December 2014

Comparison between Greulich-Pyle and Girdany-Golden methods for estimating skeletal age of children in Pakistan

Muhammad Awais
Aga Khan University, muhammad.awais@aku.edu

Naila Nadeem
Aga Khan University, naila.nadeem@aku.edu

Yousuf Husen
Aga Khan University, yousuf.husen@aku.edu

Abdul Rehman

Madiha Beg

See next page for additional authors

Authors
Muhammad Awais, Naila Nadeem, Yousuf Husen, Abdul Rehman, Madiha Beg, and Yasir Jamil Khattak

ORIGINAL ARTICLE

Comparison between Greulich-Pyle and Girdany-Golden Methods for Estimating Skeletal Age of Children in Pakistan

Muhammad Awais[1], Naila Nadeem[1], Yousuf Husen[1], Abdul Rehman[2], Madiha Beg[1] and Yasir Jamil Khattak[1]

ABSTRACT

Objective: To compare Greulich-Pyle (GP) and Girdany-Golden (GG) methods for estimation of Skeletal Age (SA) in children referred to a tertiary care hospital in Karachi, Pakistan.

Study Design: Cross-sectional study.

Place and Duration of Study: Department of Radiology, The Aga Khan University Hospital, Karachi, Pakistan, from July 2010 to June 2012.

Methodology: Children up to the age of 18 years, who had undergone X-ray for the evaluation of trauma were included. Each X-ray was interpreted using both methods by two consultant paediatric radiologists having at least 10 years experience, who were blinded to the actual Chronologic Age (CA) of children.

Results: A total of 283 children were included. No significant difference was noted in mean SA estimated by GP method and mean CA for female children (p=0.695). However, a significant difference was noted between mean CA and mean SA by GG method for females (p=0.011). For males, there was a significant difference between mean CA and mean SA estimated by both GP and GG methods. A stronger correlation was found between CA and SA estimated by GP method (r=0.943 for girls, r=0.915 for boys) as compared to GG method (r=0.909 for girls, r=0.865 for boys) respectively. Bland-Altman analysis also revealed that the two methods cannot be used interchangeably. Excellent correlation was seen between the two readers for both GP and GG methods.

Conclusion: There was no additional benefit of using GP and GG methods simultaneously over using GP method alone. Moreover, although GP was reliable in estimating SA in girls, it was unable to accurately assess SA in boys. Therefore, it would be ideal to develop indigenous standards of bone age estimation based on a representative sample of healthy native children.

Key Words: *Skeletal age. Chronological age. Greulich-Pyle. Girdany-Golden. Skeletal age measurement. Bone age measurement.*

INTRODUCTION

Skeletal maturity is one of the many objective methods of determination of age used worldwide. Radiographic investigations are the cornerstone to many such methods of age determination.[1] Accurate estimation of Skeletal Age (SA) is indispensable for social, legal and medicolegal purposes.[2] Furthermore, accurate SA estimation is also important in the realm of sports.[3] In the domain of diagnostic and therapeutic medicine, age determination has found uses in such fields as endocrinology, paediatrics and orthodontics.[4-6]

Many radiographic methods of SA determination have been described in the literature.[7] Greulich-Pyle (GP) method is one of the most popular radiologic methods used worldwide for the estimation of SA.[8] This method requires the comparison of radiographs of the hand and wrist of subjects against standardized images, published in the form of an atlas.[9] Another method frequently utilized in Pakistan is the Girdany and Golden (GG) method.[10] In this method, a plain radiograph of various large joints of the body is obtained and the SA is estimated based upon the appearance of ossification centres around those joints.

Although GP method alone is widely used for estimation of SA throughout the world, but in Pakistan, both methods are simultaneously being employed for the estimation of SA. This increases both the radiation exposure and financial burden on the patients. No study to-date has compared the two methods head-to-head and utility of using both methods simultaneously.

Therefore, in this study, the aim was to assess the efficiency of these two methods for estimation of bone age, to determine whether use of a single method could replace the simultaneous use of both.

