osteology

MDPI

Article

The Validation of the Tanner–Whitehouse 3 Method for Radiological Bone Assessments in a Pediatric Population from the Canary Islands

Sebastián Eustaquio Martín Pérez [1,2,3,4], Isidro Miguel Martín Pérez [1,*], Ruth Molina Suárez [5], Jesús María Vega González [6] and Alfonso Miguel García Hernández [1]

[1] Escuela de Doctorado y Estudios de Posgrado, Universidad de La Laguna, 38203 Santa Cruz de Tenerife, Spain; sebastian.martin@universidadeuropea.es (S.E.M.P.); almigar@ull.es (A.M.G.H.)
[2] Faculty of Health Sciences, Universidad Europea de Canarias, 38300 Santa Cruz de Tenerife, Spain
[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] Pediatric Endocrinology Unit, Pediatric Department, Hospital Universitario de Canarias, 38320 Santa Cruz de Tenerife, Spain; rmolsua@gobiernodecanarias.org
[6] Institute of Legal Medicine and Forensic Sciences of Santa Cruz de Tenerife, 38230 San Cristóbal de La Laguna, Spain; jveggonf@justiciacanarias.org
* Correspondence: alu0100713374@ull.edu.es

Abstract: Background/Objectives: Bone age assessments are essential for evaluating the growth and skeletal development of children and adolescents, influenced by factors such as genetics, ethnicity, culture, and nutrition. Clinical standards for these assessments must be adapted to the specific populations under study. This study validates the use of the Tanner–Whitehouse 3 method for determining bone age in pediatric and adolescent populations in the Canary Islands. **Methods**: This cross-sectional study analyzed 214 posteroanterior radiographs of the left hand and wrist from 80 females and 134 males, aged between 5 and 18 years. The radiographs were independently evaluated by three raters: a Radiologist Specialist (Rater 1), a General Practitioner (Rater 2), and a Medical Student (Rater 3). Intra- and inter-rater reliability were assessed using intraclass correlation coefficients (ICCs). Accuracy was evaluated by comparing estimated bone age with chronological age, stratified by sex and developmental stage. **Results**: Strong intra-rater reliability was observed across all raters. Raters 1 and 2 demonstrated excellent consistency (ICCs: 0.990–0.996), while Rater 3 exhibited slightly lower yet robust reliability (ICCs: 0.921–0.976). Inter-rater agreement was high between Raters 1 and 2 but decreased with Rater 3, reflecting the influence of experience (ICCs: 0.812–0.912). Bone age was underestimated in preschool children (mean difference: 3.712 mos.; 95% CI: 1.290–6.130; $p = 0.199$) and school-age males (mean difference: 3.978 mos.; 95% CI: −12.550 to 4.590; $p = 0.926$), with minimal discrepancies in females. In teenagers, the Tanner–Whitehouse 3 method slightly overestimated bone age (mean difference: −0.360 mos.; 95% CI: −0.770 to −0.954; $p = 0.299$). **Conclusions**: In conclusion, the Tanner–Whitehouse 3 method demonstrates overall precision and reliability but requires caution, as it underestimates bone age in preschool children and school-age males, while slightly overestimating it in adolescents.

Keywords: diagnostic imaging; skeletal maturation; Tanner–Whitehouse 3 method; bone age assessment; pediatric radiology; Canary Islands

check for updates

Academic Editor: Peter Giannoudis

Received: 9 December 2024
Revised: 28 December 2024
Accepted: 31 January 2025
Published: 6 February 2025

Citation: Martín Pérez, S.E.; Martín Pérez, I.M.; Molina Suárez, R.; Vega González, J.M.; García Hernández, A.M. The Validation of the Tanner–Whitehouse 3 Method for Radiological Bone Assessments in a Pediatric Population from the Canary Islands. *Osteology* **2025**, *5*, 6. https://doi.org/10.3390/osteology5010006

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Maturation is the process through which individuals experience physical and psychological development, transitioning from childhood to adulthood [1]. These changes occur across various systems, affecting the structure and function of organs, tissues, and bodily systems. Physical changes during maturation encompass a series of biological processes through which the body undergoes significant transformations as it grows and develops. One of the most outstanding physical changes is skeletal maturation, where the bones grow in size and strength.

Methods of evaluating skeletal maturation include several approaches, with X-ray techniques being the most common and reliable. These procedures involve the use of radiographs to visually assess the stages of bone development. One of the most widely used X-ray methods for evaluation of skeletal maturity is the Greulich and Pyle (GP) Atlas, introduced in 1959. This method uses a set of standardized wrist radiographs from children of high socioeconomic status, mainly from Euro-American backgrounds. These reference images show different stages of skeletal maturation and are used to compare an individual's X-ray to visually assess their bone age (BA) [2]. While simple and effective, it has limitations due to its focus on a specific population [3,4].

The Tanner-Whitehouse (TW) method, developed in the 1950s and 1960s, is a precise and widely used tool for assessing BA based on hand and wrist radiographs. It assigns scores according to the ossification of 20 bones, with a maximum of 1000 points upon reaching full maturity. It has evolved through three versions: TW1 (1962), based on radiographs of white European children [5]; TW2 (1975), which was updated to reflect contemporary populations [6]; and TW3 (2001), which introduces a two-phase process. First, 13 regions of interest (ROIs) are evaluated instead of the 20 bones in the original version, and then the total score is compared with reference charts to estimate BA in relation to chronological age (CA) [7].

A recent study by Martín Pérez et al. (2023) [4] evaluated the precision of BA assessment methods in children across different ethnic groups. According to their results, both the GP Atlas and TW3 method were reliable tools; however, their accuracy differs depending on the ethnicity of the child being assessed. In this regard, the GP Atlas was accurate for Caucasian and Hispanic children, but overestimated BA in Asian and Arab adolescents, while TW3 showed higher accuracy in African children. All these findings suggest that ethnic bias may influence BA estimation, highlighting the need for methods tailored to different populations.

Regarding the study population, the peoples of the Canary Islands are notably diverse, resulting from the intermixing among Indigenous peoples [8,9], Europeans, and immigrants from various regions, creating a unique ethnic mosaic [10]. With high density and urban concentration in Gran Canaria and Tenerife, the population is clustered in coastal and metropolitan areas [11], while the higher regions experience depopulation [12]. Since the 1980s, there has been demographic aging and a decline in birth rates. The cities of Las Palmas de Gran Canaria and Santa Cruz de Tenerife stand out as population and economic centers, driven by tourism and centralized services [13].

Previous studies suggest that Canarian children have higher BA than British children but slightly lower than North American. Toledo (1978) came to this conclusion in his doctoral thesis after analyzing bone growth and development in a sample of 2829 children (1421 boys and 1408 girls) from Santa Cruz de Tenerife. When comparing the Canarian and the American samples, the latter showed a slightly higher BA at the same CA. Using the TW2 method, it was observed that the BA of Canarian children was higher than those of the British population in all age groups, except during the first three years of life, where British BA exceeded the Canarian children by an average of 2.4 mos. [14]. In another place,

this author and colleagues stated that in the case of BA, both males and females showed significantly different values compared to the English and North American samples, with our data, in any case, being closer to the latter [15].

The society's composition changes that have occurred in the Canary Islands over the past 50 years could have led to variations in infant maturation patterns compared to historical data, highlighting the need to update the reference values for BA in this specific population. While the GP-Canary Atlas was recently validated for the pediatric population of the archipelago by Martín Pérez et al. (2024) [16,17], showing its accuracy in assessing bone maturation within this specific group, the latest version of the TW3 method has not yet undergone validation in this context. Despite its widespread use and adaptability across various pediatric populations, its unvalidated status for the Canary Islands raises concerns about its precision and reliability in estimating BA for this population. To address this gap, this study aims to evaluate the precision of the TW3 method by analyzing its intra- and inter-rater reliability and assessing its accuracy across developmental stages and genders in Canarian children.

