

Review

Not peer-reviewed version

Precision and accuracy of Radiological Bone Age Assessment in Children among Ethnical Groups

[Sebastián Eustaquio Martín Pérez](#)*, [Isidro Miguel Martín Pérez](#)*, Jesús Vega González, Ruth Molina Suárez, Coromoto León Hernández, Fidel Rodríguez Hernández, Mario Ulises Herrera Perez

Posted Date: 11 September 2023

doi: 10.20944/preprints202309.0670.v1

Keywords: Diagnostic imaging; Radiography; Age Determination by Skeleton; Racial Group; Children



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Precision and Accuracy of Radiological Bone Age Assessment in Children among Ethnic Groups

Sebastián Eustaquio Martín Pérez ^{1,2,3,4}, Isidro Miguel Martín Pérez ^{1,2,*}, Jesús Vega González ⁵, Ruth Molina Suárez ⁶, Coromoto León Hernández ⁷, Fidel Rodríguez Hernández ¹ and Mario Ulises Herrera Perez ^{8,9}

¹ Departamento de Farmacología y Medicina Física, Área de Radiología y Medicina Física, Sección de Enfermería y Fisioterapia, Facultad de Ciencias de la Salud, Universidad de La Laguna, 38200 Santa Cruz de Tenerife, Spain; frguezh@ull.edu.es

² Escuela de Doctorado y Estudios de Posgrado, Universidad de La Laguna, 38203 San Cristóbal de La Laguna, Santa Cruz de Tenerife, Spain. alu0100705157@ull.edu.es

³ Musculoskeletal Pain and Motor Control Research Group, Faculty of Health Sciences, Universidad Europea de Canarias, 38300 Santa Cruz de Tenerife, Spain;

⁴ Musculoskeletal Pain and Motor Control Research Group, Faculty of Sport Sciences, Universidad Europea de Madrid, 28670 Villaviciosa de Odón, Spain

⁵ Institute of Legal Medicine and Forensic Sciences of Santa Cruz de Tenerife, 38230, San Cristóbal de La Laguna, Spain; jveggonf@justiciaencanarias.org

⁶ Pediatric Endocrinology Unit, Pediatric Department, Hospital Universitario de Canarias, 38320, San Cristóbal de La Laguna, Santa Cruz de Tenerife, Spain; rmolsua@gobiernodecanarias.org

⁷ Department of Computer Engineering and Systems, Universidad de La Laguna, 38200 San Cristóbal de La Laguna, Santa Cruz de Tenerife, Spain; cleon@ull.es

⁸ School of Medicine (Health Sciences), Universidad de La Laguna, 38200 Santa Cruz de Tenerife, Spain; mherrera@ull.es

⁹ Foot and Ankle Unit, Orthopedic Surgery and Traumatology Department, 38320, San Cristóbal de La Laguna, Santa Cruz de Tenerife, Spain

* Correspondence: Isidro Miguel Martín Pérez, Departamento de Farmacología y Medicina Física, Área de Radiología y Medicina Física, Sección de Enfermería y Fisioterapia, Facultad de Ciencias de la Salud, Universidad de La Laguna, 38200 Santa Cruz de Tenerife, Spain. alu0100713374@ull.edu.es

Abstract: Introduction: Determination of radiological bone age (BA) is a diagnostic method that consists of estimating chronological age (CA) from the bone maturation of children. Although environmental and hormonal factors can interfere with bone ossification, the process is essentially dependent on ancestral inheritance. These ethnic backgrounds are not usually considered, which can lead to biased interpretations and inadequate decision-making by pediatricians or forensic experts. The objective of this study was to determine the precision and accuracy of these radiological procedures among different ethnic groups. **Methods:** A qualitative systematic review carried out following MOOSE statement and previously registered in the International Prospective Registry of Systematic Reviews PROSPERO (CRD42023449512). Search was performed in MEDLINE (PubMed) (n=561), Cochrane Library (n=261), CINAHL (n=103), Web of Science (WOS) (n=181) and public institutional repositories (n=37) from inception to 31st December 2022 using Mesh as "Diagnostic Techniques and Procedures", "Diagnostic imaging", "Radiography", "Age Determination by Skeleton" and free terms combining with booleans "AND" and "OR". PEDro scale, and Risk of Bias in non-randomized Studies of Exposure (ROBINS-E) were used to assess methodological quality and risk of bias of included studies respectively. **Results:** 51 articles (n=19,531) were included according to the inclusion criteria previously established. There was a good to moderate methodological quality and a high to very high risk of bias. Skeletal methods for determining BA were precised in terms of intra-observer and inter-observer reliability in all ethnic groups. Regarding to accuracy of Skeletal methods in Caucasians and Hispanic children, GPA was accurate at all ages, but in youths, TW3 RUS could be a consistent alternative. In Asian and Arab, GPA and TW3 overestimated BA in adolescents near adulthood. In African youths, GPA overestimated BA while TW3 was more accurate in estimating CA. Dental and Cervical radiographic methods are equally precise but lesser accurate than Skeletal BA determination. **Conclusion:** Skeletal radiographic methods GPA and TW3 are both precise for BA determination among all ethnical groups, but their accuracy in estimating CA can be altered by racial bias.

Keywords: diagnostic imaging; radiography; age determination by skeleton; racial group; children

1. Introduction

Radiographic procedure for Bone Age (BA) assessment is a procedure consisting of the estimation of the chronological age (CA) through visualization of radiographic markers of the skeletal bones (1) in the pediatrics population (2). The classical methods of BA determination are based on the recognition of changes in morphological appearance of left hand-wrist (3) or jawbone radiographs (4) by comparison with reference atlas.

The Greulich-Pyle Atlas (GP), Tanner-Whitehouse (TW) and FELS are the most popular skeletal methods that can be used for BA determination (5).

GP method is the most widely used method for skeletal age estimation in medical practice. This method relies on the shape and maturity level of the primary and secondary ossification centers, as well as the time of fusion between them (6,7)

The TW method is based on the scoring of selected radiographic regions of interest (ROI) in specific bones of the left hand, categorizing them in stages from A to I. The revised TW3 method is used to evaluate skeletal maturation of the radius, ulna, and short bones (RUS) (8) The method involves assigning a score to each bone segment that is evaluated and is more detailed than a simple comparison. It also considers gender differences, which is important in pediatric patients.

The FELS method consists of detection of indicators of maturity as radiographic features of the wrist establishing the maturity index as well as the metric maturity index. This procedure was developed using radiographs of children, with a standardized selection procedure to determine which indicators were useful, which were assigned a score in which the radius, ulna, and carpus contributed more to the final value than others (9,10)

Despite its applicability and simplicity, BA assessment is complex, even for experts (5) because, the maturation of the skeleton is not uniform and appears to depend on both non-modifiable factors such as genetic, and modifiable factors as diet or environmental living conditions. (2) In addition, previous studies have shown differences in indicators of skeletal maturity between different ethnic groups, so ethnic differences have prompted the need for development of new radiological age determination methods. (11,12)

The GP atlas, based on upper-middle-class white populations, may not be applicable to children today, especially with respect to standard development in other racial groups and late BA. (7) Also, TW consolidated the reference values based on a sample drawn from low social strata with an advancement of BA. Alternatively, faced with their generalization problem, some authors such as Eklof and Ringertz developed a method for assessing maturity in terms of the length and width of the skeleton of Scandinavian children (13), while Schmid and Moll developed criteria for white Germans (14) To overcome racial differences in determining BA, the Sugiura Nakazawa method published criteria for both sexes in Japanese children (15).

Additionally, Willems developed a method for assessing BA that was designed to reduce the influence of race and environmental factors (16) Other methods performed on North African children showed significant differences between estimated BA and CA (17).

Given the foregoing, potential racial bias in these X-ray screening procedures may alter their precision and accuracy depending on the ethnical group to which children belong. (2) Systematic use of standard radiographic methods for BA assessment can lead to incorrect decision by experts such as pediatricians following children with advanced or delayed growth or forensic for estimating CA among migrant children. (18,19)

As a consequence, the determination of CA by radiological methods of BA continues to be a challenge even today when it comes to generalizing the results among different ethnic groups. Therefore, the availability of most up-to-date and reliable information regarding metric properties of radiographic procedures for BA assessment are mandatory for experts to solve not only a medical problem but also an ethical and legal issue.

In addition, the absence of previous studies makes it necessary to carry out a systematic synthesis study that helps to identify the potential risk of interracial bias in the determination of BA. Furthermore, due to the lack of previous studies on this research question, comprehensive studies with systematic approaches are needed, which would help to obtain relevant information of applicability of these diagnostic methods for radiological determination of chronological age between the main ethnicities.

Therefore, the objective of this systematic review is to determine the precision and accuracy of radiographic procedures for BA assessment of Children among ethnic groups.

2. Materials and Methods

2.1. Study design

The systematic review study was carried out from June 1, 2023 to September 30, 2023, with the defined protocol and was subdivided into four phases based on the standards of the *MOOSE* statement (Meta-analysis of Observational Studies in Epidemiology guidelines for meta-analyses and systematic reviews of observational studies) (20).

The protocol for this systematic review was previously registered on in the International Prospective Registry of Systematic Reviews PROSPERO (CRD42023449512) and is available for consultation through this website: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023449512

2.2. Search strategy

A literature search was conducted August 1, 2023, to August 28, 2023, to identify all available studies on the precision and accuracy of BA determination by skeletal or dental radiological diagnostic methods in the *MEDLINE (PubMed)*, *Cochrane Library*, *CINAHL* and *Web of Science (WOS)* databases and other public institutional repositories.

In MEDLINE, the first search string was: "Reproducibility of results" [Mesh] OR "Dimensional Measurements Accuracy" [Mesh] OR "Diagnostic Techniques and Procedures" [Mesh] OR "Diagnostic imaging" [Mesh] OR "Radiography" [Mesh] OR "Age Determination by Skeleton" [Mesh] OR "Bone matrix" [Mesh] OR "Carpal bones" [Mesh] OR "radius" [Mesh] OR "Racial Groups" [Mesh] OR "Race factors" [Mesh] OR "White people" [Mesh] OR "Black people" [Mesh] OR "Hispanic or Latino" [Mesh] OR "Asian people" [Mesh] OR "Native Hawaiian or Other Pacific Islander"[Mesh] OR "American Indian or Alaska Native"[Mesh] OR "Pacific Island People"[Mesh] OR "Asian American Native Hawaiian and Pacific Islander"[Mesh] OR "Bone Maturity" [tw] OR "Skeletal Maturation" [tw] OR "Skeletal Age" [tw] OR "Age Measurement" [tw] OR radiograp*[tw] OR radiol *[tw].

Moreover, the second search string was: "Reproducibility of results" [Mesh] OR "Dimensional Measurements Accuracy" [Mesh] OR "Diagnostic Techniques and Procedures" [Mesh] OR "Diagnostic imaging" [Mesh] OR "Radiography" [Mesh] OR "Radiography, panoramic" [Mesh] OR "Age Determination by Teeth" [Mesh] OR "Dentition" [Mesh] OR "Teeth" [Mesh] OR "Tooth" [Mesh] OR "Molar, Third" [Mesh] OR "Incisor" [Mesh] OR "Racial Groups" [Mesh] OR "Race factors" [Mesh] OR "White people" [Mesh] OR "Black people" [Mesh] OR "Hispanic or Latino" [Mesh] OR "Asian people" [Mesh] OR "Native Hawaiian or Other Pacific Islander"[Mesh] OR "American Indian or Alaska Native"[Mesh] OR "Pacific Island People"[Mesh] OR "Asian American Native Hawaiian and Pacific Islander"[Mesh] OR "BA measurement" [tw] OR "Orthopantomography" [tw] OR "Bone Maturity" [tw] OR "Skeletal Maturation" [tw] OR "Skeletal Age" [tw] OR "Age Measurement" [tw] OR radiograp*[tw] OR radiol *[tw].

Similar research equations were used in *Cochrane Library*, *CINAHL* and *Web of Science (WOS)* database and public institutional repositories. Two independent researchers (SMP and IMP) performed the search and a blinded researcher, MHP, scored all retrieved articles by title and abstract, and then scored full-text publications to determine their eligibility. In case of discrepancies, a fourth author served as decision judge (FHR). **Table 1.** Search strategy.