METHODOLOGY

A cross-sectional study was carried out by reviewing plain X-ray films of wrist and elbow for the evaluation of trauma. This study was approved by the Ethical Review Committee of our institution and requirement of informed consent was waived. (2414-Rad-ERC-12) Data was retrieved using Radiology Information System for all children (upto 18 years of age) who underwent X-ray for

Department of Radiology[1] / Medicine[2], The Aga Khan University Hospital, Karachi.

Correspondence: Dr. Muhammad Awais, E-5, Five Star Luxury Appartment, Block 14, Gulshan-e-Iqbal, Karachi. E-mail: muhammad.awais@aku.edu

Received: August 12, 2013; Accepted: August 21, 2014.

the evaluation of trauma from July 2010 to June 2012. All children whose X-ray showed fracture, focal bone pathology or soft tissue abnormality were excluded from this analysis. Bone age was estimated for each individual using both GP and GG methods.[9,10] All X-rays were acquired using either Q-RAD (Shimadzu Medical Systems, Japan) or OTC 12S (DEL Medical Systems, USA).

Each radiograph was evaluated by two independent paediatric radiologists having at least 10 years of experience in reporting paediatric radiographs. Radiologists first estimated SA using GP method and data were recorded on a structured proforma. One week later, the same radiologist employed GG method for estimation of bone age. Both reviewers were blinded to the actual Chronologic Age (CA) of patients and to findings reported by each other.

Sample size for this study was estimated using OpenEpi® software available from the Centers for Disease Control and Prevention (CDC). The mean difference calculation method was used to detect a difference of at least 3 months between the two methods of bone age estimation at a confidence interval of 95%, power of 80% and standard deviation of 6 months (as reported by Zafar *et al*.[11] This gave a minimum sample size of 126 for this study (63 in each group).

Statistical Package for Social Sciences (SPSS) version 20 (IBM, Chicago, Illinois) was used for the purpose of data entry and analysis. Quantitative variables like age were expressed as mean ± standard deviation. Children of either gender were divided on the basis of CA into four groups based on previously established criteria by Lodler *et al*.[12] For males, these included early childhood (0-45 months), middle childhood (46 - 90 months), late childhood (91-159 months) and adolescence (160 - 216 months). For females, the age groups were early childhood (0 - 46 months), middle childhood (47 - 100 months), late childhood (101 - 159 months) and adolescence (160 - 216 months). For each age group, mean CA among either gender was compared using Student's t-test. Paired samples t-test was utilized to determine if there was a significant difference between CA and mean SA estimated by the two methods. Pearson product-moment correlation co-efficient was calculated to assess the strength of correlation between SA estimated by each method and actual CA. Level of agreement between the two methods was also determined using Bland-Altman analysis. Finally, inter-observer agreement for both methods was assessed using Intra-class Correlation Coefficient (ICC). A p-value of less than 0.05 was considered statistically significant.

RESULTS

A total of 283 children of either sex upto the age of 18 years were included in this study. Among these children, 136 (48.1%) and 147 (51.9%) were males and females respectively. Mean CA of this study subjects was 102.73 ± 57.82 months and 121.95 ± 63.47 months for boys and girls respectively. Moreover, using student's t-test, no statistically significant difference was noted between mean CA of boys and girls amongst the four age groups (p=0.775, p=0.598, p=0.315 and p=0.488 respectively).

Mean SA for boys and girls as estimated by GP method was 87.14 ± 62.39 months and 122.71 ± 70.14 months respectively. Using paired samples t-test, no statistically significant difference was noted in mean CA and mean SA estimated by GP method for girls (p=0.695). However, for boys, a significant difference existed between mean CA and mean SA estimated by GP method ($p < 0.001$). Paired samples t-test also revealed that mean SA estimated by GP method was not significantly different from mean CA for girls in all four age groups. However, for boys in their middle and late childhood, there was a statistically significant difference between mean CA and mean SA estimated by GP method. No statistically significant difference existed between mean CA and mean SA for boys in their early childhood or adolescence.