2. Materials and Methods

2.1. Study Design

A cross-sectional study was conducted between 1 September 2023 and 20 June 2024 within the Departments of Pediatrics at the Complejo Hospitalario Universitario de Canarias (Tenerife, Canary Islands, Spain), a tertiary-level referral healthcare institution located in Tenerife, Spain. This study adhered to the STARD 2015 guidelines [18], which provide a comprehensive checklist for reporting diagnostic accuracy studies, thereby ensuring adherence to high methodological and reporting standards. Ethical approval was granted by the Ethics Committee of the Complejo Hospitalario Universitario de Canarias (reference number CHUC_2023_86), with approval obtained on 13 July 2023. The study protocol conformed strictly to the ethical principles set forth in the Declaration of Helsinki.

To provide robust contextual data for analysis, sociodemographic variables, including age and gender, as well as anthropometric measurements such as height, weight, and body mass index (BMI), were meticulously extracted from the SAP Logon database (IBM®, Armonk, NY, USA) at the Complejo Hospitalario Universitario de Canarias. Standardized PA-HW radiographs, securely stored in the Centricity PACS system (GE HealthCare®, Chicago, IL, USA), were systematically analyzed for all participants. The analysis was conducted according to a rigorously predefined protocol aimed at ensuring consistency and precision in data collection and interpretation.

As part of this verification process, each PA-HW radiograph was carefully reviewed to confirm proper positioning of the patient's left hand, with the fingers slightly spread and the wrist appropriately aligned with the forearm. Radiographs were meticulously examined to ensure that all relevant anatomical landmarks, including the phalanges, metacarpals, carpal bones, and distal radius and ulna, were clearly visible and accurately captured. Furthermore, the imaging parameters—such as exposure, focus, and contrast—were thoroughly checked to confirm adherence to established protocol standards, thereby ensuring the reliability and reproducibility of imaging data.

2.2. Participants

2.2.1. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for this study were meticulously established to guarantee a representative and homogeneous cohort of healthy children and adolescents from the Canary Islands, facilitating an accurate assessment of BA through posteroanterior hand and wrist (PA-HW) radiographs. Inclusion criteria for the study required that

participants: (1) be healthy children aged 0 to 18 years; (2) be born in the Canary Islands and have lived there continuously for at least five years; and (3) have at least one parent of verified Canary Island descent, as determined through comprehensive medical and family history records, to ensure consistency in the genetic profile of the study group. Additionally, (4) participants needed to have medical records dating from 2016 onwards, and (5) their PA-HW radiographs were required to meet strict quality standards, including proper hand positioning, clear visualization of key anatomical landmarks, and adherence to standardized imaging protocols.

The exclusion criteria were designed to maintain the accuracy and reliability of bone age assessments by controlling for potential confounding factors that might disrupt normal skeletal development. (1) Children diagnosed with medical conditions, including endocrine–metabolic disorders (e.g., growth hormone deficiency, hypothyroidism), neurological disorders (e.g., cerebral palsy), or genetic syndromes (e.g., Down syndrome, Turner syndrome), were excluded, as these conditions can profoundly affect growth and bone maturation processes. Furthermore, (2) participants undergoing treatments such as growth hormone therapy, corticosteroids, or chemotherapy were not included, given their known effects on skeletal growth. Additionally, (3) radiographs were excluded if they revealed fractures, significant skeletal deformities, or (4) did not meet quality standards (e.g., poor resolution, obscured anatomical landmarks), as such issues could interfere with the reliability of the evaluation.

2.2.2. Sample Size Calculation

The sample size for this study was calculated to ensure precise and accurate BA measurements using the TW3, with a 95% confidence level and a 5% margin of error. Due to the lack of standard deviation data for BA in the Canary Islands' population, estimates from similar studies in comparable populations were used to approximate expected variability [16]. Based on these estimates, a minimum of a total sample size of n = 200 children was determined to be sufficient for validating the TW3 method in this pediatric population, ensuring reliable and generalizable findings.

2.3. Test Methods

To ensure objective and unbiased BA determination, three blinded raters independently assessed the PA-HW radiographs. Specifically, the raters included a Radiology Specialist (Rater 1), a General Practitioner (Rater 2), and a Medical Student (Rater 3), thereby representing varying levels of expertise and training in radiological evaluation. This diverse composition was intentionally chosen to analyze the influence of professional experience on the precision and accuracy of BA assessments. Consequently, each rater determined BA by comparing observed skeletal features with the reference images provided by the TW3 method.

To further evaluate intra-rater precision, each rater assessed BA at two separate time points (T1 and T2), which were spaced less than six weeks apart. Importantly, the PA-HW radiographs were presented in a randomized and blinded order during both assessments to minimize interpretation bias and prevent recall of prior evaluations. Thus, this strategy enabled a thorough examination of intra-rater reliability by comparing each rater's measurements from T1 to T2, thereby assessing the consistency of their evaluations over time.

Moreover, inter-rater precision was examined by comparing BA determinations across the three raters to analyze their levels of agreement and reliability in interpreting the same set of radiographs. This comparative analysis was essential for assessing the reproducibility of the BA assessment method across raters with differing expertise levels. Additionally, the

accuracy of the BA assessments was evaluated by comparing the estimated BA using TW3 to the subjects' CA, calculated as the difference between their birth date and the date of the radiological exam.

With respect to the BA, it was estimated by the TW3 that it is a radiological method widely recognized for assessing BA through PA-HW radiographs [5–7]. It challenges the validity of correlating age with height, dentition, and bone development, arguing that bone maturation does not follow a uniform trajectory but varies by individual developmental stages. Unlike qualitative approaches like GP Atlas, numerical methods provide a more objective quantification of bone maturation, assigning specific scores to each bone based on gender and developmental stage.

This method employs a structured scoring system for individual bones, assigning maturity scores according to predefined criteria such as shape, size, and ossification degree. This evaluation focuses on the radius, ulna, metacarpals, phalanges, and select carpal bones. The scores are tailored to the individual's sex and developmental stage, then summed to generate a total maturity score. This score is compared with established reference values corresponding to CA ranges, yielding an estimated BA. This standardized and precise approach enhances comparability across different populations and allows for accurate assessments of skeletal development.

2.4. Analysis

Statistical analyses were conducted using IBM® SPSS Statistics 29.0.1.0 software (Armonk, NY, USA). Descriptive statistics were first calculated for age (in mos.), weight (kg), height (m), and body mass index (BMI) (kg/m^2). The data were stratified according to developmental stages as defined by Fraga and Fernández (2014) [19]—preschool children (1 to 5 years), school-age children (5 to 12 years), and teenagers (12 to 18 years)—and further segmented by gender to account for potential differences in bone maturation between males and females. These descriptive measures included calculations of central tendency (mean) and dispersion (standard deviation, minimum, and maximum) for both CA and BA as estimated by the TW3 method. To confirm the suitability of the data for further statistical analyses, the Shapiro–Wilk test was applied to assess the normality of the data distribution, while Levene's test was used to evaluate homoscedasticity.