Table 1. Search strategy.

| Search Data | Database | Search equation |
|-------------|---------------------|---|
| 10/01/2023 | MEDLINE (PubMed) | <p>“Reproducibility of results” [Mesh] OR “Dimensional Measurements Accuracy” [Mesh] OR “Diagnostic Techniques and Procedures” [Mesh] OR “Diagnostic imaging” [Mesh] OR “Radiography” [Mesh] OR “Age Determination by Skeleton” [Mesh] OR “Bone matrix” [Mesh] OR “Carpal bones” [Mesh] OR “Radius” [Mesh] OR “Wrist” [Mesh] OR “Racial Groups” [Mesh] OR “Race factors” [Mesh] OR “White people” [Mesh] OR “Black people” [Mesh] OR “Hispanic or Latino” [Mesh] OR “Asian people” [Mesh] OR “Native Hawaiian or Other Pacific Islander” [Mesh] OR “American Indian or Alaska Native” [Mesh] OR “Pacific Island People” [Mesh] OR “Asian American Native Hawaiian and Pacific Islander” [Mesh] OR “Bone Maturity” [tw] “Skeletal Maturation” [tw] OR “Skeletal Age” [tw] OR “Age Measurement” [tw] OR radiograp*[tw] OR radiol *[tw]</p> |
| 10/01/2023 | MEDLINE (PubMed) | <p>“Reproducibility of results” [Mesh] OR “Dimensional Measurements Accuracy” [Mesh] OR “Diagnostic Techniques and Procedures” [Mesh] OR “Diagnostic imaging” [Mesh] OR “Radiography” [Mesh] OR “Radiography, panoramic” [Mesh] OR “Age Determination by Teeth” [Mesh] OR “Dentition” [Mesh] OR “Teeth” [Mesh] OR “Tooth” [Mesh] OR “Molar, Third” [Mesh] OR “Incisor” [Mesh] OR “Racial Groups” [Mesh] OR “Race factors” [Mesh] OR “White people” [Mesh] OR “Black people” [Mesh] OR “Hispanic or Latino” [Mesh] OR “Asian people” [Mesh] OR “Native Hawaiian or Other Pacific Islander” [Mesh] OR “American Indian or Alaska Native” [Mesh] OR “Pacific Island People” [Mesh] OR “Asian American Native Hawaiian and Pacific Islander” [Mesh] OR “bone age measurement” [tw] OR “Orthopantomography” [tw] OR “Bone Maturity” [tw] “Skeletal Maturation” [tw] OR “Skeletal Age” [tw] OR “Age Measurement” [tw] OR radiograp*[tw] OR radiol *[tw]</p> |
| 12/01/2023 | Cochrane Library | <p>([mh “Reproducibility of results”] OR [mh “Dimensional Measurements Accuracy”] OR [mh “Diagnostic Techniques and Procedures”] OR [mh “Diagnostic imaging”] OR [mh “Radiography”] OR [mh “Age Determination by Skeleton”] OR [mh “Bone matrix”] OR [mh “Carpal bone”] OR [mh “Radius”] OR [mh “Wrist”] OR [mh “Racial Groups”] OR [mh “Race factors”] OR [mh “White people”] OR [mh “Black people”] OR [mh “Hispanic or Latino”] OR [mh “Asian people”] OR [mh “Native Hawaiian or Other Pacific Islander”] OR [mh “American Indian or Alaska Native”] OR [mh “Pacific Island People”] OR [mh “Native Hawaiian or Other Pacific Islander”] OR Bone Matur*:ti,ab,kw OR Skeletal Age:ti,ab,kw OR Age Measurement:ti,ab,kw)</p> |
| 12/01/2023 | Cochrane Library | <p>([mh “Reproducibility of results”] OR [mh “Dimensional Measurements Accuracy”] OR [mh “Diagnostic Techniques and Procedures”] OR [mh “Diagnostic imaging”] OR [mh “Radiography, panoramic”] OR [mh “Age Determination by Skeleton”] OR [mh “Dentition”] OR [mh “Teeth”] OR [mh “Tooth”] OR [mh “Molar, third”] OR [mh “Incisor”] OR [mh “Racial Groups”] OR [mh “Race factors”] OR [mh “White people”] OR [mh “Black people”] OR [mh “Hispanic or Latino”] OR [mh “Asian people”] OR [mh “Native Hawaiian or Other Pacific Islander”] OR [mh “American Indian or</p> |

| | | |
|------------|----------------------|--|
| | | Alaska Native"]OR [mh "Pacific Island People"] OR [mh "Native Hawaiian or Other Pacific Islander"] OR Orthopantomography:ti,ab,kw OR Bone Matur*:ti,ab,kw OR Skeletal Age:ti,ab,kw OR Age Measurement:ti,ab,kw) |
| 14/01/2023 | CINAHL | (MH "Reproducibility of results" OR MH "Dimensional Measurements Accuracy OR MH "Diagnostic Techniques and Procedures" OR MH "Diagnostic imaging" OR MH "Radiography" OR MH "Age Determination by Skeleton" OR MH "Bone matrix" OR MH "Carpal bones" OR MH "Radius" OR MH "Wrist" OR MH "Racial Groups" OR MH "Race factors" OR MH "White people" OR MH "Black people" OR MH "Hispanic or Latino" OR MH "Asian people" OR MH "Native Hawaiian or Other Pacific Islander" OR MH "American Indian or Alaska Native" OR MH "Pacific Island People" OR MH "Asian American Native Hawaiian and Pacific Islander" OR bone matur* OR Skeletal Matur* OR Skeletal Age OR Age Measurement) |
| 14/01/2023 | CINAHL | (MH "Reproducibility of results" OR MH "Dimensional Measurements Accuracy OR MH "Diagnostic Techniques and Procedures" OR MH "Diagnostic imaging" OR MH "Radiography, panoramic" OR MH "Age Determination by Skeleton" OR MH "Dentition" OR MH "Teeth" OR MH "Tooth" OR MH "Molar, Third" OR MH "Incisor" OR MH "Racial Groups" OR MH "Race factors" OR MH "White people" OR MH "Black people" OR MH "Hispanic or Latino" OR MH "Asian people" OR MH "Native Hawaiian or Other Pacific Islander" OR MH "American Indian or Alaska Native" OR MH "Pacific Island People" OR MH "Asian American Native Hawaiian and Pacific Islander" OR "Orthopantomography" OR bone matur* OR Skeletal Matur* OR Skeletal Age OR Age Measurement) |
| 20/01/2023 | Web of Science (WOS) | "Reproducibility of results" [Mesh] OR "Dimensional Measurements Accuracy" [Mesh] OR "Diagnostic Techniques and Procedures" [Mesh] OR "Diagnostic imaging" [Mesh] OR "Radiography" [Mesh] OR "Age Determination by Skeleton" [Mesh] OR "Bone matrix" [Mesh] OR "Carpal bones" [Mesh] OR "Radius" [Mesh] OR "Wrist" [Mesh] OR "Racial Groups" [Mesh] OR "Race factors" [Mesh] OR "White people" [Mesh] OR "Black people" [Mesh] OR "Hispanic or Latino" [Mesh] OR "Asian people" [Mesh] OR "Native Hawaiian or Other Pacific Islander" [Mesh] OR "American Indian or Alaska Native" [Mesh] OR "Pacific Island People" [Mesh] OR "Asian American Native Hawaiian and Pacific Islander" [Mesh] OR Bone Maturity [tw] OR Skeletal Maturation [tw] OR Skeletal Age [tw] OR Age Measurement [tw] |
| 28/01/2023 | Web of Science (WOS) | "Reproducibility of results" [Mesh] OR "Dimensional Measurements Accuracy" [Mesh] OR "Diagnostic Techniques and Procedures" [Mesh] OR "Diagnostic imaging" [Mesh] OR "Radiography, panoramic" [Mesh] OR "Age Determination by Skeleton" [Mesh] OR "Dentition" [Mesh] OR "Teeth" [Mesh] OR "Tooth" [Mesh] OR "Molar, Third" [Mesh] OR "Incisor" [Mesh] OR "Racial Groups" [Mesh] OR "Race factors" [Mesh] OR "White people" [Mesh] OR "Black people" [Mesh] OR "Hispanic or Latino" [Mesh] OR "Asian people" [Mesh] OR "Native Hawaiian or Other Pacific Islander" [Mesh] OR "American Indian or Alaska Native" [Mesh] OR "Pacific |

Island People" [Mesh] OR "Asian American Native Hawaiian and Pacific Islander" [Mesh] OR Bone Maturity [tw] OR Skeletal Maturation [tw] OR Skeletal Age [tw] OR Age Measurement [tw]

2.3. Selection and Data Extraction

The selection criteria were: (1) observational studies (cohorts and cross-sectional), bibliographies, case report, classical articles, clinical conferences, comment, comparative studies, evaluation studies, congress proceedings, consensus development conference, dictionaries, editorial, letters, government publications, guidelines, historical articles, lectures, legal case, legislation, (2) published in English, Spanish, French and Portuguese, (3) recruiting children (6 to 12 yrs.), adolescent (13 to 18 yrs.) and young adults (19 to 24 yrs.) (4) of any ethnic group (5) undergoing BA determination by skeleton, dental or cervical radiography procedures (6) published in *MEDLINE (PubMed)*, *Cochrane Library*, *CINAHL*, *Web of Science (WOS)* database and public institutional repositories (7) available in full text and (8) have measured at least outcomes related to precision or accuracy or measures related to radiographic BA assessment.

Data extraction was performed independently by two authors (IMP and SMP), and in case of disagreement, a third author (FHR) was responsible for resolving discrepancies.

A standardized work template based on PECO question was used to extract and detail all the information related to authors, year and country of publication, study design, outcomes, participants (sample size, gender, type of radiological projection, institutional information, etc.), radiographic BA determination method and results of measured outcomes.

The *Cochrane Handbook for Systematic Reviews of Interventions-v.5.1.0* was used to develop these sections. The reliability of the table was tested using a representative sample of the studies to be reviewed.

2.4. Methodological quality assessment

Non-randomized clinical trials or observational studies were assessed using the *Newcastle Ottawa Scale (NOS)* (21) This analysis instrument is based on different domains that include: the selection of the group study (4 points), the compatibility between the data (2 points) and the evaluation of the results (3 points).

For the evaluation of the study through NOS, each of the 7 questions asked is awarded stars in the categories of sample selection and evaluation of results, and a maximum of two stars in the compatibility section, obtaining 9 points as the maximum score.

2.5. Risk of bias assessment

Risk of bias analysis of observational studies trials were independently performed by MHP using the *Cochrane Risk of Bias Tool for observational studies of exposures (ROBINS-E)* (22). This assessment instrument includes flagging questions that should be addressed within each confounding domain, selection of study participants, classification of exposures, deviations from expected exposures, missing data, outcome measurement, and selection of reported outcomes. The response options are: "Low risk", "Some concerns", "High risk", "Very risk" and "No information".

Based on the score obtained in the analysis of the domains of the tool, the existence of a low, some concern, high and very high risk of bias is interpreted globally. Any disagreement between the authors was resolved by discussion, and in case of conflicting scores, the third reviewer (FRH) resolved to make the decision.

3. Results

3.1. Study Selection

A total of 229 studies were detected and analyzed by performing the agreed searches in the detailed databases *MEDLINE* (PubMed) (n=561), *Cochrane Library* (n=261), *CINAHL* (n=103), *Web of Science* (WOS) (n=181) and public institutional repositories (n=37).

After eliminating duplicates of n=671 articles, the remaining 472 papers, were screened, eliminating a total of 393 after reading the title and abstract. Afterwards, the remaining 79 articles were evaluated in full text, eliminating 28 because they did not match our previously established eligibility criteria.

88 papers were eliminated for having a different study design, 5 for not including the reference population, 9 papers were excluded for not following procedures for determining BA, and 6 papers were excluded for not measuring required outcome variables. Finally, a total of n=51 articles were included for the qualitative synthesis. **Figure 1.** MOOSE flowchart of observational studies selection process.

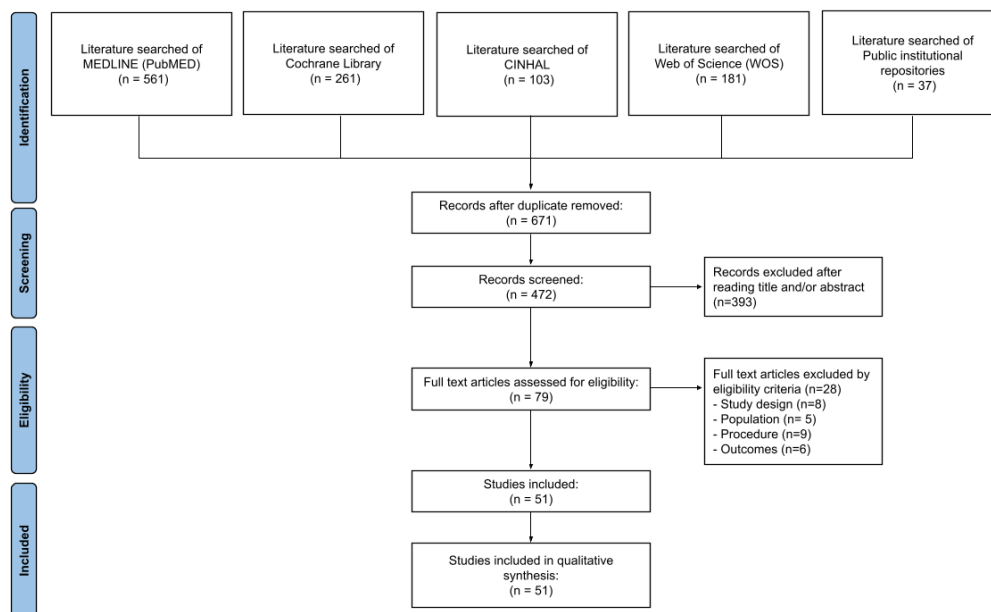


Figure 1. MOOSE flowchart of observational studies selection process.

3.2. Characteristics of included studies

The publications included were published between 1984 and 2023. Out of the 51 studies that were included in the analysis, 21 of them, making up approximately 41.18%, were carried out in various regions of Asia. Specifically, 5 studies were conducted in India (23–27), 4 in Turkey (28–31), 3 in Pakistan (18,32,33), 2 each in Saudi Arabia (34,35), in China (36,37), in South Korea (38,39), and 1 each in Taiwan (40), Iran (41), and Israel (42).