Mean SA estimated by GG method was 80.6 ± 71.1 months for boys and 130.2 ± 85.7 months for girls. In contrast with GP method, mean SA estimated by GG method was significantly different from mean CA for both boys ($p < 0.001$) and girls (p=0.011). Using paired samples t-test, a statistically significant difference existed between SA estimated by GG method and CA for girls in their adolescent age group only. In contrast, a statistically significant difference existed between mean CA and mean SA estimated by GG method for boys in their early, middle and late childhood groups.

There was a statistically significant difference between SA estimated by both methods for girls in middle childhood and adolescent groups. For boys in early childhood group, a statistically significant difference existed between mean SA estimated by the two methods. Overall for girls and boys, there existed a significant difference between the two methods for SA estimation. Overall comparison of CA and SA estimated by GP and GG methods for children of all age groups and both genders are depicted in Table I.

The correlation between mean CA and mean SA for both methods in either gender is shown in Figure 1. A stronger correlation was found between mean CA and mean SA determined by GP method as compared to GG method for both male and female children. In girls, Pearson's moment product correlation coefficient was 0.943 ($p < 0.001$) for GP method and 0.909 ($p < 0.001$) for GG method. Pearson's moment product correlation coefficients in boys for GP and GG methods were 0.915 ($p < 0.001$) and 0.865 ($p < 0.001$) respectively.

Bland-Altman plot was utilized to assess the agreement between the two methods as shown in Figure 2. The

Table I: Overall comparison of CA and mean SA estimated by both GP and GG methods in boys and girls (all ages in months).

Age group	Mean CA ± SD	Mean SA by GP method ± SD (p-value)	Mean SA by GG method ± SD (p-value)	p-value for SA by GP vs. GG methods
Boys	102.7 ± 57.8	87.1 ± 62.4 (p<0.001)	80.6 ± 71.1 (p<0.001)	p=0.004
0 - 45	31.3 ± 9.5	27.7 ± 8.5 (p=0.317)	11.9 ± 19.9 (p<0.001)	p<0.001
46 - 90	71.2 ± 17.9	45.1 ± 16.3 (p<0.001)	40.6 ± 32.7 (p<0.001)	p=0.309
91 - 159	129.8 ± 15.6	110.7 ± 28.3 (p<0.001)	106.3 ± 36.6 (p<0.001)	p=0.315
160 - 216	184.8 ± 12.5	174.8 ± 14.8 (p=0.061)	172.8 ± 51.8 (p=0.255)	p=0.735
Girls	121.9 ± 63.5	122.7 ± 70.1 (p=0.695)	130.2 ± 85.7 (p=0.011)	p=0.001
0 - 46	32.5 ± 18.9	28.4 ± 21.6 (p=0.189)	25.7 ± 27.9 (p=0.189)	p=0.561
47 - 100	69.3 ± 17.2	68.8 ± 34.9 (p=0.940)	59.3 ± 47.2 (p=0.129)	p=0.014
101 - 159	124.3 ± 19.0	124.0 ± 30.7 (p=0.947)	129.9 ± 46.1 (p=0.391)	p=0.128
160 - 216	189.6 ± 11.1	194.0 ± 13.8 (p=0.101)	215.6 ± 33.7 (p<0.001)	p<0.001

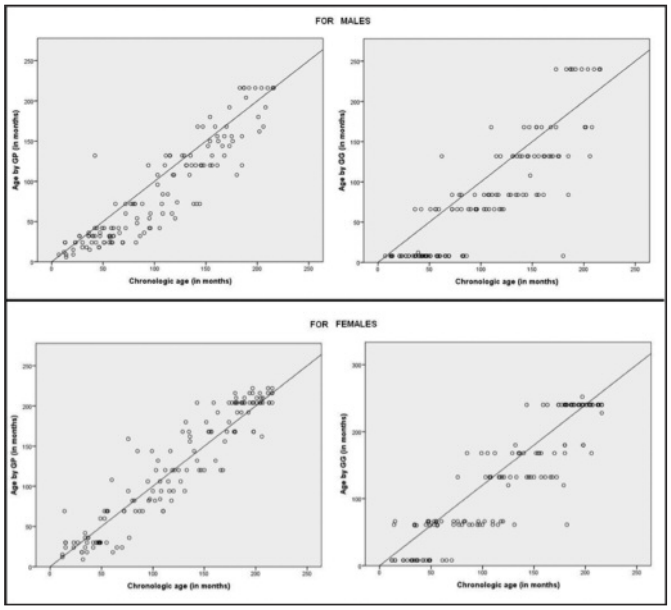

Figure 1: Correlation between CA and SA estimated by both methods in boys and girls.