In order to assess precision, the intra-class correlation coefficient (ICC) was calculated to evaluate both intra-rater and inter-rater agreement. The ICC was chosen as it provides a robust and quantitative measure of reliability, indicating the degree of consistency within and between raters when using the TW3 method for BA assessment. Bland–Altman plots were constructed to visually examine inter-rater reliability, offering a detailed evaluation of potential systematic bias and the limits of agreement between raters' measurements. This method was selected for its ability to graphically highlight discrepancies and assess agreement across the range of values. Additionally, a paired samples test was employed to evaluate the accuracy of BA estimations by comparing discrepancies between BA estimated through the TW3 method and the children's CA.

3. Results

3.1. Characteristics of the Sample

A total of 214 posteroanterior hand and wrist (PA-HW) radiographs from healthy children were analyzed, including 80 females and 134 males. Among females in the preschool group, the mean age was 39.33 mos. (SD = 15.18), with an average weight of 14.52 kg (SD = 2.05) and an average height of 0.91 m (SD = 0.07). Conversely, males had a mean age of 46.49 mos. (SD = 13.33), a mean weight of 13.09 kg (SD = 2.17), and a mean height of 0.94 m (SD = 0.05). In the school-age group, the female group showed a mean age

of 92.00 mos. (SD = 26.08), a mean weight of 29.58 kg (SD = 7.14), and a mean height of 1.14 m (SD = 0.07).

Conversely, the male group presented with a mean age of 100.16 mos. (SD = 20.33), a mean weight of 23.67 kg (SD = 4.85), and a mean height of 1.16 m (SD = 0.05). Among teenager, both females and males recorded a mean height of 1.33 m. Females had an average age of 144.17 mos. (SD = 23.81) and a mean weight of 33.84 kg (SD = 4.62), while males demonstrated a mean age of 151.53 mos. (SD = 20.17) and a mean weight of 34.21 kg (SD = 3.19). The Shapiro–Wilk test confirmed the normal distribution of all variables within these groups. Further details are presented in Table 1.

Table 1. Characteristics of sample. Abbreviation: BMI = Body Mass Index, mos. = months. A *p*-value lower than these thresholds indicates a statistically significant deviation from normality.

	Stage	Gender	N	Mean	SD	Min	Max	*p*-Value
Age (mos.)	Preschool	Female	24	39.33	15.18	20.00	67.00	0.235
		Male	45	46.49	13.33	18.00	69.00	0.105
	Scholar	Female	40	92.00	26.08	85.00	118.00	0.310
		Male	62	100.16	20.33	75.00	109.00	0.089
	Teenager	Female	16	144.17	23.81	102.00	168.00	0.150
		Male	27	151.53	20.17	107.00	192.00	0.080
Weight (kg)	Preschool	Female	24	14.52	2.05	9.80	18.60	0.215
		Male	45	13.09	2.17	7.40	18.00	0.175
	Scholar	Female	40	29.58	7.14	17.60	40.00	0.200
		Male	62	23.67	4.85	14.20	44.00	0.115
	Teenager	Female	16	33.84	4.62	22.00	39.50	0.250
		Male	27	34.21	3.19	23.80	45.70	0.140
Height (m)	Preschool	Female	24	0.91	0.07	0.77	1.05	0.289
		Male	45	0.94	0.05	0.80	1.10	0.175
	Scholar	Female	40	1.14	0.07	0.99	1.30	0.200
		Male	62	1.16	0.05	0.94	1.40	0.115
	Teenager	Female	16	1.33	0.04	1.21	1.37	0.250
		Male	27	1.33	0.03	1.16	1.45	0.140
BMI (kg/m^2)	Preschool	Female	24	17.53	2.47	8.32	19.49	0.180
		Male	45	14.81	2.45	18.81	18.87	0.120
	Scholar	Female	40	22.76	5.49	13.45	20.29	0.175
		Male	62	17.59	3.60	12.57	20.92	0.150
	Teenager	Female	16	19.13	2.66	15.02	21.73	0.240
		Male	27	19.33	1.80	14.61	20.99	0.130

3.2. Main Results

3.2.1. Precision

- intra-rater Agreement

The ICC results showed strong precision and consistency in intra-rater reliability across all three raters when assessing BA using the TW3 technique, with slight gender-based variations. Specifically, Radiologist Specialist (Rater 1) exhibited high consistency, achieving an ICC of 0.996 (95% CI: 0.991–0.998) for females and 0.994 (95% CI: 0.989–0.997) for males. Similarly, General Practitioner (Rater 2) displayed strong reliability, with an ICC of 0.988 (95% CI: 0.976–0.994) for females and 0.993 (95% CI: 0.980–0.997) for males. Medical Student (Rater 3), while showing slightly lower but still strong ICC values, recorded 0.928 (95% CI: 0.845–0.968) for females and 0.970 (95% CI: 0.935–0.986) for males. Additional details are provided in Table 2.

Table 2. Intra-rater agreement across time points and gender. The table presents mean BA values, intra-class correlation coefficients (ICC), and 95% confidence intervals (CI) for the lower and upper bounds for each rater (Rater 1, Rater 2, and Rater 3) during two measurement periods (T1 and T2) for female and male participants.

Group	Time of Measurement	Gender	Mean	ICC	95% CI Lower	95% CI Upper
Rater 1	T1	Female	76.82			
		Male	77.94			
	T2	Female	74.89	0.996	0.991	0.998
		Male	75.67	0.994	0.989	0.997
Rater 2	T1	Female	73.50			
		Male	81.32			
	T2	Female	71.12	0.988	0.976	0.994
		Male	79.58	0.993	0.980	0.997
Rater 3	T1	Female	77.45			
		Male	79.02			
	T2	Female	79.34	0.845	0.968	0.964
		Male	80.21	0.935	0.986	0.989

- Inter-rater agreement

The inter-rater agreement in determining the BA using the TW3 method showed distinct differences between female and male participants. For females, there was excellent agreement between the Radiologist Specialist (Rater 1) and the General Practitioner (Rater 2), with an ICC of 0.976 (95% CI: 0.950, 0.987). However, the agreement was lower between the Radiologist Specialist and the Medical Student (Rater 3), as well as between the General Practitioner and the Medical Student, with ICCs of 0.812 (95% CI: 0.702, 0.896) and 0.880 (95% CI: 0.820, 0.922), respectively.

For males, the Radiologist Specialist and the General Practitioner demonstrated strong consistency, with an ICC of 0.968 (95% CI: 0.940, 0.982). In contrast, the agreement between the Radiologist Specialist (Rater 1) and the Medical Student (Rater 3), and between the General Practitioner (Rater 2) and the Medical Student (Rater 3), was lower, with ICCs of 0.857 (95% CI: 0.770, 0.921) and 0.912 (95% CI: 0.860, 0.946), respectively. These findings emphasize the high level of agreement between trained evaluators, while the reduced consistency with the Medical Student highlights the need for standardized training to ensure consistent use of the TW3 method. Further details are provided in Table 3.

The Bland–Altman plots in Figure 1 display the agreement among the three raters (Rater 1, Rater 2, and Rater 3) for the BA assessment using the TW3. For female participants, Rater 1 and Rater 2 exhibited high agreement, reflected by a narrow range of differences, indicating strong consistency. In contrast, the agreements between Rater 1 and Rater 3, as well as Rater 2 and Rater 3, were moderate, with wider limits of agreement, suggesting greater variability due to differences in training. A similar pattern emerged for male participants: Rater 1 and Rater 2 demonstrated strong agreement, while Rater 1 and Rater 3, and particularly Rater 2 and Rater 3, showed lower levels of agreement with broader ranges of differences, emphasizing the challenges of achieving consistent assessments among less experienced raters.

Table 3. Inter-rater agreement of bone age assessment using the TW3 method by gender, showing mean bone age values, Intraclass Correlation Coefficient (ICC), and 95% Confidence Intervals (CI) for agreement between rater pairs (Radiologist Specialist vs. General Practitioner, Radiologist Specialist vs. Medical Student, and General Practitioner vs. Medical Student) for female and male participants.