Out of all the studies that were included in the analysis, 16 of them, accounting for 31.37% of the total, were conducted in Europe. Specifically, 3 studies were conducted in the United Kingdom (43–45), 2 in Spain (46,47), 2 in Portugal (48,49), 2 in Italy (50,51), 2 in France (52,53), and 1 each in Austria (54), 1 Germany (55), 1 The Netherlands (56,57), 1 Sweden (57) and 1 Denmark (58).

Out of the studies that were analyzed, a total of 5 (9.80%) were carried out specifically in Africa, with 2 in South Africa (59,60), 1 in Zimbabwe (61) 1 in Botswana (62) and 1 in Ethiopia (63).

Out of the 51 studies that were analyzed, a total of 5 studies (9.80%) were carried out in the region of Oceania. Specifically, there were 2 studies conducted in Australia (64,65), 1 study conducted in Malaysia (66) and another one from Malaysia with radiographies of children from United States (12) and 1 study conducted in Thailand (67)

Out of all the studies that were ultimately included, 4 of them (7.84%) were carried out in America; 1 in the United States of America (68), 2 in Venezuela (69,70) and 1 in Chile (71).

Out of the total included studies, a majority of 25 were retrospective (49.02%) (12,23,28,30,31,36,38,40,42,43,45,46,49–57,64,67,71,72), followed by 19 cross-sectional studies (37.25%) (18,24–26,32–34,39,41,47,48,60–63,65,66,69,70) and a smaller number of 7 prospective studies (13.73%) (27,29,43,44,58,59,68)

A total of 30 papers among total included (58.8%) studied the accuracy (12,23–25,28,29,31,33,36,37,39–41,43,45,47,52–54,57,58,60–62,64–68,70), among them, 1 calculated separately the sensitivity and specificity (50).

A total of n=27 (52.94%) studied precision (23,24,30,32,34,35,38,40,44–46,48–50,52,53,57,59–62,64–68,71) of which 23 studies (45.1%) evaluated repeatability (23,30,32,34,35,38,40,44–46,48–50,52,57,59–62,65–68) On the other hand, n=18 (35.29%) were those in which reproducibility were assessed (24,30,32,34,35,38,40,46,48–50,53,61,62,64,66,67,71).

The studies included X-rays of postero-anterior projection of the hand and left wrist, antero-posterior panoramic orthopantomographies and lateral cervical spine (n=20,100) for the application of BA assessment methods. The sample consisted of children between the age of 0 yrs. (45,46,59) and 22 yrs. (59).

The total number of studies carried out in Caucasians children was 23 (45.90%) studies that represented a total of n=9,777 wrist-carpal radiograph or panoramic radiography used to estimate the BA of this ethnic group (12,28,29,31,42–47,49–56,58,64,65,67,68). In particular, Spain accounts for a total of 1310 cases, 13.39% among Caucasians and approximately 6.51% of the total number of radiographs. (47)

Secondly, the number of studies conducted on children from Asia amounted to n=3097 radiographs (15.40%). Among these studies, a portion focused specifically on Asiatic children (n=2,366, 11.77%) (12,35–40,68), while another subset centered on Indian children (n=731, 3.63%) (23–27) and Indonesian people (66,67).

Thirdly, A total of n=4,674 (n=23.25%) were the radiographs used to estimate BA in children of any Arab ethnic group (18,28–35,41,42) Fourthly, regarding to Latin America, research has been conducted on a total of 1728 radiographs, making up approximately 8.59% of the overall sample. (12,68–71) Fifthly, the total number of studies carried out in African children was 8 studies that represented a total of n=810 (4.02%) wrist-carpal radiographs used to estimate the BA of this ethnic group. (12,43,59–63,68). Lastly, 14 radiographies belonged to other ethnic group such as Caucasian/Asiatic (n=5) (0.02%) (43) and others (n=9) (0.04%) (68).

The estimation of BA involves the utilization of both skeletal and dental techniques. Regarding the skeletal methods, a significant percentage (92.2%) of studies employed the manually applied Greulich and Pyle (GPA) method (n=47) (12,23–25,27–29,31–69,71) followed by the Tanner-Whitehouse-3 method (n=9) (26,35–39,43,50,61) Tanner-Whitehouse-2 (n=4) (44,47,50,58) Alternative techniques employed for the determination of BA included the FELS method (48), and the evaluation of cervical vertebra maturation (CVM), as conducted by Mito et al. (26,30). The latter approach involves assessing BA through the examination of the radiometric surface of the spinous processes.

Other techniques were employed in the included articles, such as Girdany and Golden's method (67), the Fishman method (36,67), the RUS-CHN approach (36), McKay's Method (23,38), the Korean Standard BA method (38), the Thiemann and Nitz Atlas method (55), the Maturos method (30,49), and the Hand and wrist maturation-Ru stage (30). The characteristics of the included studies are presented in **Table S1**. Characteristics of included studies.

3.3. Methodological quality assessment (NOS)

The methodological quality assessment of the included studies ranged from good to moderate with a mean of 6.2 (SD=0.9) out of 9 obtained by the NOS. From a content analysis, it was found that the availability of data (n=48, 80.39%), the verification of the intervention (n=35, 68.62%) and the evaluation of the result (n=30,58.62%) were the domains of the scale that obtained the worst score. Furthermore, there were three studies that achieved the highest methodological quality was marked 8 out of 9 (33,61,66) while eleven studies showed the lowest quality with an overall score of 5 out of

9 (18,26,37,44,47,51,53,56,63,64,71) The detailed of methodological quality assessment with the NOS can be seen at Table 3. Methodological quality assessment (NOS).

Table 3. Methodological quality assessment (NOS).

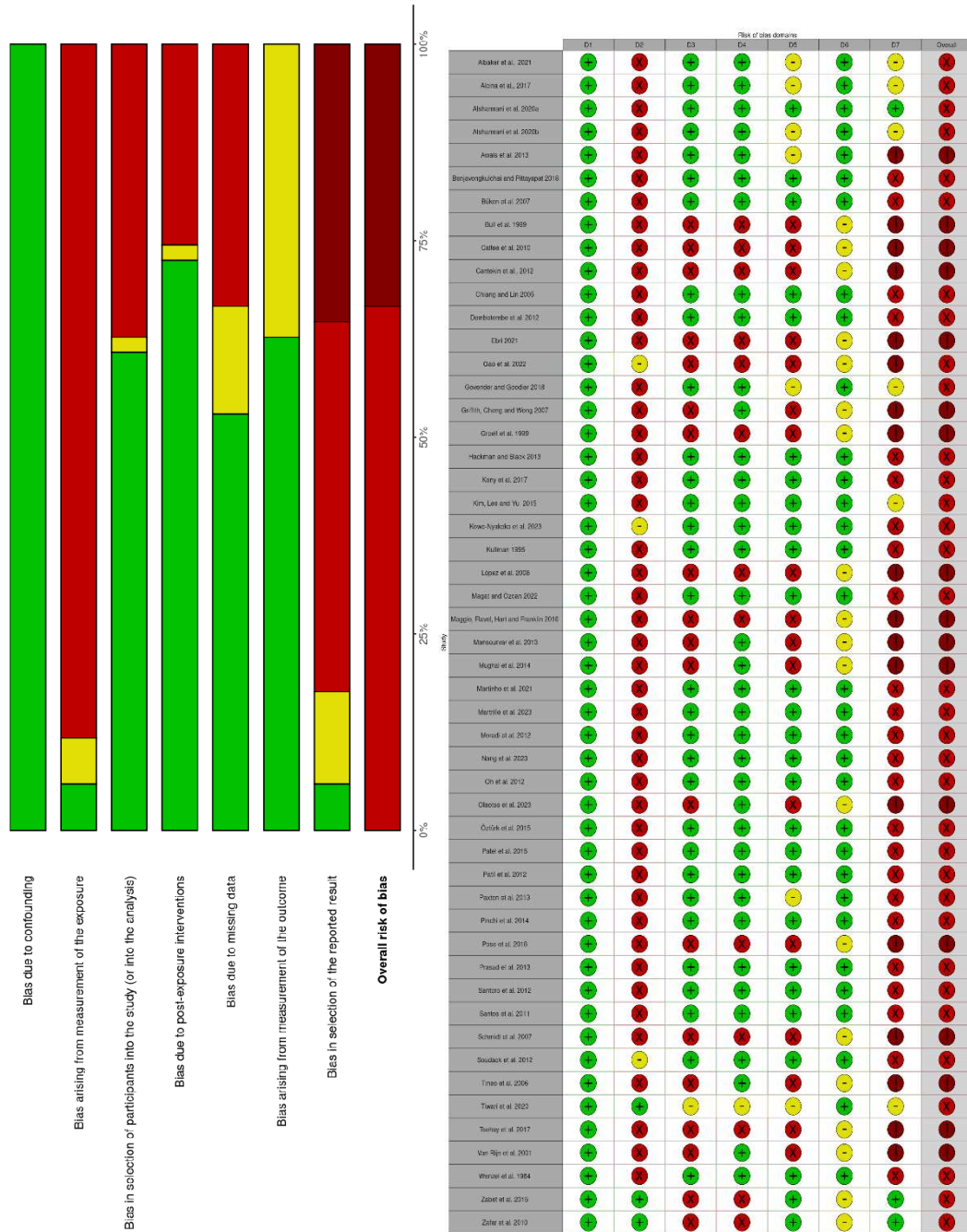
| Authors (yrs.) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|---|---|---|---|---|----|---|---|---|-------|
| Albaker et al. (2021) (34) | * | * | * | * | ** | | * | | 7 |
| Alcina et al. (2017) (46) | * | * | * | * | * | | * | | 6 |
| Alshamrani et al. (2020) (43) | * | * | | * | ** | | * | | 6 |
| Alshamrani et al. (2020) (35) | * | * | * | * | * | | * | | 6 |
| Awais et al. (2014) (32) | * | * | | * | ** | * | * | | 7 |
| Benjavongkulchai and Pittayapat (2018) (67) | * | * | | * | * | * | * | | 6 |
| Büken et al. (2007) (28) | * | * | * | * | ** | | * | | 7 |
| Bull et al. (1999) (44) | * | * | | * | * | | * | | 5 |
| Calfee et al. (2010) (68) | * | * | | * | ** | * | * | | 7 |
| Cantekin et al. (2012) (29) | * | * | | * | * | | * | | 5 |
| Chiang and Lin (2005) (40) | * | * | | * | * | | * | | 6 |
| Dembetembe et al. (2012) (59) | * | * | | * | * | | * | | 6 |
| Ebri (2021) (47) | * | * | | * | * | | * | | 5 |
| Gao et al. (2022) (36) | * | * | | * | ** | | * | * | 7 |
| Govender and Goodier (2018) (60) | * | * | | * | * | * | * | | 6 |
| Griffith, Cheng and Wong (2007) (37) | * | * | | * | * | | * | | 5 |
| Groell et al. (1999) (54) | * | * | | * | ** | * | * | | 7 |
| Hackman and Black (2013) (45) | * | * | | * | * | * | * | | 6 |
| Keny et al. (2017) (23) | * | * | | * | ** | | * | | 6 |
| Kim, Lee and Yu (2015) (38) | * | * | | * | ** | | * | | 6 |
| Kowo-Nyakoko et al. (2023) (61) | * | * | * | * | ** | | * | * | 8 |
| Kullman (1995) (57) | * | * | | * | ** | | * | | 6 |
| López et al. (2008) (70) | * | * | * | * | ** | | * | | 7 |
| Magat and Ozcan (2022) (30) | * | * | | * | ** | | * | | 6 |
| Maggio, Flavel, Hart and Franklin (2016) (64) | * | * | | * | * | | * | | 5 |
| Mansourvar et al. (2014) (12) | * | * | | * | ** | * | * | | 7 |
| Martinho et al. (2021) (48) | * | * | * | * | ** | | * | | 7 |
| Martrille et al. (2023) (52) | * | * | * | * | ** | * | * | | 8 |
| Moradi et al. (2012) (41) | * | * | | * | * | * | * | | 6 |
| Mughal et al. (2014) (18) | * | * | | * | * | | * | | 5 |
| Nang et al. (2023) (66) | * | * | * | * | ** | * | * | | 8 |
| Oh et al. (2012) (39) | * | * | | * | ** | | * | | 6 |
| Olaotse et al. (2023) (62) | * | * | | * | * | * | * | | 6 |
| Öztürk et al. (2015) (31) | * | * | | * | ** | * | * | | 7 |
| Patel et al. (2015) (24) | * | * | * | * | ** | | * | | 7 |
| Patil et al. (2012) (25) | * | * | | * | * | * | * | | 6 |
| Paxton et al. (2013) (65) | * | * | | * | * | * | * | | 6 |
| Pinchi et al. (2014) (50) | * | * | * | * | * | * | * | | 7 |
| Pose et al. (2018) (71) | * | * | | * | * | | * | | 5 |
| Prasad et al. (2013) (26) | * | * | | * | * | | * | | 5 |
| Santoro et al. (2012) (51) | * | * | | * | * | | * | | 5 |
| Santos et al. (2011) (49) | * | * | * | * | * | * | * | | 7 |
| Schmidt et al. (2007) (42,55) | * | * | | * | ** | * | * | | 7 |

| | | | | | | | | | |
|-----------------------------|---|---|---|---|----|---|---|---|---|
| Soudack et al. (2012) (55) | * | * | * | * | * | * | * | * | 7 |
| Tineo et al. (2006) (69) | * | * | * | * | * | | * | | 6 |
| Tiwari et al. (2020) (27) | * | * | * | * | * | * | * | | 7 |
| Tsehay et al. (2017) (63) | * | * | | * | * | | * | | 5 |
| Van Rijn et al. (2001) (56) | * | * | | * | * | | * | | 5 |
| Wenzel et al. (1984) (58) | * | * | | * | ** | | * | | 6 |
| Zabet et al. (2015) (53) | * | * | | * | * | | * | | 5 |
| Zafar et al. (2010) (33) | * | * | * | * | * | * | * | * | 8 |

Note: *Newcastle Ottawa Scale (NOS)* domains: (1) Representativeness of Exposed Cohort (*); (2) Selection of Non-Exposed Cohort (*); (3) Ascertainment of Intervention (*); (4) Demonstrate Outcome Assessed before Intervention (*); (5) Comparability of Cohorts on the Basis of Design or Analysis (**); (6) Assessment of Outcome (*); (7) Adequacy of Follow-Up (*); (8) Data available (No Missing Data)(*).