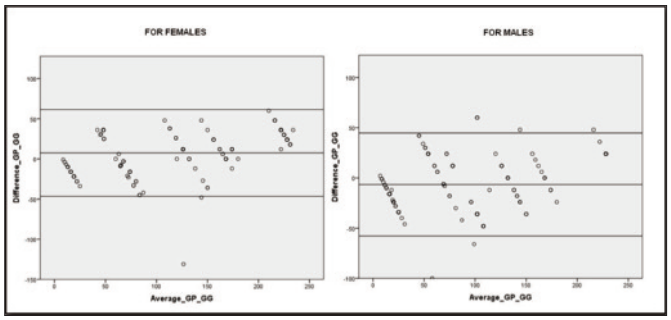

Figure 2: Bland Altman plot for GP and GG methods showing clinically significant discrepancy between the two methods in boys and girls.

analysis illustrate that within 95% confidence limits, the agreement between GP and GG methods fell between -57.8 to +44.7 months for boys and -46.5 to +61.5 months for girls. This implies that a clinically significant discrepancy existed up to 58 months for boys and 62 months for girls in determining SA between these two methods.

Overall, there was an excellent agreement between the two radiologists in estimating SA using both methods. However, it was better for GP (ICC = .998) as compared to GG method (ICC = .974), which suggests lesser variation for GP method among different readers.

DISCUSSION

Bone age determination is of considerable interest for various medical, medico-legal as well as non-medical fields. Consequently, extensive research has been done to develop accurate and reliable tools for the estimation of SA. GP and GG methods are two notable examples of such methods of bone age estimation. Given the fact that both these instruments were developed in the Western world, their applicability to children of South-Asian descent remains undetermined.

The present study was done with the aim of substantiating or refuting the applicability of the two methods to the population of Pakistan. Moreover, as both GP and GG methods are being used simultaneously in many institutions in Pakistan, an attempt was made to determine the added benefit (if any) of using both these methods together. This study participants were young children of up to 18 years of age, who were referred to a prominent tertiary care centre located in Karachi. Being the world's sixth largest urban conglomeration and housing more than twenty million people of various ethnicities, Karachi provides a mixed representative sample of the whole country of Pakistan.

Gender disparity in studies conducted in developing countries, especially those in male-dominated societies, is a well-recognized phenomenon. Moreover, given the physiological differences in male and female skeletal maturation,[12] a sample consisting of equitable number of male and female study subjects is highly desirable and essential for making valid conclusions. This was adequately covered with the present sample size.

GP method is a popular method of bone age determination used in many parts of the world for this purpose. Albeit the original atlas was based on a sample of children of North American descent,[9] it has been extensively studied in children of various ethnicities across the globe. Studies on American, European,

Korean, Moroccan and Spanish children have shown good agreement between actual CA and SA predicted by GP method.[14-18] However, reports on Austrian, Turkish, Chinese, Indian, Hispanic and South Asian children have suggested that GP method may not be applicable to children of such populations.[19-23] Moreover, one study which was done in Erie Basin (area from where the Greulich and Pyle standards originated) suggests that GP method may not be reliable in black children.[24]

With reference to Pakistan, at least two independent studies have contended that GP method may not be wholly reliable for the estimation of SA in this population.[11,25] In this study, GP method was reliable in predicting the SA for girls in all age groups. However, we observed that SA estimated by GP method was not accurate for boys overall and it underestimated SA in boys, thus rendering this method unreliable for use in boys.