Groups	Gender	Mean	ICC	95% CI Lower	95% CI Upper
Rater 1–Rater 2	Female	76.82 73.50	0.976	0.950	0.987
	Male	77.94 81.32	0.968	0.940	0.982
Rater 1–Rater 3	Female	76.82 77.45	0.812	0.702	0.896
	Male	77.94 79.02	0.857	0.770	0.921
Rater 2–Rater 3	Female	73.50 77.45	0.880	0.820	0.922
	Male	81.732 79.02	0.912	0.860	0.946

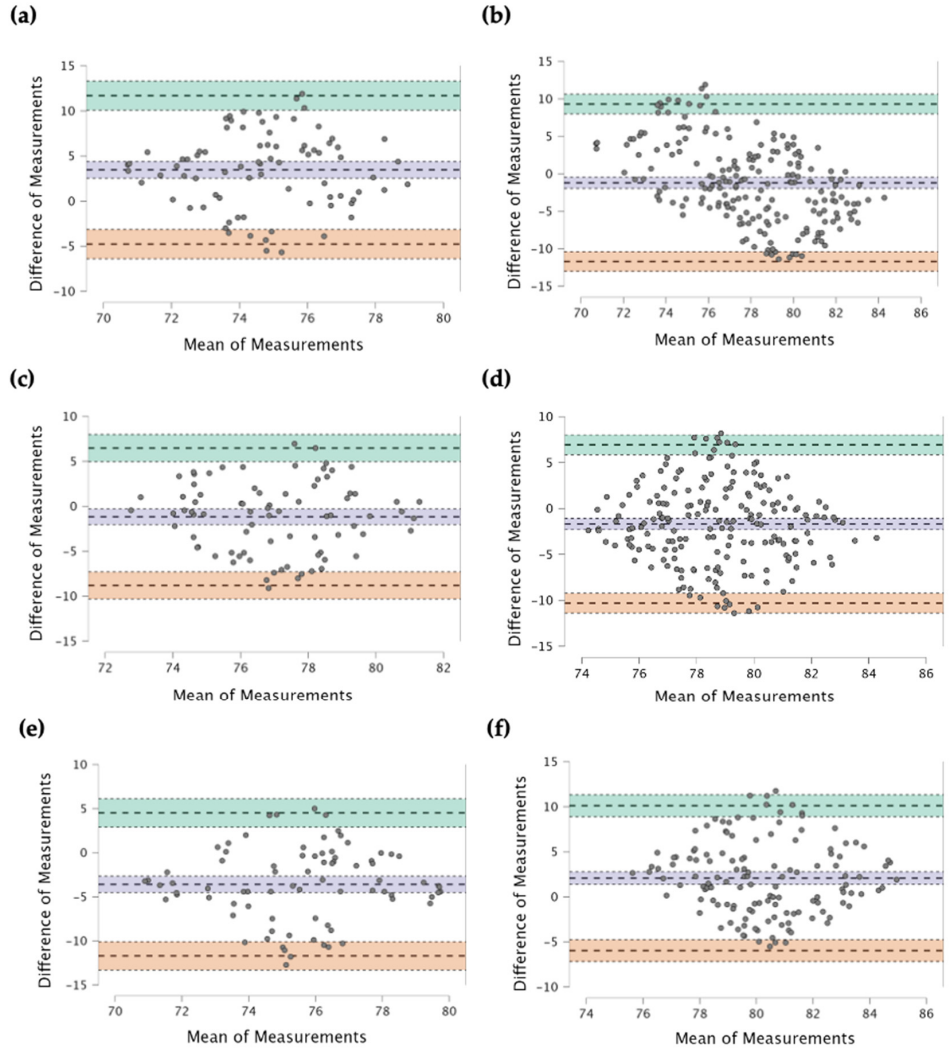


Figure 1. Bland–Altman plots illustrating BA assessments using the Tanner–Whitehouse 3 (TW3). The plots compare the assessments of Rater 1 with Rater 2 for both females (**a**) and males (**b**), Rater 1 with Rater 3 for females (**c**) and males (**d**), and Rater 2 with Rater 3 for females (**e**) and males (**f**). The dashed lines represent the mean differences, while the shaded areas in orange and green show the limits of agreement (±1.96 standard deviations). The purple lines represent the confidence intervals for the limits of agreement.

3.2.2. Accuracy

The TW3 method demonstrated limited accuracy in estimating BA compared to CA in the preschool cohort. Specifically, for children aged 1 to 5 years, the method yielded a slight underestimation of BA, with a mean difference MD = 3.712 mos. (95% CI: 1.29 to 6.13 mos., $p = 0.199$). This underestimation was less pronounced in males (MD = 0.106 mos., 95% CI: −4.83 to 5.05 mos., $p = 0.935$) compared to females, who exhibited a larger difference (MD = 4.200 mos., 95% CI: 0.78 to 7.62 mos., $p = 0.164$).

In the school-age group (ages >5 to 12 years), the TW3 method similarly demonstrated a minor underestimation of BA, with an MD of 3.522 mos. (95% CI: 1.05 to 5.99 mos., $p = 0.888$). The magnitude of this underestimation was less significant in males (MD = 3.978 mos., 95% CI: −12.55 to 4.59 mos., $p = 0.926$) compared to females (MD = 4.110 mos., 95% CI: −8.25 to 16.47 mos., $p = 0.945$).

In the teenager group, the TW3 method showed a slight overestimation of BA relative to CA, with an MD = −0.360 mos. (95% CI: −0.77 to −0.954 mos., $p = 0.299$). When stratified by sex, females exhibited a mean difference of −0.380 mos. (95% CI: −4.84 to 5.50 mos., $p = 0.256$), indicating no statistically significant difference. Similarly, in males, the mean difference was −0.330 mos. (95% CI: −1.03 to 1.78 mos., $p = 0.222$), also reflecting no significant discrepancy between BA and CA. Further details are provided in Table 4 and visually summarized in Figure 2.

Table 4. Accuracy of BA assessments using TW3. Abbreviation: BA = bone age, CA = chronological age, CI = confidence interval, MD = mean difference CA–BA, SD = standard deviation; W = Paired Samples Test or Wilcoxon signed-rank statistic.

Stage		Mean	SD	MD	Lower CI 95%	Superior CI 95%	*p*
Preschool (n = 69)	CA	43.485	14.476				
	BA	39.772	15.409	3.712	1.290	6.130	0.199
Female	CA	39.331	15.182				
	BA	35.131	16.896	4.200	0.780	7.620	0.164
Male	CA	46.496	13.333				
	BA	46.390	24.881	0.106	−4.830	5.05	0.935
Scholar (n = 102)	CA	95.684	23.906				
	BA	92.161	35.572	3.522	1.050	5.990	0.888
Female	CA	92.001	26.086				
	BA	87.891	37.203	4.110	−8.250	16.47	0.945
Male	CA	100.168	20.338				
	BA	96.190	33.876	3.978	−12.55	4.59	0.926
Teenager (n = 43)	CA	148.883	23.665				
	BA	149.242	29.943	−0.360	−0.770	−0.954	0.299
Female	CA	144.170	23.810				
	BA	144.549	24.231	−0.380	−4.840	5.500	0.256
Male	CA	151.53	20.176				
	BA	151.86	18.179	−0.330	−1.03	1.78	0.222