3.4. Risk of bias assessment (ROBINS-E)

The overall risk of bias determined by the ROBINS-E instrument was high to very high. A high risk of bias of measurements of the exposure were detected in 88.23% (n=45) of included papers (12,18,23–26,28–32,34,35,37–41,43,45–52,54–60,62–71). In addition, the selection risk of the reported outcome was scored high in 82.35% of the studies (n=42).(12,18,23–26,28–32,36,37,39–42,44,45,47–52,54–59,61–71) Finally, the selection bias of the participants who were part of the study samples was high in 37.25% (n=19) of the total articles included in this review. (12,18,29,33,36,37,44,47,53–56,62–64,68–71) Risk of bias analysis is detailed in Table 4. Risk of Bias in non-randomized Studies of Exposure (ROBINS-E).



Note: Cochrane Risk of Bias Tool for observational studies of exposures (ROBINS-E) domains: (1) Bias due to confounding; (2) Bias arising from measurements of the exposure; (3) Bias in selection of participants into the study (or into analysis); (4) Bias due to post-exposure interventions; (5) Bias due to missing data; (6) Bias arising from measurement of the outcome; (7) Bias in selection of the reported result.

3.5. Data synthesis

3.5.1. Precision and Accuracy of Skeletal Method for BA Assessment Among Caucasian ethnicities Children

3.5.1.1. Precision of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Intra-examiner reliability (Pearson Correlation Coefficient)

The intra-examiner reliability of GPA for determining BA in Caucasian children was moderate to excellent in the included studies, indicating high repeatability.

In Mediterranean countries, Pinchi et al. (2014) (50) in a retrospective observational study carried out in Florence (Italy) showed that the intra-examiner reliability of this method to determine CA was very high for both Caucasian boys $r=0.907$ (95%CI= 0.761-0.966, $p<0.05$) and girls $r=0.928$ (95% CI= 0.789-0.977, $p<0.05$) In this same country, Santoro et al. (2012) (51) with southern children aged 7 to 15 yrs. also found that the inter-examiner reliability of the GPA was moderate $r=0.88$ ($p<0.0001$) in boys and somewhat lower $r=0.81$ ($p<0.0001$) in girls. Moreover, in Portuguese children aged 12 to 10 yrs. belonging to this ethnic group, Santos et al. (2011) (49) corroborated the high intra-examiner reliability $r=0.99$ ($p<0.05$) when BA was scored with the GPA method.

In other studies, conducted in northern Europe with Caucasian population, Kullman (1995) (57) found that the GPA had a moderate intra-examiner reliability ($r=0.64-0.74$) in determining the CA following Swedish children aged 12 to 19 yrs. On the other hand, Hackman and Black (2013) (45,50) found that in Scottish children and adolescents under 21 yrs. of age, the intra-examiner reliability of the GPA to quantify age was excellent $r=0.969$ ($p<0.001$).

In Lower Saxony (Germany), Schmidt et al. (2007) (55) showed a high correlation of GPA as a method to identify changes in CA in both boys $r=0.96$ ($p<0.05$) and girls $r=0.96$ ($p<0.05$). In children aged 5 to 19 yrs. from Rotterdam in the Netherlands, Van Rijn et al. (2001) (56) identified a Pearson's correlation coefficient of $r=0.979$ for males ($p<0.001$) and $r=0.974$ ($p<0.001$) for female girls, indicating high precision in estimating CA.

These results are similar to other studies carried out in other Anglo-Saxon countries.

On the one hand, in the prospective cohort of Caucasian children in the United States, Calfee et al. (2010) (68) identified that the intra-examiner reliability of the GPA to estimate CA was moderate, reaching $r=0.890$ ($p<0.001$).

On the other hand, in Australia, Maggio, Flavel, Hart and Franklin (2016)(68) reported a very high repeatability of GPA in Caucasian boys, showing a very strong Pearson's correlation coefficient between BA and CA in boys of $r=0.970$ and in girls $r=0.972$.

Inter-examiner reliability or concordance (Intraclass correlation coefficient)

Inter-examiner reliability of the GPA used to assess BA in Caucasian children ranged from low to high in the included studies, implying that its reproducibility was controversial.

In France, the agreement of the GPA method obtained by intraclass correlation coefficient (ICC) was excellent for the Caucasian child population $ICC=0.94$ (95% CI: 0.91-0.96, $p<0.05$) (52) In the case of the United Kingdom, Alshamrani et al. (2020) (43) detected slight differences in concordance attributed to sex in a sample of British Caucasian children finding that women scored a lower intraclass correlation coefficient ($ICC=0.984$) than males ($ICC=0.991$).

In America, Calfee et al. (2010) found that the agreement of this method was very good, $ICC=0.982$ when CA was estimated in a sample of Caucasian children living in the northwestern United States. (68)

Inter-examiner reliability or concordance (Cohen's kappa coefficient)

On the other hand, the concordance of the GPA method estimated by Cohen's Kappa Coefficient seems to give rise to some controversy.

In Europe, when this procedure was used to determine CA in Caucasian Portuguese girls under 13 yrs. of age, Martinho et al. (2021) (48) found that inter-observer reliability was low $k=0.48$ ($p<0.05$) Also, in France, Zabet et al. (2014) (53) found in a sample of Caucasian children from the city of Tours aged 10 to 19 yrs. that Cohen's Kappa coefficient applied to the GPA showed an inter-examiner reliability of $k=0.96$ ($p=0.0177$).

In the middle east, Soudack et al. (2012) (42) also reported low agreement among examiners for GPA, $k=0.371$ ($p=0.0177$) in a sample of Caucasian children from Edmond and Lily Safra Children's Hospital in Tel Aviv (Israel). Within the same sample, the degree of agreement in girls was significantly higher $k=0.4667$ ($p=0.005$). Along the same lines, Büken et al. (2007) (28) published a low agreement among the evaluators when they applied the GPA to estimate the CA in boys $k=0.275$ ($p<0.001$) and for Turkish girls $k=0.143$ ($p<0.001$) of Caucasian ethnicity.

In contrast, Maggio, Flavel, Hart and Franklin (2016) (64) published in a study of Caucasian children in the Perth region (Australia) that showed that when CA was determined from the GPA method, regarding reproducibility, there was great agreement among examiners ($k=0.887$, $p<0.001$).

Inter-examiner reliability or concordance (Lin's coefficient of agreement)

Finally, Alcina et al. (2017) (46) showed through Lin's correlation coefficient of agreement ($\rho_c = 0.99$) that the repeatability of the GPA method in Spanish Caucasian children from 0 to 18 yrs. old is excellent regardless of the sex of the sample.

3.5.1.2. Accuracy of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Mean differences

With regard to accuracy, that is, the property of the BA estimation method to determine the CA of a person, studies generally support its use in Caucasians although they recognized slight underestimation of some months with respect to the CA of children.

Mansourvar et al. (2013) (12), in a retrospective study with Caucasian living in Malaysia, identified that the accuracy of the method was excellent for children aged 10 to 16 yrs. ($MD=0.044$ yrs., $p>0.05$). In contrast, although Kullman (1995) (57) confirmed its accuracy for Swedish children, found that GPA underestimated CA ($MD=0.4$ yrs., $p>0.05$). Similarly, in Caucasian children from Montpellier (France) GPA underestimated CA by a magnitude of $MD=1.27$ mths. ($SD=1.56$, $p<0.05$) (52). Also in northern France, Zabet et al. (2014) showed that in Caucasian children, GPA slightly underestimated the CA ($MD= 2.29$ mths., $SD=10$, $p<0.05$).

With the sample obtained from the retrospective study of Santoro et al. in 2012 conducted in southern Italy with Caucasian population found that GPA slightly underestimates CA for both males ($MD=0.1$ yrs., $SD=1.3$, $p=0.18$) and females $MD=0.4$ yrs. ($SD=1.0$, $p<0.0001$) aged between 7 and 15 yrs. (51). Santos et al. (2011) (49) conducted with Caucasian Portuguese children aged 12 to 20 yrs. living in the city of Coimbra and in which it was found that the GPA underestimated in a range of 2 to 7 mths. the CA of the participant ($p<0.05$).

Likewise, in Caucasian children from northwestern Germany, Schmidt et al. (2007) (55) showed that this method of estimating CA underestimated the age of boys $MD=0.49$ yrs. ($SD=2.02$, $p<0.05$) and girls $MD=0.39$ yrs ($SD=2.16$, $p<0.05$).

Wenzel et al. (1984) found statistically significant differences between chronological and BA for GPA in Austrian boys from Graz ($p<0.01$) but there were no differences in adolescent girls aged between 7 and 16 yrs. ($p=0.4$). Groell et al. (1999) (54) also found no statistically significant differences between bone and CA, although they recognized underestimation of BA in the same ethnic group ($MD= 0.4$ mths., $SD=4.0$ in boys; $MD=1.1$ mths., $SD= 5.9$ in girls, $p=0.20$).

Agreeing in results, Alshamrani et al. (2020) with a sample of Caucasian children from Sheffield (United Kingdom) observed that GPA underestimated CA by 4 mths. ($p<0.01$) in males (43). When it is analyzed by age, the Scottish study of Hackman and Black (2013) published that in the case of males aged 0-2 yrs., the GPA underestimates the CA from 0.2 to 10 mths. ($p<0.05$). A similar situation occurs in the case of children aged 0 to 10 yrs. for whom the method underestimates their age from 2.44 to 3.54 mths. ($p<0.05$).

Paradoxically, the same authors pointed out that the situation is reversed from 11-15 yrs., an interval in which the GPA overestimates the CA by 1.74 mths. ($p<0.05$). The average difference observed increases in the order of 1.62-11.05 mths. in the case of male adolescents aged 13 to 17 yrs. ($p<0.05$) and in girls aged 9-17 yrs. overestimates from 0.20 to 5.73 mths. ($p<0.05$) (45).

Also, the Spanish study of Ebri (2021) (47) when comparing the accuracy of GPA in relation to the Ebri carpal index (EOIC) it was observed that the GPA overestimates BA by almost 6 mths. In this same work, when comparing the accuracy of GPA in relation to the carpo-metacarpal-phalangeal index EOICMF it was observed that the GPA overestimates BA by almost 6.5 mths. If the accuracy of GPA is compared in relation to the metacarpophalangeal index EOIMF, it could be observed that the GPA overestimates BA by almost 5 mths.

In the Middle East, the study by Cantekin et al. (2012) (29) conducted among Caucasian children from eastern Turkey observed how the GPA method slightly underestimated the CA of participants

in MD=0.13 yrs. (95% CI: 0.31-0.70 yrs., $p>0.05$). In relation to children aged 10 to 17 yrs. of Caucasian children from eastern Turkey, there was a delay between the age that scored the GPA and the CA of the child in mean difference values between 0.02 yrs. for the youngest ages and 0.24 yrs. for the limit with adulthood. In the case of Turkish girls of Caucasian ethnicity, these differences were higher in this same age range, reducing to 0.03 yrs the differences between the age obtained by the GPA and the CA of 17 yrs.

In this same sense, the GPA underestimated the CA in Caucasian children aged 9 to 17 yrs in the regions of the Anatolian peninsula of Malatya and Sivas, finding a difference in the results for males MD=1.19 mths. (95%CI: 12.81 ± 2.3 mths., 13.71 ± 2.6 mths., $p<0.05$) compared to women MD=0.90 mths. (95%CI= 12.91 ± 2.3 mths., 14.11 ± 2.6 mths., $p<0.05$) (31)

In this same region, the GPA underestimated CA in Caucasian children from Tel Aviv (Israel) aged 15 to 18 yrs., finding a difference in results for males MD=2.9 mths. (95%CI, $p<0.0043$) (42,65)

If we analyze the results of Australia, the accuracy of the GPA was measured for Caucasian children of this country finding a slight underestimation of BA for both boys (MD = 1.5 mths., $p = 0.142$) and girls (MD = 3.7mths., $p = 0.002$).(65) Strikingly, if we analyze by age during early childhood the GPA method underestimates (MD=0.81 mths., $p=0.719$) but as they grow GPA begins to overestimate the BA of young people (MD = 3.8 mths, $p=0.001$).