The Girdany-Golden method is also being employed in various institutions throughout Pakistan for the purpose of SA determination. Scarce literature is available regarding GG method and its applicability to children of various ethnicities. Virtually no data exists regarding the applicability of this method to the children of Pakistan. In the present study, it was observed that GG method was not reliable for the determination of SA in girls and boys of all ages. It was also observed that GG method tended to underestimate CA for boys up to the late childhood group, while it underestimates CA in the adolescent age group in girls. Consequently, this method does not appear to be reliable for the determination of SA in Pakistani children of either gender.

To the best of authors' knowledge, this study is the first of its kind making a head-to-head comparison of GP and GG methods for estimation of SA. This study revealed that although Pearson product-moment correlation coefficient was significant for both GP and GG methods, it was stronger for GP as compared to GG method in either gender. Also, in terms of intra-class correlation coefficients, GP was better than GG method, even though both showed excellent agreement between the two readers. Bland Altman analysis also illustrated that for 95% confidence intervals, a clinically significant discrepancy existed between GP and GG methods of up to 58 and 62 months for boys and girls respectively. This suggests that these two methods cannot be used interchangeably. These facts were further expounded by paired samples t-test which showed that although GP method was only accurate in estimating SA for girls, GG method overall was unreliable in estimating SA for either gender. A statistically significant difference was also found between SA estimated by GP and GG methods for either gender. Keeping these statistics in mind along with the fact that there was not even a single age group in either gender for which GG method was more accurate than GP method, it may be inferred that the simultaneous use of both GG and GP methods does not provide any added benefit over the use of GP method alone.

However, before coming to any definite conclusions, the limitations of this study must be borne in mind. First and foremost, this study was based on a sample of children who had presented for the evaluation of trauma at a tertiary care hospital. This hospital-based sample of children may differ significantly from the general population and thus may not be truly representative of healthy children. Furthermore, convenience sampling was used in this study, which again may pose problems when generalizing the results of this study to the general population. Ethnicity may be a significant confounder in our analysis, which was not controlled for and thus may have influenced the results of this study.

Despite the limitations of this study, the strengths of this study also merit attention. First and foremost, this study is the first study to analyze the use of GG method for the determination of SA in Pakistani children. Secondly, this study is the first to provide a head-to-head comparison between GP and GG methods, which has not been reported in the published literature before. Last but not the least, as both GP and GG methods are being utilized concurrently in Pakistan, the results of this study are the first to suggest that the use of GP method alone may be equally effective. This has important implications for clinical practice in terms of cost, radiation exposure and feasibility.

Another important finding of this study is that neither GP nor GG method is reliable for the estimation of SA in Pakistani boys. This finding can have major implications for medical and legal domains. Given the fact that the findings of this study are in line with previous reports from the country,[11,25] it is imperative that indigenous standards of SA be developed by studying a larger representative sample of healthy children from our own population.

CONCLUSION

There was no additional benefit of using GP and GG method simultaneously over using GP method alone. Moreover, although GP was reliable in estimating SA in girls, it too was unable to accurately assess SA in boys. Therefore, it would be ideal to develop indigenous standards of bone age estimation based on a representative sample of healthy native children.

REFERENCES

1. Fishman LS. Radiographic evaluation of skeletal maturation. *Angle Orthodont* 1982; **52**:88-112.
2. Black S, Aggrawal A, Payne-James J, editors. Age estimation in the living: the practitioner's guide. West Sussex: *Wiley-Blackwell;* 2010.