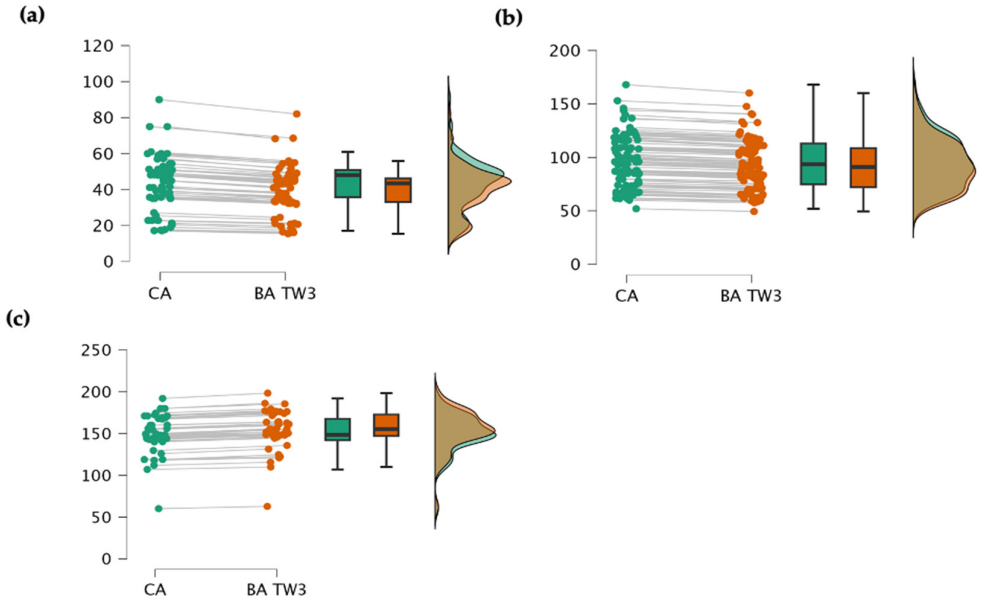


Figure 2. Accuracy of BA determination using TW3 across different developmental stages. Raincloud plots display BA accuracy in (**a**) preschool (1 to 5 years), (**b**) school-age (>5 to 12 years), and (**c**) teenager (>12 to 18 years) groups. Green points represent chronological age (CA), orange points represent bone age (BA) estimated using the TW3 method, and the brown density plots depict the distribution of CA and BA. Method shows significant BA underestimation and variability in preschool and school-age groups, while accuracy improves in the teenager group with no significant overestimation between CA and BA.

4. Discussion

4.1. Precision of TW3

The assessment of an instrument's precision ensures that it meets essential criteria for accuracy and consistency, which are necessary for the generalization of research findings. Reliability refers to the degree of precision with which an instrument measures what it is intended to measure.

4.1.1. intra-rater Agreement

Our findings on the precision of the TW3 method align with the existing literature for diverse populations, including Caucasian cohorts. Our study showed a high intra-rater precision in TW3-based BA assessments across various raters. The radiology specialist (Rater 1) achieved near-perfect intra-class correlation coefficients ICC = 0.995 for females and 0.996 for males. Similarly, the general practitioner (Rater 2) demonstrated high precision with ICC = 0.990 for females and ICC = 0.992 for males. While medical students (Rater 3) also exhibited strong precision; their ICCs were slightly lower at ICC = 0.921 for females and ICC = 0.976 for males.

These findings are consistent with other studies that have previously studied TW3 precision. In Europe, Pinchi et al. (2014) [20] reported high intra-examiner reliability among Caucasian Italian children, with correlation coefficients of $r = 0.843$ for boys and $r = 0.910$ for girls. Although these values are somewhat lower than ours, such variations may reflect differences in rater training, population characteristics, or even the sample sizes.

Furthermore, in South America, the evidence for TW3 precision is provided by López et al. (2008) [21], who documented high accuracy for the TW3 RUS method in Venezuelan children aged 7 to 14 years, with a correlation of $r = 0.91$ for the entire sample and $r = 0.93$ for girls. In contrast, the TW3 Carpal method showed reduced accuracy, with $r = 0.89$ for boys and $r = 0.82$ for girls.

Comparable precision results have been observed by Kowo-Nyakoko et al. (2023) [22] in Zimbabwean children, where the TW3-RUS method displayed a very strong intra-examiner reliability, with correlation coefficients of $r = 0.95$ for boys and $r = 0.93$ for girls. On the other hand, Kim et al. (2015) [23] similarly demonstrated strong intra-examiner agreement for TW3, with a Pearson's correlation coefficient of $r = 0.93$ for Korean children aged 7 and 12 years.

The results obtained for children from the Canary Islands are consistent with findings from other studies, particularly among experienced raters, reinforcing the robustness of the TW3 method across various ethnic and geographic contexts [24]. However, these findings also point to potential variability influenced by the rater's level of training and cognitive biases, underscoring the critical need for standardized training and protocols to enhance reliability and consistency across diverse populations.

4.1.2. Inter-Rater Agreement

Our findings on the inter-rater agreement of the TW3 method for BA assessment align with previous studies in various populations, while also highlighting unique differences. Specifically, we observed excellent agreement between the Radiologist Specialist (Rater 1) and the General Practitioner (Rater 2), with ICC = 0.976 for females and ICC = 0.968 for males. Concordance was lower when the Medical Student (Rater 3) was involved, indicating that experience and training significantly affect precision.

Comparatively, Benjavongkulchai et al. (2018) [25] reported kappa values ranging from $k = 0.66$ to $k = 0.88$ for the TW3-RUS method in Thai children, reflecting moderate to high agreement but with greater variability compared to our findings among professional raters. Similarly, Geng et al. (2024) [26] found Kendall's W values of W = 0.833 for the radius, ulna, and short finger (TW3-RUS) system, and W = 0.944 for the carpal system in a sample of children living in Beijing, China.

Moreover, Alshmarani et al. (2019) [27] reported strong inter-examiner reliability for the TW3 method in Saudi children, with intraclass correlation coefficients (ICCs) of 0.972 for females and 0.963 for males. This finding aligns with the high agreement observed among our trained raters and highlights consistent gender-related differences in reliability. Similarly, Yuh et al. (2023) [28] documented comparable inter-rater agreement for the TW3-RUS method, ranging from 0.985 to 0.984, and for the TW3-Carpal method, ranging from 0.981 to 0.973, in a sample of Taiwanese children.

Our findings underscore the fundamental role of rater training and experience in achieving consistent assessments using the TW3 method [29]. The variability observed among less experienced raters, such as medical students, can be attributed to the complexity introduced by this method in radiological BA assessments.

While the hand offers a large number of areas of interest for studying radiological maturity, inter-observer variation or systematic error tends to be more pronounced in non-expert raters with the TW3 method. Unlike the GP Atlas, which relies on a single general judgment due to its primarily visual approach, the TW3 method requires at least 20 distinct evaluations for each radiological region of interest, thereby increasing the potential for variability in results.

4.2. Accuracy of TW3

The evaluation of an instrument's validity ensures precision and accuracy [30], with reliability targeting consistency and validity assessing accuracy. This study seeks to develop a tool for skeletal maturity assessment, emphasizing the need for updated, population-specific normative data due to the influence of hereditary, ethnic, behavioral, and nutritional factors. While Toledo's 1978 study established early TW2 references for the Canary Is-

lands [14], outdated data and methods limited its relevance [7,31]. Updated standards are crucial for applicability, as high population variability challenges the generalization of methods, necessitating tailored normative data.

Our study is the first to assess the suitability of the TW3 method in the contemporary pediatric population of the Canary Islands. Differentiating between normal and pathological variations is essential, and our study expands the threshold for skeletal maturity from 2 to 2.5 standard deviations to identify normal variants more flexibly.