3.5.1.3. Sensitivity and specificity of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Regarding the sensitivity of the GPA, a figure of 90% was found for boys and 87.71% for women point in terms of the specificity of this method was reached 87.18% for boys and 82.76% for Caucasian girls from Italy. (50)

3.5.1.4. Precision of radiographic skeletal methods Tanner-Whitehouse 2 and 3 (TW2 and TW3)

Intra-examiner reliability (Pearson Correlation Coefficient)

The intra-examiner reliability of TW2 and TW3 in determining BA in Caucasian children was excellent in the included studies indicating high repeatability. Some retrospective studies such as that of Pinchi et al. (2014) (50) found that the intra-examiner reliability of TW2 was very high for both male children $r=0.862$ (95% CI= 0.759-0.949, $p<0.05$) and for female $r=0.929$ (95% CI= 0.793-0.978, $p<0.05$)

In the case of TW3, this same work carried out with Italian Caucasian children showed that the intra-examiner reliability of TW3 was also high for both male children $r =0.843$ (95%CI= 0.617-0.942, $p<0.05$) as for female $r=0.910$ (95%CI= 0.817-0.956, $p<0.05$).

3.5.1.5. Accuracy of radiographic skeletal methods Tanner-Whitehouse 2 and 3 (TW2 and TW3)

Mean differences

Ebri (2021) showed the accuracy of TW2 to assess the CA of Hispanic Caucasian children from the mean differences between TW2 and different anthropometric indices validated in this population. For the carpo-metacarpal-phalangeal index (EOICMF) it was observed that TW2 overestimated CA by almost 4 mths. and 6 mths. ($p>0.05$), with little difference between sexes.

Similar results were found by analyzing the differences in CA of TW2, metacarpal-phalangeal index (EOIMF) and Ebrí-carpal index (EOIC) in which an overestimation of 5 mths in the age of male children was found. (47)

3.5.1.6. Sensitivity and specificity of radiographic skeletal methods Tanner-Whitehouse 2 and 3 (TW2 and TW3)

Pinchi et al. (2014) published that sensitivity of TW2 in Italian Caucasian children was 100% for men and 87.50% percent for women while specificity reached 72.92% in boys and 72.41% in girls. This same study conducted by the University of Florence (Italy), published that the sensitivity of TW3 in Italian Caucasian children was 90% for men and 71.42% percent for women while specificity reached 87.5% in boys and 83.87% in girls. (50)

3.5.1.7. Precision of radiographic dental method of Demirjian

Dental methods of estimating CA are not precise and accurate enough to become an alternative to skeletal radiological methods. However, despite the above, they can be effective alternatives to specify the CA in the absence or impossibility of interpretation of the AP X-ray of the carpus and left wrist.

Intra-examiner reliability (Pearson Correlation Coefficient)

The work of Santoro et al. (2012) on the accuracy of the Demirjian method found an intra-examiner reliability calculated through the Pearson correlation coefficient of $r=0.77$ which would indicate a moderate accuracy of this dental method to detect changes in CA. (51)

3.5.1.8. Accuracy of radiographic dental method of Kullman

Mean differences

When analyzing the accuracy of Kullman's (1995) (57) dental method of determining CA in a sample of Caucasian children, it was observed that the method significantly underestimated CA by finding statistically significant differences between them (MD=1.2 yrs, SD=1.0-1.4, $p<0.05$).

3.5.2. Precision and Accuracy of Skeletal Method for BA Assessment Among Asian ethnicities Children

3.5.2.1. Precision of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Intra-examiner reliability (Pearson Correlation Coefficient)

The inter-examiner reliability of the GPA method applied to Asian children demonstrated a very high intra-examiner reliability with a Pearson correlation coefficient of $r=0.94$ ($p<0.001$) for a sample of Korean children between 7 and 12 yrs old. (38)

Inter-examiner reliability or concordance (Cohen's kappa coefficient)

Regarding concordance, Chiang and Lin (2005) (40) demonstrated that GPA has excellent inter-observer reliability $k=0.997$ ($p<0.05$) when used to calculate the CA of 10-year-old Taiwanese children.

3.5.2.2. Accuracy of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Mean differences

Regarding accuracy, studies maintain applicability of GPA in this ethnic group although they recognize deviations towards the overestimation of BA with respect to the CA of the child.

The GPA radiological method does not seem to be sufficiently accurate when it is intended to estimate the CA of children. In a sample of X-rays of the carpal and left wrist bones of Chinese children aged 3 to 6 yrs. in Zhejiang Province, Gao et al. (2022) (36) found that the GPA method has an accuracy of 12.02% for boys and 25.76% for girls in determining CA.

In Korean children, GPA on a sample of carpal and left wrist radiographs has been shown to slightly overestimate CA (MD=0.45 mths., SD= 1.79) in children (38).

In addition, Mansourvar et al. (2014) (12) in a retrospective study found that GPA significantly overestimated CA in 4-year-old Malay children with mean differences between CA and BA of 2.3 mths. ($p<0.05$).

In South Korean children, the GPA markedly overestimated the CA of participants under 18 during puberty. According to this work published by Oh et al. (2012), in the case of boys the BA method overestimated 54.6% with respect to their CA and in girls 74.3%. (39) Also, Chiang and Lin (2005) (40) suggested, in line with the above, that when applied to a cohort of Taiwanese girls aged 9 to 17 yrs, the GPA overestimates the CA by 0.18 to 1.48 mths. ($p<0.05$).

In contrast to these results, these same authors demonstrated under the same study conditions as in Taiwanese children aged 13 to 18 yrs, the GPA underestimated the CA between 0.13 and 1.28 yrs ($p<0.05$).

3.5.2.3. Precision of radiographic skeletal method Tanner-Whitehouse 3 (TW3)

Intra-examiner reliability (Pearson Correlation Coefficient)

Regarding repeatability, TW3 demonstrated a very strong intra-examiner agreement as evidenced by a Pearson's correlation coefficient of $r=0.93$ ($p<0.001$) for a sample of Korean boys and girls between 7 and 12 yrs old. (38)

3.5.2.4. Accuracy of radiographic skeletal method Tanner-Whitehouse 3 (TW3)

Mean differences

On the other hand, in samples of radiographs of carpus and left wrist of Chinese children from 3 to 6 yrs of age Gao et al. (2022) (36) published that the TW3 presents a low accuracy of this method with 32.24% for boys and 24.15% for girls.

In South Korean boys and girls, TW3 overestimated the CA of participants under 18 by 59.6% in boys and 72.2% in girls. (39) In this same region, TW3 slightly overestimated the CA of Korean children whose CA was assessed with this method MD=0.45 mths (SD= 1.81 mths) (38).

In a sample of Chinese children, Griffith, Cheng and Wong (2007) (37) observed that when radiographs of the carpus and left wrist of children aged 6 to 18 were analyzed with TW3, it was found that the TW3 statistically significantly overestimated CA compared to the GPA reference method ($p<0.0001$).

3.5.2.5. Precision of radiographic skeletal method Korean Standard Chart (KS)

Intra-examiner reliability (Pearson Correlation Coefficient)

The Korean Standard BA Chart (KS) method studied by Griffith, Cheng and Wong (2007) (37) demonstrated high intra-examiner reliability with a Pearson correlation coefficient of $r=0.94$ ($p<0.001$) for a sample of Korean children aged 7 to 12 yrs.

3.5.2.6. Accuracy of radiographic skeletal method Korean Standard Chart (KS)

Mean differences

Regarding accuracy, Kim, Lee and Yu (2015) showed that in a sample of anteroposterior radiographs of the carpus and left wrist of Korean children, the KS slightly overestimates the CA MD=0.21 mths. (SD= 1.19 mths., $p<0.05$). (38)

3.5.2.7. Accuracy of radiographic skeletal methods RUS-CHN (China 05)

Mean differences

Other specific methods for determining CA for Asian child populations such as RUS-CHN (China 05) had a low accuracy with 12.02% for boys and 21.26% for girls. (36)

3.5.3. Precision and Accuracy of Skeletal Method for BA Assessment Among Indian ethnicities Children

3.5.3.1. Precision of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Intra-examiner reliability (Pearson Correlation Coefficient)

The precision of the GPA is high as it was detected in the study by Patel et al. (2015) (24) in the case of Indian children aged 6 to 16 yrs. living in the region of Gandhinagar (India) an accuracy of 90.65% or what is the same a very strong correlation between CA and BA ($r=0.921$, $p<0.001$). In the case of girls, repeatability is similar reaching a mean value of 89.04% with a Pearson correlation coefficient of $r=0.960$ ($p < 0.001$).

Inter-examiner reliability or concordance (Cohen's kappa coefficient)

The inter-examiner reliability of the GPA method is very high $k= 0.82$ ($p < 0.01$) when looking at the scores provided by different observers after determining CA from a sample of wrist and left-hand radiographs of 10-year-old Thai children (67).

Analyzing the reproducibility of GPA in Hindu children from Mumbai, Keny et al. (2017) (23) showed that this method presents a good agreement between raters with a $k=0.68$ (95% CI = 0.504–0.848, ($p < 0.001$)) but substantially lower than other Indo-European ethnicities.

Inter-examiner reliability or concordance (Intraclass correlation coefficient)

The agreement of the GPA measured from the intra-class correlation coefficient for Malay boys and girls aged between 9 and 18 yrs. was excellent when detecting an inter-examiner reliability $ICC=0.947$ ($p = 0.86$) for boys and somewhat lower for women $ICC= 0.93$ ($p=0.33$).(66)

3.5.3.2. Accuracy of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Mean differences

In terms of accuracy, the studies included in this review acknowledge that the skeletal methods GPA and TW3 slightly overestimate BA when compared to the CA of the Indian child.

In a sample of Indian children between 1 and 15 yrs. old, Keny et al. (2017) (23) found that GPA overestimates in those aged 1 to 6 yrs. at $MD=10$ mths the CA for males. In the same lines, the differences seem to be slightly reduced in the case of girls in whom the differences between CA and BA up to 8 mths.

Similarly, in their work Patil et al. (2012) (25) concluded that GPA overestimates CA in Indian children. In the case of boys aged 8 to 9 yrs. there is a marked difference on average between CA and BA ($MD=2.11$ yrs., $p < 0.05$) while this difference seems to be reduced over the maturation timeframe ($MD = 1.33$ yrs., $p < 0.05$). A similar result has been found in the young girls' population aged 4–8 yrs. ($MD = 0.52$ yrs., $p < 0.05$) and $MD=0.22$ yrs., ($p < 0.05$) at 18 yrs. of CA.

In Hindu children from Eastern Uttar Pradesh aged 1 to 19 yrs., GPA slightly overestimates CA ($MD= 0.56$ mths, $SD= 1.33$ yrs., $p=0.001$). These differences are not homogeneous since for males were $MD=9.03$ mths ($SE= 0.25$, $t = 2.98$, $p \leq 0.05$) and the case of females were noticeably smaller ($MD=4.33$ mths, $SE= 0.18$, $p \leq 0.05$) (27). As children grew, Tiwari et al. (2020) showed that the mean differences decreased slightly upon 0.89 yrs. ($SD= 0.85$ yrs., $p=0.03$), from 0 to 5 yrs., and 0.81 yrs. ($SD= 1.57$ yrs., $p=0.03$) for 0 to 15 yrs.

The accuracy of GPA for determining BA in Malay children in relation to CA is weak as at least one mean difference has been detected that underestimates $MD=0.6$ yrs. (95% CI, $p < 0.05$) in males and $MD=0.7$ yrs. (95% CI, $p < 0.05$) in females. (66)

If we analyze it by age, overestimation of CA has been identified as children grown. Therefore, difference between CA and BA increased from $MD=0.6$ yrs. ($p < 0.05$) for the ages of 13 to 13.9 yrs. to $MD=1.5$ yrs. ($p < 0.05$) in children from 18 to 18.9 yrs. (66)

3.5.3.3. Precision of radiographic skeletal method Tanner-Whitehouse 3 (TW3)

Inter-examiner reliability or concordance (Cohen's kappa coefficient)

The inter-observer concordance reliability of the TW3 RUS method is very high at $k=0.66-0.88$ ($p < 0.01$) in a wrist and left-hand radiograph sample of 10-year-old Thai children (67).

3.5.3.4. Precision of radiographic skeletal method Fishman

Intra-examiner and Inter-examiner reliability or concordance (Cohen's kappa coefficient)

The degree of intra-examiner agreement for the Fishman method of skeletal determination of CA was very good $k = 0.91$ ($p < 0.01$). On the other hand, the inter-observer reliability of the Fishman method of skeletal determination of CA is very good at $k=0.85$ ($p < 0.01$) in 18-year-old Thai children (67).

3.5.3.5. Accuracy of radiographic skeletal method McKay's Method (MK)

Mean differences

Keny et al. (2017) (23) studying the CA accuracy of McKay's skeletal radiological method showed that in Mumbai Indian children aged 1 to 6 yrs. overestimates age by 22 mths for boys and 17 mths for girls.

3.5.3.6. Precision of radiographic dental method of Demirjian

Intra-examiner reliability (Pearson Correlation Coefficient)

The Demirjian's dental method demonstrated great accuracy with a moderate linear correlation coefficient between CA and BA of $r = 0.882$ ($p < 0.001$) in boys and very strong $r = 0.956$ ($p < 0.001$) in the case of girls evaluated with this method of CA. (24)

Mean differences

The accuracy of Demirjian's dental method was analyzed by Patel et al. (2015) (24) who found that it overestimates the CA of Indian children between 6 and 10.99 yrs. ($p > 0.05$) while underestimating somewhat older children during puberty from 11 to 14.99 yrs. ($p > 0.05$).

3.5.3.7. Precision of radiographic dental method of Willem

Intra-examiner reliability (Pearson Correlation Coefficient)

Patel et al. (2015) (24) studied another dental method such as Willem's that presented a high accuracy to detect changes in CA from the age of tooth wear morphology with a correlation coefficient of $r = 0.959$ ($p > 0.05$).

