1400
6. Nusman CM, Anink J, Otten MH et al (2015) Bone health of patients with juvenile idiopathic arthritis: a comparison between dual-energy X-ray absorptiometry and digital X-ray radiogrammetry. Eur J Radiol 84:1999–2003
7. Schündeln MM, Marschke L, Bauer JJ et al (2016) A piece of the puzzle: the bone health index of the BoneXpert software reflects cortical bone mineral density in pediatric and adolescent patients. PLoS One 11:e0151936
8. Neelis E, Rijnen N, Sluimer J et al (2018) Bone health of children with intestinal failure measured by dual energy X-ray absorptiometry and digital X-ray radiogrammetry. Clin Nutr 37:687–694
9. Evans J (1996) Straightforward statistics for the behavioral sciences. Brooks/Cole Publishing Co., Pacific Grove, CA
10. Arundel P, Bishop N (2010) Diagnosing osteogenesis imperfecta. Paediatr Child Health (Oxford) 20:225–231
11. Rauch F, Travers R, Plotkin H, Glorieux FH (2002) The effects of intravenous pamidronate on the bone tissue of children and adolescents with osteogenesis imperfecta. J Clin Invest 110:1293–1299