3. Engebretsen L, Steffen K, Bahr R, Broderick C, Dvorak J, Janarv PM, *et al*. The International Olympic Committee Consensus Statement on age determination in high-level young athletes. *Br J Sport Med* 2010; **44**:476-84.
4. Seeman E, Wahner HW, Offord KP, Kumar R, Johnson WJ, Riggs BL. Differential effects of endocrine dysfunction on the axial and the appendicular skeleton. *J Clin Investigat* 1982; **69**:1302.
5. Beker L. Principles of growth assessment. *Pediatr Rev* 2006; **27**:196-8.
6. Fishman LS. Chronological versus skeletal age, an evaluation of craniofacial growth. *The Angle Orthodontist* 1979; **49**:181-9.
7. Cole AJL, Webb L, Cole TJ. Bone age estimation: a comparison of methods. *Br J Radiol* 1988; **61**:683-6.
8. Gilli G. The assessment of skeletal maturation. *Horm Res* 1996; **45**:49-52.
9. Greulich WW, Pyle SI, editors. Radiographic atlas of skeletal development of the hand and wrist. Stanford: *Stanford University Press;* 1959.
10. Girdany BR, Golden R. Centers of ossification of the skeleton. The American journal of roentgenology, radium therapy and nuclear medicine 1952; **68**:922-42.
11. Zafar AM, Nadeem N, Husen Y, Ahmad MN. An appraisal of Greulich-Pyle Atlas for skeletal age assessment in Pakistan. *J Pak Med Assoc* 2010; **60**:552-5.
12. Loder R, Estle D, Morrison K, Eggleston D, Fish DN, Greenfield ML, *et al*. Applicability of the Greulich and Pyle skeletal age standards to black and white children of today. *Am J Dis Child* 1993; **147**:1329-33.
13. Patel B, Reed M, Patel S. Gender-specific pattern differences of the ossification centres in the pediatric elbow. *Pediatr Radiol* 2009; **39**:226-31.
14. Roche AF, Davila GH, Eyman SL. A comparison between Greulich-Pyle and Tanner-Whitehouse assessments of skeletal maturity. *Radiology* 1971; **98**:273-80.
15. Groell R, Lindbichler F, Riepl T, Gherra L, Roposch A, Fotter R. The reliability of bone age determination in central European children using the Greulich and Pyle method. *Br J Radiol* 1999; **72**:461-4.
16. Kim SY, Oh YJ, Shin JY, Rhie YJ, Lee KH. Comparison of the Greulich-Pyle and Tanner Whitehouse (TW3) methods in bone age assessment. *J Korean Soc Pediatr Endocrinol* 2008; **13**: 50-5.
17. Garamendi PM, Landa MI, Ballesteros J, Solano MA. Reliability of the methods applied to assess age minority in living subjects around 18 years old: a survey on a Moroccan origin population. *Forens Sci Int* 2005; **154**:3-12.
18. Jiménez-Castellanos J, Carmona A, Catalina-Herrera CJ, Viñuales M. Skeletal maturation of wrist and hand ossification centers in normal Spanish boys and girls: a study using the Greulich-Pyle method. *Acta Anatomica* 1996; **155**:206-11.
19. Wenzel A, Droschl H, Melsen B. Skeletal maturity in Austrian children assessed by the GP and the TW-2 methods. *Ann Hum Biol* 1984; **11**:173-7.
20. Koc A, Karaoglanoglu M, Erdogan M, Kosecik M, Cesur Y. Assessment of bone ages: Is the Greulich-Pyle method sufficient for Turkish boys? *Pediatr Int* 2001; **43**:662-5.
21. Lee MMC. Maturation disparity between hand-wrist bones in Hong Kong Chinese children. *Am J Phys Anthropol* 1971; **34**:385-95.
22. Patil ST, Parchand MP, Meshram MM, Kamdi NY. Applicability of Greulich and Pyle skeletal age standards to Indian children. *Forensic Sci Int* 2012; **216**: 200.e1-200.e4.
23. Ontell FK, Ivanovic M, Ablin DS, Barlow TW. Bone age in children of diverse ethnicity. *AJR* 1996; **167**:1395-8.
24. Loder RT, Estle DT, Morrison K, Eggleston D, Fish DN, Greenfield ML, *et al*. Applicability of the Greulich and Pyle skeletal age standards to black and white children of today. *Am J Dis Child* 1993; **147**:1329-33.
25. Rikhasor RM, Qureshi AM, Rathi SL, Channa NA. Skeletal maturity in Pakistani children. *J Anat* 1999; **195**:305-8.