3.5.3.8. Precision of other of radiographic methods

Cervical vertebrae maturation (CVM)

Intra-examiner reliability (Pearson Correlation Coefficient)

In a sample of cervical radiographs of Indian children aged 8 to 14 yrs. from the Andhra Pradesh region, Prasad et al. (2013) (26) observed that the cervical ripening method was highly accurate in detecting possible changes in CA with a linear correlation coefficient of $r = 0.915$ ($p = 0.000$).

Mean differences

The mean difference between the estimated CA of cervical ripening (CMV) studied by Prasad et al. (2013) (26) slightly overestimated the CA of Indian children $MD = 0.097$ yrs. ($SD = 0.793$ yrs., $p > 0.05$). The same author found something more of differences when estimating from BA with TW3 method the CA of children of this ethnic group $MD = 0.170$ yrs. ($SD = 1.08$ yrs., $p > 0.05$).

3.5.4. Precision and Accuracy of Skeletal Method for BA Assessment Among Arab ethnicities Children

3.5.4.1. Precision of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Intra-examiner reliability (Pearson Correlation Coefficient)

The intra-rater reliability of GPA in determining the CA of Arab children ranged from moderate to high. X-rays of Pakistani children from Karachi were strong with a positive linear association of $r = 0.915$ ($p < 0.001$) and $r = 0.943$ ($p < 0.001$) for boys and girls respectively. (32)

On the other hand, in Arab children from Saudi Arabia, the repeatability of the GPA was demonstrated by the identification of a strong association strength between CA and BA $r = 0.873$ ($p < 0.001$) and $r = 0.872$ ($p < 0.001$) for boys and girls, respectively. (34)

In addition, in a sample of young Pakistanis aged 0 and 18 yrs., the GPA showed excellent accuracy in quantifying the association between CA and BA obtained by this method $r = 0.992$ ($p < 0.001$). (33)

However, the correlation between CA and BA estimated by the GPA determination method in native Pakistani children showed a positive and moderate association for both boys and girls $r = 0.778$ ($p < 0.001$). (18,34)

Inter-examiner reliability or concordance (Intraclass correlation coefficient)

The intra-examiner agreement was excellent, reaching $ICC = 0.995$ in boys and $ICC = 0.996$ in girls. (34) The intra-examiner agreement was $ICC = 0.991$ for boys and $ICC = 0.984$ for girls. (35) In a sample of Pakistani children, an excellent intra-examiner concordance was detected, reaching $ICC = 0.998$. (32)

An Israeli study indicated that in a sample of male children from Tel Aviv the method of determining BA, GPA demonstrated an excellent degree of intra-examiner agreement $ICC = 0.9846$.

These results are similar in women as they have quantified an excellent degree of intra-examiner agreement $ICC=0.9787$. (42)

3.5.4.2. Accuracy of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Mean differences

The accuracy of estimates of the GPA skeletal radiographic method among children of Arab ethnicity is controversial as there is no consensus on the results obtained in the studies included in this review.

On the one hand, accuracy for boys aged 4 to 8 yrs. in Saudi Arabia tended to significantly overestimate CA (MD=143.5 mths., SD=44.0, $p<0.001$), while for girls (MD=116.9 mths., SD=41.8, $p<0.001$)(34) Along the same lines, Moradi et al. (2012) (41) in a cross-sectional study found that, in a sample of Iranian boys aged 6 to 18 yrs., GPA overestimated CA more markedly in MD = 0.37 males (SD= 0.98 yrs., $p>0.05$) than in girls MD = 0.04 (SD= 0.78 yrs., $p>0.05$).

Likewise, the accuracy of the GPA to quantify CA in Pakistani children living in Karachi is good although it slightly overestimates MD=0.4 mths. ($p=0.584$) when studied in young population up to 18 yrs. of age. (33)

On the other hand, and contrary to the above, in a similar sample, but with Saudi children aged 10.48 ± 4.8 yrs., Alshamrani et al. (2020) (35) found that GPA underestimated the CA of participants by 4 mths ($p<0.01$).

3.5.4.3. Precision of radiographic skeletal method Tanner-Whitehouse 3 (TW3)

Inter-examiner reliability or concordance (Intraclass correlation coefficient)

The agreement between examiners of the TW3 method in Saudi children was very good, although it differed between sexes, as indicated by a lower intraclass correlation coefficient for men $ICC=0.963$ compared to women $ICC=0.972$. (35)

3.5.4.4. Accuracy of radiographic skeletal method Tanner-Whitehouse 3 (TW3)

Mean differences

In terms of accuracy, the TW3 for samples of children from Saudi Arabia with a mean of 10.21 to 10.48 yrs found an underestimation of CA of 2.5 mths. ($p<0.01$). (35)

3.5.4.5. Precision of other of radiographic methods

Girdany and Golden method

Intra-examiner reliability (Pearson Correlation Coefficient)

If we analyze the precision of Girdany and Golden's method, there is a slight difference between sexes in CA determination finding $r=0.865$ for boys, and a greater correlation between BA and CA for girls, $r=0.909$. (32)

Inter-examiner reliability or concordance (Intraclass correlation coefficient)

In Saudi children, the Girdany and Golden method showed a very high agreement between raters of $ICC=0.974$. (32)

Cervical vertebrae maturation (CVM)

Inter-examiner reliability or concordance (Cohen's kappa coefficient)

In a sample of 14-year-old Turkish children, the inter-examiner agreement of the BA estimation method based on cervical maturity (CVM) was very good at $k=0.862-0.958$ ($p<0.05$). Other methods that studied inter-examiner agreement based on the stages of bone maturation of the hand and wrist (HWM) of Turkish children found that the degree of agreement between the different evaluators of the method was moderate $k=0.812-0.961$ ($p<0.05$). (30)

3.5.4.6. Precision of radiographic dental method of Demirjian

Inter-examiner reliability or concordance (Cohen's kappa coefficient)

On the other hand, in this same sample of Turkish children, the Demirjian dental method of determining BA obtained a moderate degree of inter-examiner agreement $k= 0.823-0.928$ ($p<0.05$). (30)

3.5.5. Precision and Accuracy of Skeletal Method for BA Assessment Among Hispanic ethnicities Children

3.5.5.1. Precision of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Intra-examiner reliability (Pearson Correlation Coefficient)

Regarding the accuracy of the GPA method, a strong and positive linear correlation was found $r=0.890$ ($p<0.001$) that associated the increase in CA with an increase in BA scores obtained by this method of determining GPA in Hispanic children residing in the United States.(68) As such, the accuracy of the GPA determination method was very good, with an association coefficient between CA and GPA score of $r=0.918$ ($p<0.05$) found in a sample of antero-posterior radiographs of the left hand and wrist of Venezuelan children aged 6 to 12 yrs.(69)

In Chilean children under 16 yrs. of age, the manual GPA score against an automated expert system, Pose et al. (2018) (71) found a very strong and positive linear correlation ranging from $r=0.91-0.93$ ($p<0.05$) which would indicate the reliability of the procedure even when using machine learning.

Inter-examiner reliability or concordance (Intraclass correlation coefficient)

In an American study conducted by Calfee et al. (2010) (68) with a sample of Hispanic children aged 12 to 18 yrs., an excellent inter-examiner reliability $ICC= 0.982$ for the GPA method was found.

3.5.5.2. Accuracy of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Mean differences

The accuracy of GPA from a sample of left hand and wrist radiographs from Hispanic children is consistent across the various studies included in this review.

In Hispanic children aged 9.96 to 11.12 yrs. Pose et al. (2018) (71) found GPA underestimated CA (MD=0.19 yrs., 95%CI: 0.13-0.25, $p<0.05$) in a large sample of children treated in an orthopedic clinic in Santiago de Chile. In the case of Mansourvar et al. (2014) whom applied this method in sample of Hispanic children aged 15 to 18 yrs. living in California (USA) also detected an underestimation of CA MD = 0.094 yrs. (95% CI, $p>0.05$) (12).

3.5.5.3. Precision of radiographic skeletal method Tanner-Whitehouse 3

Intra-examiner reliability (Pearson Correlation Coefficient)

Regarding the TW3 RUS method, López et al. (2008) (70) found in a sample of Venezuelan children a high accuracy in children aged 7 to 14 yrs. $r=0.91$ ($p<0.05$). Similar results were also obtained for girls in this sample with a Pearson correlation coefficient of $r=0.93$ ($p<0.05$). As for TW3 Carpal that evaluates the regions of interest of the carpal bones, a lower accuracy was found $r=0.89$ ($p < 0.05$) in boys and $r=0.82$ ($p<0.05$) for radiographs of girls.

3.5.5.4. Precision of radiographic dental method of Demirjian

Intra-examiner reliability (Pearson Correlation Coefficient)

The accuracy of the Demirjian dental method found a very strong correlation coefficient $r=0.929$ ($p<0.05$) that associated CA with dental age obtained with the Demirjian method in a sample of Venezuelan children from Maracaibo in the State of Zulia between 6 and 12 yrs old. (69)

3.5.6. Precision and Accuracy of Skeletal Method for BA Assessment Among African ethnicities Children

3.5.6.1. Precision of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Intra-examiner reliability (Pearson Correlation Coefficient)

When intra-examiner agreement was studied, a very strong correlation was observed between CA and GPA scores in boys $r=0.93$ ($p>0.05$) and girls $r=0.94$ ($p>0.05$) in the central district of Botswana. (62)

With respect to the degree of intra-observer agreement, the GPA presents a very strong correlation in radiographic samples of boys $r=0.96$ ($p<0.05$) and girls from Zimbabwe $r=0.96$ ($p<0.05$). (61)

In South Africa, in the prospective cohort study involving young Bantu people, Dembetembe et al. (2012) (59) found a moderate correlation when analyzing the accuracy of the GPA method of $r=0.76$. However, intra-examiner reliability decreased to a zero linear correlation $r=0.02$ when individuals between 13 and 18.5 yrs. of age were examined.

Inter-examiner reliability or concordance (Intraclass correlation coefficient)

When studying the inter-examiner concordance of the GPA method in South African children, an intraclass coefficient $ICC=0.99$ was found with a statistical significance of $p<0.001$ (60) In Botswana, the intra-observer agreement of the GPA with left hand-wrist radiographs of male children between 5 and 18 yrs. also registered an excellent $ICC=0.97$ ($p>0.05$), being somewhat higher for girls $ICC=0.98$ ($p>0.05$). (62)

3.5.6.2. Accuracy of radiographic skeletal method Greulich and Pyle Atlas (GPA)

Mean differences

The accuracy of GPA and TW3 among children of African ethnicity is controversial because it is the ethnic group in which these radiological methods present the greatest overestimation.

The accuracy of GPA for determining BA in relation to CA is weak in South African children as at least one mean difference overestimates $MD=7.4$ mths. ($SD=15.7$ mths., $p<0.05$) has been detected. (60)

In African males aged up to 19 yrs., the mean differences are even greater overestimated than reported by previous studies, finding a CA of $MD=4.4 \pm 14.5$ mths. (95% CI, $p<0.05$). In the case of African females up to 18 yrs. the mean differences are $MD = 2.4 \pm 12.8$ mths. (95% CI, $p<0.05$). (60)

Another study notes that the GPA method overestimates the age of African pubertal males from Zimbabwe finding important differences between CA and BA ($MD= 0.76$ yrs, 95% CI: $-0.95, -0.57$, $p<0.05$) (61) Similar results were found in a study conducted in Botswana, when they published that the differences between CA and BA after using the GPA were accentuated as the CA increased from ($MD=0.25$ yrs., $p<0.05$) from 5 to 10 yrs. to ($MD=0.94$ yrs, 95% CI, $p<0.05$) in children aged between 15 and 18 yrs. (62)

In Ethiopia, it has also been detected that GPA overestimated CA for males ($MD = 8.7$ mths. ($p<0.05$) and for females ($MD = 11.8$ mths. ($p<0.05$) between 10 to 22 yrs. (63) Even in the African-American population, the GPA method overestimated CA at 15 yrs. of age ($MD=2.4$ yrs., 95% CI, $p>0.05$). (12)

3.5.6.3. Precision of radiographic skeletal method Tanner-Whitehouse 3

Intra-examiner reliability (Pearson Correlation Coefficient)

The TW3 RUS intra-examiner reliability is similar to that found in GPA presenting a strong to very strong correlation for boys $r=0.95$ ($p<0.05$) and for girls $r=0.93$ ($p<0.05$) Zimbabweans. (61)

3.5.6.4. Accuracy of radiographic skeletal method Tanner-Whitehouse 3

Mean differences

When studying the accuracy of this method for age determination in Zimbabwean children, overestimation was detected means ($MD=-0.43$ yrs., 95% CI: $- 0.61, -0.24$, $p<0.05$). (61)

4. Discussion

Radiological BA Assessment such as GPA or TW3 were generally accepted as precise for all ethnic groups as evidenced by the results of intra-examiner reliability estimated in most studies from

Pearson's linear correlation coefficient.(73) Also, it has been demonstrated the excellent inter-examiner concordance of radiological methods of determination of BA for the ethnicities analyzed from calculation of Cohen's kappa coefficient (k), intraclass correlation coefficient (ICC) or Lin's concordance (ρ_c).

Although the results justify its use for the determination of BA as part of pediatric professional performance and in legal medicine, we must not ignore the fact that currently the determination of BA based on the score of milestones of skeletal development through an X-ray may involve inadequate decision-making due to the existence of racial biases that alter its interpretation. (74)

The intra-examiner reliability of GPA for determining BA in Caucasian children was moderate to excellent in the included studies, indicating high repeatability.