Previous studies have demonstrated that Canarian children show a higher BA than British children but a slightly lower BA compared to North American children. The abovementioned study of Toledo (1978) [14] carried out a comparative analysis of Canarian and United States and Canada populations, finding that North American children showed a somewhat higher BA. Using the TW2 method, he observed that children from the Canary Islands had a higher BA than their British counterparts, except during the first three years of life, when British children exceeded the Canarian cohort by an average of 2.4 mos.

Our findings align with those of Toledo (1978) [14], particularly concerning the underestimation of BA in Canarian preschool children (1–5 yrs.). Moreover, consistent with Toledo's observations, we confirmed a slight overestimation of BA using the TW method in teenagers (12–18 yrs.). This overestimation, however, was notably lower than that reported in other regions of Spain, such as Aragón, where Ebrí-Torné (2021) observed that the TW2 method overestimated BA by 4 to 6 mos. in a comparable sample of Aragonese children [32]. However, in the school-age population (5–12 yrs.), we have identified an underestimation of BA using the TW3 method, in contrast with the overestimation reported by Toledo (1978) in the late Seventies [14]. Therefore, despite the small differences that exist, we can affirm that the accuracy of the TW method identified in our study for Canarian children is consistent with historical findings for the same population. Thus, variants detected within the school-age population should be interpreted as physiological and within the normal range, i.e., reflecting constitutional developmental differences without clinical significance.

Regarding the differences in bone maturation seen in the present study, while not significant, it may be attributed to factors such as natural individual variability and sexual dimorphism [33]. Studies utilizing databases such as those from the Colorado Longitudinal Twin Study (Colorado, 1985) and the Fels Longitudinal Study (Ohio, 1929) suggest a hereditary predisposition, as specific ossification sequences tend to recur within families, pointing to a genetic pattern [34]. This process is primarily influenced by genetic factors through multifactorial polygenic inheritance, which, under ideal conditions, explains between 80% and 90% of bone maturation [35]. It is estimated that 25% of the variability in hand bone ossification is explained by genetic factors, and a similar phenomenon is observed in the order of ossification [34].

Furthermore, it is important to consider that during the preschool stage, growth deceleration is observed, influenced by both environmental and genetic factors [34,36]. Linkage studies suggest that, although genetic factors affect skeletal maturation during childhood, some have a limited impact over time [37,38]. In early childhood, bone ossification predominates, and structural genes are believed to be responsible for the initial formation of bones [39–41]. The growth deceleration after 6 months may be partly due to environmental factors, although genetic factors may also influence this process [34,42].

In the preschool stage (1–5 yrs.), the TW3 method tends to slightly underestimate BA, with an MD = 3.712 mos. This underestimation is more pronounced in girls compared to boys, potentially reflecting early sexual dimorphism. Although this dimorphism is subtle in the early years, it becomes increasingly evident with development, likely due to differences in the timing and sequence of ossification processes between sexes [43,44]. Sex-based discrepancies in BA estimation may arise from variations in the interplay of

genetic and hormonal factors. Estrogen and testosterone, even at this early stage of life, play pivotal roles in skeletal development. Estrogen, for example, promotes epiphyseal plate closure and modulates endochondral ossification, while testosterone stimulates periosteal apposition, contributing to differences in bone structure and density between sexes [45]. These hormonal influences could partly explain the greater deviation observed in girls during this stage.

Thomis et al. (2006) [36] underscore the importance of systemic hormones, such as growth hormone (GH) and thyroid hormones, in regulating bone maturation and linear growth. Thyroid hormones, particularly triiodothyronine (T3), have a dual role in promoting bone ossification while also influencing cartilage proliferation. They directly stimulate osteoblast activity and interact with GH through the hypothalamic-pituitary-thyroid (HPT) axis to regulate somatomedin production, which is critical for bone growth and development [46–48]. Notably, thyroid hormones have a stronger effect on ossification compared to linear growth, which could amplify the discrepancies observed in BA estimation [46]. Furthermore, the HPT axis tightly regulates these hormonal interactions. Thyrotropin-releasing hormone (TRH) from the hypothalamus induces the secretion of thyroid-stimulating hormone (TSH) from the pituitary, which in turn stimulates thyroid hormone synthesis. The existence of dysregulations in this axis, even minor ones, could disproportionately affect bone maturation processes, further influencing BA estimation [49,50].

In the school stage (5–12 yrs.), small sexual differences in BA were observed, with an underestimation of BA identified, showing a mean discrepancy of 3.522 mos., which was more pronounced in girls. Contrary to previous studies, our findings suggest an acceleration in bone maturation in girls, leading to a convergence in skeletal development between the sexes, thereby overcoming the delay observed in earlier stages. Recent research has established that global differences in bone maturation rates between genders are narrowing, resulting in more synchronized developmental trajectories [51]. However, in contrast to this general trend, and in line with other studies [52,53], we report a consistent developmental gap during the preschool stage (5–12 yrs.), with girls maturing faster than boys. It is well established that changes in bone maturation occur in both sexes, characterized by increases in length (proceritas prima and secunda) and weight (turgor secundus). These findings may be further explained by the pubertal growth spurt, which, at least in the Canarian pediatric population, occurs earlier in girls (ages 11–13), leading to greater height and strength compared to boys [53].

In our study, we observed that girls had an average BMI of 22.76 kg/m^2, which was significantly higher than that of boys (17.59 kg/m^2). This difference may contribute to earlier puberty, which is often associated with obesity [54,55]. Leptin, a hormone associated with body fat profile, is believed to play a role in stimulating puberty in girls. However, its relationship with the onset of puberty in boys remains less clear, suggesting that it may not be the primary trigger for them [56]. Although the specific mechanisms behind this acceleration are not fully understood, it is suggested that factors such as exposure to endocrine disruptors, chronic stress, and calorie-rich diets may play a significant role in influencing this phenomenon [57,58]. These complex factors require further research to understand their impact on pubertal development and bone maturation in the child population of the Canary Islands.

Lastly, in the adolescent stage (12–18 years), the TW3 method demonstrated a slight overestimation of bone age (BA), with a mean difference (MD) of −0.360 mos. Girls presented an MD of −0.380 mos., while boys exhibited an MD of −0.330 mos.; however, these differences were not statistically significant. During this stage, the role of sex hormones becomes more prominent, driving differences in skeletal maturation between sexes. In boys, the surge of androgens, particularly testosterone, accelerates bone maturation by

promoting periosteal expansion and increasing bone mineral density. This process is critical for the completion of skeletal growth and the attainment of peak bone mass [49].

In girls, estrogens play a central role in regulating bone metabolism and maturation. Their influence is greater on bone maturation than on linear growth, as estrogen promotes the closure of the growth plates (epiphyseal fusion) while modulating bone turnover rates. This hormonal influence contributes to the earlier completion of skeletal growth in females compared to males [59]. Furthermore, the interaction between estrogens and other growth-related factors, such as insulin-like growth factor-1 (IGF-1), enhances osteoblast activity and bone remodeling, ensuring the structural and functional optimization of the skeleton during adolescence [60].

4.3. Limitations

A significant limitation of this study lies in its inability to fully capture bone maturation across all stages of childhood and adolescence, particularly during the early preschool years and late adolescence. This limitation reduces its clinical applicability for precise estimations during these critical developmental periods. The reliance on TW3 reference standards, which may not adequately reflect population diversity, introduces potential biases, particularly in underrepresented groups. Additionally, the manual nature of bone age assessment poses inherent challenges, such as observer variability, which can compromise precision. Simplifications in the study design, including its cross-sectional approach and the lack of advanced imaging techniques for validation, further limit the generalizability of the findings. To address these challenges, future research should focus on integrating longitudinal data to track growth trajectories, establishing population-specific reference standards, and utilizing automated processes and advanced imaging technologies to enhance precision and minimize variability. Such advancements are essential to improving the accuracy and broader applicability of bone age estimation methods across diverse clinical and research settings.