GPA precision for British Caucasian children (43) and Southwestern Australian, with British ancestors (64), show a high concordance. When the agreement among observers is measured only with the Cohen's Kappa Coefficient it is observed that for Caucasians from Israel (42) and Portugal(48) it is weaker compared to the previous ones. This could be explained because the Caucasian American population of Northern European origin that Greulich and Pyle used to prepare their atlas differs from Mediterranean Caucasians.

When it is analyzed by sex, differences are found in intraclass correlation coefficient of the GPA. Based on the study by Alshamrani et al. (2020b) (43), this coefficient is higher in boys than in girls, which in principle would suggest that GPA is more accurate in boys than in girls. These results are also consistent with the findings provided by Nang et al. (2023) (66) for Indian children, however these results contrast with those obtained by several authors for the Arab (32,34,35) and African (62) population in which the agreement between observers was stronger in girls than boys. This could be due, among other causes, to the differences between the study samples, finding that, unlike Alshamrani et al. (2020b) (43) whose girls were on average 8.8 yrs. (SD=3.6), had included older girls who, by presenting more ossification points, allowed increasing the precision of the measurement.

Regarding accuracy, we can point out that, in general, the application of GPA to the Caucasian population generates an underestimation with respect to CA. We found that the underestimation is lower in the study by Mansouvar et al. (2014) (12) carried out on a child population from the Children's Hospital Los Angeles (United States), which is coherent considering that the reference population of the GPA is also American (45). Likewise, the underestimation of GPA continues to be low for the population of Scotland, although in this case the interval of difference in means is greater. This could be explained by the existence of common ancestors between the Scottish population and the northern American population from which the Greulich and Pyle method originates.

Conversely, the underestimation of chronological age is much more pronounced in Caucasian children originating from Central European (54,55,58) and Scandinavian countries.(57) If we analyze it by sex, we observe that the underestimation in girls is lower in the Turkish population (31) compared to the German population (55,58), in which the greatest deviation is observed. In the case of male children, Turkish children have the least underestimation (31), while Austrians have the most downward deviation (47) On the contrary, the GPA method, and also Tanner-Whitehouse 2 and 3, produces an overestimation in Spanish children's CA (47) as well as Scottish children between the ages of 9 and 17, which could be due to a particular growth pattern of this population.(45)

Although the precision of GPA measured through intra- and inter-examiner reliability is strong in the Asian population, regarding accuracy, we found an overestimation of chronological age in this ethnic group. The overestimation is highest in the population of Malaysian (12) children with South Korea (38) being the smallest. By age, in the Taiwanese population (40), the overestimation is highest in children between 9 and 17 years old and slightly lower for children between 13 and 18 years old. In an analysis by sex, GPA tends to overestimate Korean girls (39) more and Chinese boys less. (36)

In Indians, the GPA application produces an overestimation of CA as it happens with precision of Demirjian's Dental Method and Cervical Vertebrae Maturation (CVM). This overestimation is higher in the Hindus in the study by Patil et al. (2012) (25) while in children from the Eastern Uttar Pradesh region the method presents one of the lowest overestimations. (27) In the case of the Indian population, we also observe that due to the size of the sample, this method produces an

overestimation depending on the age period to which we refer, with said mean difference being greater in children aged 0-5 and 0- 15 years (27) and younger in the range 4-8 yrs.(25) The overestimation of CA is generally greater in boys than in girls, with sexual differences between children in eastern Uttar Pradesh being especially notable.(27) It stands out, in turn, that only in the group of girls this overestimation is minimized when they reach 18 years of age. (25) Regarding the Malay ethnicity, the GPA overestimates the CA of those who have reached the age of majority, doubling the overestimation of children aged 13 to 13.9 yrs. (66) In this population, unlike what happens in Hindu children, there is an underestimation of chronological age for both sexes, which is significantly higher for girls of this ethnic group. (66)

In the Pakistani population there is discrepancy regarding the precision of the GPA, given the differences found in intra-examiner reliability. (18,34) It is also observed that when analyzing this same parameter but using the Girdany and Golden method, a greater strength of correlation is found between girls than boys.(32) Furthermore, with respect to accuracy, we observe that the test significantly overestimates the CA of children from Saudi Arabia (34), however, Alshamrani et al. (2020) found that GPA, as it happens when TW3 is applicated, underestimates CA. (35) These results are in line with those obtained from a sample of wrist and left-hand radiographs from Pakistan in which overestimation is minimal. (33) In Iranian Arab children, this overestimation is significantly lower than the reported by Albaker et al. (2021) for Saudi people, and in this case, there is a marked difference between the overestimation of CA in boys, which is 10 times higher than that in girls. (41)

The accuracy of GPA from a sample of left hand and wrist radiographs from Hispanic children is consistent across the various studies included in this review. As age increases, the underestimation of the GPA method decreases (12,71). If we analyze by sex, the TW3 RUS application is more accurate than the TW3 Carpal applied to radiography of girls.

Furthermore, almost half of the included studies (n=23, 45.90%) have been carried out with samples of Caucasian children while on the opposite side we find the studies carried out in African children (n=8, 4.18%).

In general, the accuracy when applying the GPA and TW3 is low. This is attributed to the overestimation that is generated after their application to the African population. African adolescents living in the United States are those with the greatest overestimation(12) compared to residents of Botswana (62) although the former has a small sample of x-rays, which makes comparability difficult.

As age increases, the difference in means with respect to CA seems to increase, with greater overestimation found among adolescents than among children.(12,60–63) If we analyze it by sex, this previously described phenomenon is maintained, seeing that as the samples include older boys and girls, the overestimation of CA increases.(60,63) Specifically, we observed that by ethnic group, it is children with African ancestors who generate the greatest overestimation when applying radiological bone age diagnostic methods. We also noticed that in this population there is a deviation with respect to the average age, which would be attributable to the strikingly low number of radiographs that make up the sample as well as the wide spectrum of ages of the African children included in these studies. (60)

Regarding the size of the samples, in general it was small and biased, since they were obtained in a unicentric and non-randomized manner with sequential or intentional sampling. This limitation of the sampling technique requires us to carefully consider the results presented, as the external validity of this review could be affected.

Regarding the cross-sectional design used, in many of the included publications, it only allows identifying the main milestones of skeletal maturation but not the temporal sequence or the spatial references that ossification takes over time.

Another of the limitations of these studies is that they do not allow us to know if the deviation of the BA measurement with respect to CA observed in each and every one of the ethnicities is really significant from a clinical point of view. This encourages disinterest among the recipients of the tools for radiological determination of the child's age, who are essentially Legal Medicine and Pediatric professionals.

With this design, it is not clear to us the level of organization to which these accuracy errors can be attributed, that is, whether they are the product of the individual, the family, or the ethnic group. What does seem evident is that the result of maturation, and consequently that of ossification, is an extremely and at the same time unknown complex phenomenon in which environmental, hormonal and genetic factors occur and whose interaction would explain the differences in accuracy found in radiological bone age determination tools.

4.1. Limitations

In our study there is a set of methodological limitations that may alter the external validity of the results presented. In relation to the study design of the included studies, we must note the existence of a high proportion of cross-sectional studies $n=19$ (37.25%) which could become a limitation by preventing the study of the maturational changes of the bone throughout the period time. As a consequence of these designs, the results obtained on the metric properties of radiological BA assessment methods may vary significantly.

In relation to the sampling of the studies we observe some relevant aspects. Firstly, the sampling techniques chosen are, on the one hand, non-consecutive non-probabilistic sampling and, on the other hand, convenience sampling. Furthermore, in some ethnic groups, there is an imbalance in the sample size in favor of the number of male children recruited compared to that of girls, which could imply a gender bias that would alter the interpretation of the metric properties of the radiological methods. (75)

Moreover, almost half of the included studies ($n=23$, 45.90%, $n=9,777$ wrist-carpal radiographs) have been carried out with samples of Caucasian children while on the opposite side we find the studies carried out in African children ($n=8$, 4.18%, $n=810$ wrist-carpal radiographs). This imbalance could lead to interpretation biases by not having homogeneous groups that allow comparability of the precision of the methods. (76)

In addition to the above, poor sampling robustness is especially important when the authors analyze the consistency of the method in multiethnic samples such as African Americans or Hispanic Americans. With respect to this tenor, we consider that the estimation of BA in these groups in which there is no dominant ethnicity is necessarily inaccurate. (77)

5. Conclusions

Skeletal radiographic methods GPA and TW3 are both precise for BA determination among all ethnical groups, but their accuracy in estimating CA can be altered by racial bias.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Table S1. Characteristics of included studies.

Author Contributions: Theoretical conceptualization, F.R.H., J.V.G., C.L.H.; Methodology, F.R.H., M.H.P.; literature searching, I.M.P., S.M.P., Data analysis, I.M.P., S.M.P., Writing—Original Draft, I.M.P., S.M.P.; review, M.H.P., R.M.S.; Supervision, F.R.H.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Complejo Hospitalario Universitario de Canarias (CHUC_2023_86—07/13/2023).

Data Availability Statement: Record data for this review is available at: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023449512.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Martin DD, Calder AD, Ranke MB, Binder G, Thodberg HH. Accuracy and self-validation of automated bone age determination. *Sci Rep.* 2022;12(1). doi:10.1038/s41598-022-10292-y

2. Mughal AM, Hassan N, Ahmed A. Bone age assessment methods: A critical review. *Pak J Med Sci.* 2014;30(1). doi:10.12669/pjms.301.4295
3. Phudowski P, Lebidowski M, Lorenc RS. Evaluation of practical use of bone age assessments based on DXA-derived hand scans in diagnosis of skeletal status in healthy and diseased children. *Journal of Clinical Densitometry.* 2005;8(1). doi:10.1385/JCD:8:1:048
4. Makkad R, Balani A, Chaturvedi S, Tanwani T, Agrawal A, Hamdani S. Reliability of panoramic radiography in chronological age estimation. *J Forensic Dent Sci.* 2013;5(2). doi:10.4103/0975-1475.119781
5. Lee BD, Lee MS. Automated Bone Age Assessment Using Artificial Intelligence: The Future of Bone Age Assessment. *Korean J Radiol.* 2021;22(5). doi:10.3348/kjr.2020.0941
6. Chaumoitre K, Saliba-Serre B, Adalian P, Signoli M, Leonetti G, Panuel M. Forensic use of the Greulich and Pyle atlas: prediction intervals and relevance. *Eur Radiol.* 2017;27(3). doi:10.1007/s00330-016-4466-4
7. Prokop-Piotrkowska M, Marszałek-Dziuba K, Moszczyńska E, Szalecki M, Jurkiewicz E. Traditional and new methods of bone age assessment-an overview. *JCRPE Journal of Clinical Research in Pediatric Endocrinology.* 2021;13(3). doi:10.4274/jcrpe.galenos.2020.2020.0091
8. Vignolo M, Milani S, Cerbello G, Coroli P, Di Battista E, Aicardi G. FELS, Greulich-Pyle, and Tanner-Whitehouse bone age assessments in a group of Italian children and adolescents. *American Journal of Human Biology.* 1992;4(4). doi:10.1002/ajhb.1310040408
9. Nahhas RW, Sherwood RJ, Chumlea WC, Duren DL. An update of the statistical methods underlying the FELS method of skeletal maturity assessment. *Ann Hum Biol.* 2013;40(6). doi:10.3109/03014460.2013.806591
10. Chumela WC, Roche AF, Thissen D. The FELS method of assessing the skeletal maturity of the hand-wrist. *American Journal of Human Biology.* 1989;1(2). doi:10.1002/ajhb.1310010206
11. Alshamrani K, Messina F, Offiah AC. Is the Greulich and Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis. *Eur Radiol.* 2019;29(6). doi:10.1007/s00330-018-5792-5
12. Mansourvar M, Ismail MA, Raj RG, et al. The applicability of Greulich and Pyle atlas to assess skeletal age for four ethnic groups. *J Forensic Leg Med.* 2014;22. doi:10.1016/j.jflm.2013.11.011
13. Cao F, Huang HK, Pietka E, Gilsanz V. Digital hand atlas and web-based bone age assessment: System design and implementation. *Computerized Medical Imaging and Graphics.* 2000;24(5). doi:10.1016/S0895-6111(00)00026-4
14. Grave KC, Brown T. Skeletal ossification and the adolescent growth spurt. *Am J Orthod.* 1976;69(6). doi:10.1016/0002-9416(76)90143-3
15. Ashizawa K, Asami T, Anzo M, et al. Standard RUS skeletal maturation of Tokyo children. *Ann Hum Biol.* 1996;23(6). doi:10.1080/03014469600004682
16. Mohammed RB, Krishnamraju P V., Prasanth PS, Sanghvi P, Reddy MAL, Jyotsna S. Dental age estimation using Willems method: A digital orthopantomographic study. *Contemp Clin Dent.* 2014;5(3). doi:10.4103/0976-237X.137954
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. *Forensic Sci Int.* 2005;154(1). doi:10.1016/j.forsciint.2004.08.018
18. Mughal AM, Hassan N, Ahmed A. The applicability of the Greulich & Pyle Atlas for bone age assessment in primary school-going children of Karachi, Pakistan. *Pak J Med Sci.* 2014;30(2). doi:10.12669/pjms.302.4296
19. Wang X, Zhou B, Gong P, et al. Artificial Intelligence-Assisted Bone Age Assessment to Improve the Accuracy and Consistency of Physicians With Different Levels of Experience. *Front Pediatr.* 2022;10. doi:10.3389/fped.2022.818061
20. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting - Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group B. *JAMA Neurol.* 2000;283.
21. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses. (Available from: URL: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp). Published online 2012. doi:10.2307/632432
22. Bero L, Chartres N, Diong J, et al. The risk of bias in observational studies of exposures (ROBINS-E) tool: Concerns arising from application to observational studies of exposures. *Syst Rev.* 2018;7(1). doi:10.1186/s13643-018-0915-2