4.4. Recommendations for Clinical Practice

The findings suggest that the TW3 method has limitations in estimating BA across developmental stages in the pediatric population of the Canary Islands. Specifically, BA tends to be underestimated in preschool and school-aged children and slightly overestimated in adolescent girls. To address these discrepancies, clinicians are encouraged to complement TW3 assessments with additional diagnostic tools, such as radiographic or anthropometric evaluations, and to adjust for demographic and population-specific factors inherent to the multiethnic nature of the Canary Islands. For preschool-aged children, implementing longitudinal monitoring can help account for individual growth variability, while for adolescent girls, tailored correction factors may mitigate overestimation biases. Furthermore, training healthcare professionals on the strengths and limitations of the TW3 method and incorporating visual aids, such as tables and figures, can enhance the method's clinical applicability.

5. Conclusions

In summary, the TW3 method proves to be a reliable tool for estimating BA in the pediatric population, with particular accuracy in preschool and school-aged children. However, its precision varies across sex, with underestimations being more pronounced in females, and across developmental stages, where heterogeneity within age groups, such as 5–12 years, may influence outcomes. These findings underscore the need for further refinement of the method to improve its applicability across diverse clinical and demo-

graphic settings, ultimately enhancing its value as a diagnostic tool. Future research should prioritize addressing these variations to ensure greater consistency and broader utility.

Author Contributions: Conceptualization, I.M.M.P. and S.E.M.P.; methodology, I.M.M.P.; validation, I.M.M.P., S.E.M.P. and R.M.S.; formal analysis, I.M.M.P.; investigation, I.M.M.P.; resources, R.M.S.; data curation, I.M.M.P.; writing—original draft preparation, I.M.M.P.; writing—review and editing, I.M.M.P.; I.M.M.P., S.E.M.P., R.M.S., J.M.V.G. and A.M.G.H. visualization, I.M.M.P.; supervision, A.M.G.H. and J.M.V.G.; project administration, I.M.M.P. All authors have read and agreed to the published version of the manuscript.

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) (CHUC_2023_86, 13 July 2023).

Informed Consent Statement: Our study is retrospective. Informed consent will not be requested.

Data Availability Statement: Data are contained within the article.

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

References

1. Tanner, J.M. *Growth at Adolescence*; Blackwell Scientific Publications: Oxford, UK, 1962; pp. 1–350.
2. Greulich, W.W.; Pyle, S.I. *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, 2nd ed.; Stanford University Press: California, CA, USA, 1959; pp. 1–272.
3. Ebrí Torné, B. *Maduración Ósea: Metodología Numérica Sobre Tarso y Carpo*; Heraldo de Aragón: Zaragoza, Spain, 1988; pp. 1–320.
4. Martín Pérez, S.E.; Martín Pérez, I.M.; Vega González, J.M.; Molina Suárez, R.; León Hernández, C.; Rodríguez Hernández, F.; Herrera Pérez, M. Precision and Accuracy of Radiological Bone Age Assessment in Children among Different Ethnic Groups: A Systematic Review. *Diagnostics* **2023**, *13*, 3124. [CrossRef] [PubMed]
5. Tanner, J.M.; Whitehouse, R.; Healy, M. *A New System for Estimating Skeletal Maturity from the Hand and Wrist, with Standards Derived from a Study of 2600 Healthy British Children. Part II: The Scoring System*; Centre International de L'Enfance: Paris, France, 1962; pp. 1–100.
6. Tanner, J.M.; Whitehouse, R.H.; Marshall, W.A.; Healey, M.J.R.; Goldstein, H. *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*; Academic Press: New York, NY, USA, 1975; pp. 1–142.
7. Tanner, J.M.; Realy, J.; Goldstein, H. *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method)*; Harcourt Publishers: New York, NY, USA, 2001; pp. 1–200.
8. Fregel, R.; Ordóñez, A.C.; Serrano, J.G. The Demography of the Canary Islands from a Genetic Perspective. *Hum. Mol. Genet.* **2021**, *30*, R64–R71. [CrossRef] [PubMed]
9. Tejera Gaspar, A. *Tenerife y los Guanches*; Centro de la Cultura Popular Canaria: Tenerife, Spain, 2002; p. 17.
10. Afonso, L. *Esquema de Geografía Física de las Islas Canarias*; Ediciones Idea: Tenerife, Spain, 2004; p. 210.
11. Martín Ruiz, J.F. *La Población de Canarias: Análisis Sociodemográfico y Territorial (El Debate Actual)*; Anroart Ediciones: Las Palmas, Spain, 2005; p. 116.
12. Burriel de Orueta, E.L. *Canarias: Población y Agricultura en una Sociedad Dependientes*; Oikos-Tau: Barcelona, Spain, 1982; pp. 46–47.
13. Hernandez Hernández, P. *Natura y Cultura de las Islas Canarias*, 8th ed.; Tafor Publicaciones: La Laguna, Spain, 1997; pp. 56–63.
14. Toledo Trujillo, F.M. Maduración Ósea en una Muestra de Población Urbana de las Islas Canarias. Ph.D. Thesis, Universidad de La Laguna, San Cristóbal de La Laguna, Spain, 1978.
15. Toledo, F.; Cruz, M.; Pastor, S.; Paz, G.; Fernández, J.; Machado, M. Estudios radio-antropométrico de una muestra preadulta de la población canaria. In *Actualizaciones en Medicina del Hospital Universitario de Canarias*; Cabildo de Tenerife: Tenerife, Spain, 1992; pp. 25–35.
16. Martín Pérez, I.M.; Martín Pérez, S.E.; Vega González, J.M.; Molina Suárez, R.; García Hernández, A.M.; Rodríguez Hernández, F.; Herrera Pérez, M. The Validation of the Greulich and Pyle Atlas for Radiological Bone Age Assessments in a Pediatric Population from the Canary Islands. *Healthcare* **2024**, *12*, 1847. [CrossRef] [PubMed]
17. Martín Pérez, S.E.; Martín Pérez, I.M.; Vega González, J.M.; Molina Suárez, R.; León Hernández, C.; Rodríguez Hernández, F.; Herrera Pérez, M. Análisis comparativo de dos métodos radiográficos para determinar la edad ósea en la población pediátrica canaria. In Proceedings of the III Congreso Internacional de Jóvenes por la Investigación, San Cristóbal de La Laguna, Spain, 14–16 November 2024; Universidad de La Laguna: Tenerife, Spain, 2024; pp. 15–20.