23. Keny SM, Sonawane D V., Pawar E, et al. Comparison of two radiological methods in the determination of skeletal maturity in the Indian pediatric population. *Journal of Pediatric Orthopaedics Part B*. 2018;27(4). doi:10.1097/BPB.0000000000000475
24. Patel P, Chaudhary A, Dudhia B, Bhatia P, Jani Y, Soni N. Accuracy of two dental and one skeletal age estimation methods in 6-16 year old Gujarati children. *J Forensic Dent Sci*. 2015;7(1). doi:10.4103/0975-1475.150298
25. Patil ST, Parchand MP, Meshram MM, Kamdi NY. Applicability of Greulich and Pyle skeletal age standards to Indian children. *Forensic Sci Int*. 2012;216(1-3). doi:10.1016/j.forsciint.2011.09.022
26. Krishna Prasad CMS, Reddy VN, Sreedevi G, Ponnada SR, Padma Priya K, Raveendra Naik B. Objective evaluation of cervical vertebral bone age-its reliability in comparison with hand-wrist bone age: By TW3 method. *Journal of Contemporary Dental Practice*. 2013;14(5). doi:10.5005/jp-journals-10024-1407
27. Tiwari PK, Gupta M, Verma A, Pandey S, Nayak A. Applicability of the Greulich–Pyle Method in Assessing the Skeletal Maturity of Children in the Eastern Utter Pradesh (UP) Region: A Pilot Study. *Cureus*. Published online 2020. doi:10.7759/cureus.10880
28. Büken B, Şafak AA, Yazici B, Büken E, Mayda AS. Is the assessment of bone age by the Greulich-Pyle method reliable at forensic age estimation for Turkish children? *Forensic Sci Int*. 2007;173(2-3). doi:10.1016/j.forsciint.2007.02.023
29. Cantekin K, Celikoglu M, Miloglu O, Dane A, Erdem A. Bone Age Assessment: The Applicability of the Greulich-Pyle Method in Eastern Turkish Children. *J Forensic Sci*. 2012;57(3). doi:10.1111/j.1556-4029.2011.02035.x
30. Magat G, Ozcan S. Assessment of maturation stages and the accuracy of age estimation methods in a Turkish population: A comparative study. *Imaging Sci Dent*. 2022;52. doi:10.5624/ISD.20210231
31. Öztürk F, Karataş OH, Mutaf HI, Babacan H. Bone age assessment: comparison of children from two different regions with the Greulich–Pyle method In Turkey. *Australian Journal of Forensic Sciences*. 2016;48(6). doi:10.1080/00450618.2015.1119311
32. Awais M, Nadeem N, Husen Y, Rehman A, Beg M, Khattak YJ. 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*. 2014;24(12).
33. 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(7).
34. Albaker AB, Aldhilan AS, Alrabai HM, et al. Determination of Bone Age and its Correlation to the Chronological Age Based on the Greulich and Pyle Method in Saudi Arabia. *J Pharm Res Int*. Published online 2021. doi:10.9734/jpri/2021/v33i60b34731
35. Alshamrani K, Hewitt A, Offiah AC. Applicability of two bone age assessment methods to children from Saudi Arabia. *Clin Radiol*. 2020;75(2). doi:10.1016/j.crad.2019.08.029
36. Gao C, Qian Q, Li Y, et al. A comparative study of three bone age assessment methods on Chinese preschool-aged children. *Front Pediatr*. 2022;10. doi:10.3389/fped.2022.976565
37. Griffith JF, Cheng JCY, Wong E. Are western skeletal age standards applicable to the Hong Kong Chinese population? A comparison of the Greulich and Pyle method and the tanner and whitehouse method. *Hong Kong Medical Journal*. 2007;13(3).
38. Kim JR, Lee YS, Yu J. 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*. 2015;16(1). doi:10.3348/kjr.2015.16.1.201
39. Oh Y, Lee R, Kim HS. Evaluation of skeletal maturity score for Korean children and the standard for comparison of bone age and chronological age in normal children. *Journal of Pediatric Endocrinology and Metabolism*. 2012;25(3-4). doi:10.1515/jpem-2012-0028
40. Chiang KH, Chou AS Bin, Yen PS, et al. The reliability of using Greulich-Pyle method to determine children's bone age in Taiwan. *Tzu Chi Med J*. 2005;17(6).
41. Moradi M, Sirous M, Morovatti P. The reliability of skeletal age determination in an Iranian sample using Greulich and Pyle method. *Forensic Sci Int*. 2012;223(1-3). doi:10.1016/j.forsciint.2012.08.030
42. Soudack M, Ben-Shlush A, Jacobson J, Raviv-Zilka L, Eshed I, Hamiel O. Bone age in the 21st century: UIs Greulich and Pyle's atlas accurate for Israeli children? *Pediatr Radiol*. 2012;42(3). doi:10.1007/s00247-011-2302-1

43. Alshamrani K, Offiah AC. Applicability of two commonly used bone age assessment methods to twenty-first century UK children. *Eur Radiol.* 2020;30(1). doi:10.1007/s00330-019-06300-x
44. Bull RK, Edwards PD, Kemp PM, Fry S, Hughes IA. Bone age assessment: A large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods. *Arch Dis Child.* 1999;81(2). doi:10.1136/adc.81.2.172
45. Hackman L, Black S. The reliability of the greulich and pyle atlas when applied to a modern scottish population. *J Forensic Sci.* 2013;58(1). doi:10.1111/j.1556-4029.2012.02294.x
46. Alcina M, Lucea A, Salicrú M, Turbón D. Reliability of the Greulich and Pyle method for chronological age estimation and age majority prediction in a Spanish sample. *Int J Legal Med.* 2018;132(4). doi:10.1007/s00414-017-1760-x
47. Ebrí B. Comparative study between bone ages: Carpal, Metacarpophalangeal, Carpometacarpophalangeal Ebrí, Greulich and Pyle and Tanner Whitehouse2. *Med Res Arch.* 2021;9(12). doi:10.18103/mra.v9i12.2625
48. Martinho D V., Coelho-e-Silva MJ, Valente-dos-Santos J, et al. Assessment of skeletal age in youth female soccer players: Agreement between Greulich-Pyle and Fels protocols. *American Journal of Human Biology.* 2022;34(1). doi:10.1002/ajhb.23591
49. Santos C, Ferreira M, Alves FC, Cunha E. Comparative study of Greulich and Pyle Atlas and Matusos 4.0 program for age estimation in a Portuguese sample. *Forensic Sci Int.* 2011;212(1-3). doi:10.1016/j.forsciint.2011.05.032
50. Pinchi V, De Luca F, Ricciardi F, et al. Skeletal age estimation for forensic purposes: A comparison of GP, TW2 and TW3 methods on an Italian sample. *Forensic Sci Int.* 2014;238. doi:10.1016/j.forsciint.2014.02.030
51. Santoro V, Roca R, De Donno A, et al. Applicability of Greulich and Pyle and Demirjian aging methods to a sample of Italian population. *Forensic Sci Int.* 2012;221(1-3). doi:10.1016/j.forsciint.2012.04.014
52. Martrille L, Papadodima S, Venegoni C, et al. Age Estimation in 0–8-Year-Old Children in France: Comparison of One Skeletal and Five Dental Methods. *Diagnostics.* 2023;13(6). doi:10.3390/diagnostics13061042
53. Zabet D, Rérolle C, Pucheux J, Telmon N, Saint-Martin P. Can the Greulich and Pyle method be used on French contemporary individuals? *Int J Legal Med.* 2015;129(1). doi:10.1007/s00414-014-1028-7
54. 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. *British Journal of Radiology.* 1999;72(MAY). doi:10.1259/bjr.72.857.10505010
55. Schmidt S, Koch B, Schulz R, Reisinger W, Schmeling A. 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.* 2007;121(4). doi:10.1007/s00414-007-0165-7
56. van Rijn RR, Lequin MH, Robben SGF, Hop WCJ, van Kuijk C. Is the Greulich and Pyle atlas still valid for Dutch Caucasian children today? *Pediatr Radiol.* 2001;31(10). doi:10.1007/s002470100531
57. Kullman L. Accuracy of two dental and one skeletal age estimation method in Swedish adolescents. *Forensic Sci Int.* 1995;75(2-3). doi:10.1016/0379-0738(95)01792-5
58. 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(2). doi:10.1080/03014468400007041
59. Dembetembe KA, Morris AG. Is Greulich-Pyle age estimation applicable for determining maturation in male Africans? *S Afr J Sci.* 2012;108(9-10). doi:10.4102/sajs.v108i9/10.1036
60. Govender D, Goodier M. Bone of contention: The applicability of the Greulich- Pyle method for skeletal age assessment in South Africa. *South African Journal of Radiology.* 2018;22(1). doi:10.4102/SAJR.V22I1.1348
61. Kowo-Nyakoko F, Gregson CL, Madanhire T, et al. Evaluation of two methods of bone age assessment in peripubertal children in Zimbabwe. *Bone.* 2023;170. doi:10.1016/j.bone.2023.116725
62. Olaotse B, Norma PG, Kaone PM, et al. Evaluation of the suitability of the Greulich and Pyle atlas in estimating age for the Botswana population using hand and wrist radiographs of young Botswana population. *Forensic Science International: Reports.* 2023;7. doi:10.1016/j.fsir.2023.100312
63. Tsehay B, Afework M, Mesifin M. Assessment of Reliability of Greulich and Pyle (GP) Method for Determination of Age of Children at Debre Markos Referral Hospital, East Gojjam Zone. *Ethiop J Health Sci.* 2017;27(6). doi:10.4314/ejhs.v27i6.8
64. Maggio A, Flavel A, Hart R, Franklin D. 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.* 2018;50(4). doi:10.1080/00450618.2016.1251970

65. Paxton ML, Lamont AC, Stillwell AP. The reliability of the Greulich-Pyle method in bone age determination among Australian children. *J Med Imaging Radiat Oncol*. 2013;57(1). doi:10.1111/j.1754-9485.2012.02462.x
66. Nang KM, Ismail AJ, Tangaperumal A, et al. Forensic age estimation in living children: how accurate is the Greulich-Pyle method in Sabah, East Malaysia? *Front Pediatr*. 2023;11. doi:10.3389/fped.2023.1137960
67. Benjavongkulchai S, Pittayapat P. Age estimation methods using hand and wrist radiographs in a group of contemporary Thais. *Forensic Sci Int*. 2018;287. doi:10.1016/j.forsciint.2018.03.045
68. Calfee RP, Sutter M, Steffen JA, Goldfarb CA. Skeletal and chronological ages in American adolescents: Current findings in skeletal maturation. *J Child Orthop*. 2010;4(5). doi:10.1007/s11832-010-0289-z
69. Tineo F, Espina de Ferreira A, Barrios F, Ortega A, Ferreira J. Estimación de la edad cronológica con fines forenses, empleando la edad dental y la edad ósea en niños escolares en maracaibo, estado zulía. *Acta Odontol Venez*. 2006;44(2).
70. López P, Morón A, Urdaneta O. Maduración ósea de niños escolares (7-14 años) de las etnias Wayúu y Criolla del Municipio Maracaibo, Estado Zulia. Estudio Comparativo. *Ciencia Odontológica*. 2020;5(2):99-111. Accessed August 16, 2023. <https://produccioncientificaluz.org/index.php/cienciao/article/view/33940>
71. Pose Lepe G, Villacrés F, Fuente-Alba CS, Guiloff S. Correlation in radiological bone age determination using the Greulich and Pyle method versus automated evaluation using BoneXpert software. *Rev Chil Pediatr*. 2018;89(5). doi:10.4067/S0370-41062018005000705
72. Griffith JF. Musculoskeletal complications of severe acute respiratory syndrome. *Semin Musculoskelet Radiol*. 2011;15(5). doi:10.1055/s-0031-1293500
73. De Sanctis V, Di Maio S, Soliman A, Raiola G, Elalaily R, Millimaggi G. Hand X-ray in pediatric endocrinology: Skeletal age assessment and beyond. *Indian J Endocrinol Metab*. 2014;18. doi:10.4103/2230-8210.145076
74. Grave K. The use of the hand and wrist radiograph in skeletal age assessment; and why skeletal age assessment is important. *Aust Orthod J*. 1994;13(3).
75. Wingerd J, Peritz E, Sproul A. Race and stature differences in the skeletal maturation of the hand and wrist. *Ann Hum Biol*. 1974;1(2). doi:10.1080/03014467400000211
76. Loder RT, Estle DT, Morrison K, et al. Applicability of the Greulich and Pyle Skeletal Age Standards to Black and White Children of Today. *American Journal of Diseases of Children*. 1993;147(12). doi:10.1001/archpedi.1993.02160360071022
77. Ontell FK, Ivanovic M, Ablin DS, Barlow TW. Bone age in children of diverse ethnicity. *American Journal of Roentgenology*. 1996;167(6). doi:10.2214/ajr.167.6.8956565

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.