18. Cohen, J.F.; Korevaar, D.A.; Altman, D.G.; Bruns, D.E.; Gatsonis, C.A.; Hooft, L.; Irwig, L.; Levine, D.; Reitsma, J.B.; de Vet, H.C.; et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open* **2016**, *6*, e012799. [CrossRef] [PubMed]
19. Fraga Bermúdez, J.M.; Fernández Lorenzo, J.R. La Pediatría, el Niño y el Pediatra: Una Aproximación General. In *Tratado de Pediatría*, 1st ed.; Moro Serrano, M., Málaga Guerrero, S., Madero López, L., Eds.; Editorial Médica Panamericana: Madrid, Spain, 2014; Volume 1, pp. 1–18.
20. Pinchi, V.; De Luca, F.; Ricciardi, F.; Focardi, M.; Piredda, V.; Mazzeo, E.; Norelli, G.-A. Skeletal age estimation for forensic purposes: A comparison of GP, TW2, and TW3 methods on an Italian sample. *Forensic Sci. Int.* **2014**, *238*, 83–90. [CrossRef] [PubMed]
21. 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. *Cienc. Odontol.* **2020**, *5*, 99–111. Available online: https://produccioncientificaluz.org/index.php/cienciao/article/view/33940 (accessed on 12 November 2024).
22. Kowo-Nyakoko, F.; Gregson, C.L.; Madanhire, T.; Stranix-Chibanda, L.; Rukuni, R.; Offiah, A.C.; Micklesfield, L.K.; Cooper, C.; Ferrand, R.A.; Rehman, A.M.; et al. Evaluation of two methods of bone age assessment in peripubertal children in Zimbabwe. *Bone* **2023**, *170*, 116725. [CrossRef] [PubMed]
23. Kim, J.R.; Lee, Y.S.; 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*, 201–205. [CrossRef]
24. Shin, N.Y.; Lee, B.D.; Kang, J.H.; Kim, H.R.; Oh, D.H.; Lee, B.I.; Kim, S.H.; Lee, M.S.; Heo, M.S. Evaluation of the Clinical Efficacy of a TW3-Based Fully Automated Bone Age Assessment System Using Deep Neural Networks. *Imaging Sci. Dent.* **2020**, *50*, 237–243. [CrossRef]
25. Benjavongkulchai, S.; Pittayapat, P. Age estimation methods using hand and wrist radiographs in a group of contemporary Thais. *Forensic Sci. Int.* **2018**, *287*, 218.e1–218.e8. [CrossRef] [PubMed]
26. Geng, J.; Zhang, W.; Ge, Y.; Wang, L.; Huang, P.; Liu, Y.; Shi, J.; Zhou, F.; Ma, K.; Blake, G.M.; et al. Inter-Rater Variability and Repeatability in the Assessment of the Tanner-Whitehouse Classification of Hand Radiographs for the Estimation of Bone Age. *Skelet. Radiol.* **2024**, *53*, 2635–2642. [CrossRef]
27. Alshamrani, K.; Offiah, A.C. Applicability of Two Commonly Used Bone Age Assessment Methods to Twenty-First Century UK Children. *Eur. Radiol.* **2019**, *30*, 504–513. [CrossRef] [PubMed]
28. Yuh, Y.S.; Chou, T.Y.; Tung, T.H. Bone Age Assessment: Large-Scale Comparison of Greulich-Pyle Method and Tanner-Whitehouse 3 Method for Taiwanese Children. *J. Chin. Med. Assoc.* **2023**, *86*, 246–253. [CrossRef] [PubMed]
29. Toledo, F.; Rodríguez, I. *Atlas Radiológico de Referencia de la Edad Ósea en la Población Canaria*; Fundación Canaria Salud y Sanidad, Cabildo de Tenerife: Tenerife, Spain, 2009; p. 22.
30. Ferrante, L.; Cameriere, R. Statistical methods to assess the reliability of measurements in the procedures for forensic age estimation. *Int. J. Legal Med.* **2009**, *123*, 277–283. [CrossRef]
31. Ahmed, M.L.; Warner, J.T. TW2 and TW3 Bone Ages: Time to Change? *Arch. Dis. Child.* **2007**, *92*, 371–372. [CrossRef]
32. Ebrí-Torné, B. Comparative Study Between Bone Ages: Carpal, Metacarpophalangic, Carpometacarpophalangic Ebrí, Greulich and Pyle, and Tanner Whitehouse 2. *Med. Res. Arch.* **2021**, *9*, 1–8. [CrossRef]
33. Noirrit-Esclassan, E.; Valera, M.-C.; Tremollieres, F.; Arnal, J.-F.; Lenfant, F.; Fontaine, C.; Vinel, A. Critical Role of Estrogens on Bone Homeostasis in Both Male and Female: From Physiology to Medical Implications. *Int. J. Mol. Sci.* **2021**, *22*, 1568. [CrossRef]
34. Roberts, D.F. Race, Genetics and Growth. *J. Biosoc. Sci.* **1969**, *1*, 43–67. [CrossRef]
35. Díaz Gómez, N.M. *Crecimiento y Desarrollo Físico del Niño*; Santa Cruz de Tenerife, Spain, 1992; p. 18.
36. Lejarraga, H. Growth in Infancy and Childhood: A Pediatric Approach. In *Human Growth and Development*, 2nd ed.; Academic Press (Elsevier Science): London, UK, 2002; pp. 21–44.
37. Thomis, M.A.; Towne, B. Genetic Determinants of Prepubertal and Pubertal Growth and Development. *Food Nutr. Bull.* **2006**, *27* (Suppl. 4), S257–S278. [CrossRef] [PubMed]
38. Bianchi, M.L.; Sawyer, A.J.; Bachrach, L.K. Rationale for Bone Health Assessment in Childhood and Adolescence. In *Bone Health Assessment in Pediatrics*; Fung, E., Bachrach, L., Sawyer, A., Eds.; Springer: Cham, Switzerland, 2016; pp. 1–21.
39. López, J.M. Bone Development and Growth. *Int. J. Mol. Sci.* **2024**, *25*, 6767. [CrossRef]
40. Perrone, S.; Caporilli, C.; Grassi, F.; Ferrocino, M.; Biagi, E.; Dell'Orto, V.; Beretta, V.; Petrolini, C.; Gambini, L.; Street, M.E.; et al. Prenatal and Neonatal Bone Health: Updated Review on Early Identification of Newborns at High Risk for Osteopenia. *Nutrients* **2023**, *15*, 3515. [CrossRef] [PubMed]
41. Boyanov, M.A. Bone Development in Children and Adolescents. In *Puberty*; Kumanov, P., Agarwal, A., Eds.; Springer: Cham, Switzerland, 2016; pp. 105–123. [CrossRef]
42. Womack, S.R.; Beam, C.R.; Giangrande, E.J.; Scharf, R.J.; Tong, X.; Ponnapalli, M.; Davis, D.W.; Turkheimer, E. Nonlinear Catch-Up Growth in Height, Weight, and Head Circumference from Birth to Adolescence: A Longitudinal Twin Study. *Behav. Genet.* **2023**, *53*, 385–403. [CrossRef]

43. Iuliano-Burns, S.; Hopper, J.; Seeman, E. The Age of Puberty Determines Sexual Dimorphism in Bone Structure: A Male/Female Co-Twin Control Study. *J. Clin. Endocrinol. Metab.* **2009**, *94*, 1638–1643. [CrossRef] [PubMed]
44. Plotkin, L.I.; Bruzzaniti, A.; Pianeta, R. Sexual Dimorphism in the Musculoskeletal System: Sex Hormones and Beyond. *J. Endocr. Soc.* **2024**, *8*, bvae153. [CrossRef] [PubMed]
45. Sallam, A.A.; Briffa, N.; Mahmoud, S.S.; Imam, M.A. Normal Wrist Development in Children and Adolescents: A Geometrical Observational Analysis Based on Plain Radiographs. *J. Pediatr. Orthop.* **2020**, *40*, e860–e872. [CrossRef]
46. Bakos, B.; Takacs, I.; Stern, P.H.; Lakatos, P. Skeletal Effects of Thyroid Hormones. *Clin. Rev. Bone Miner. Metab.* **2018**, *16*, 57–66. [CrossRef]
47. Szulc, P. Biochemical Bone Turnover Markers in Hormonal Disorders in Adults: A Narrative Review. *J. Endocrinol. Investig.* **2020**, *43*, 1409–1427. [CrossRef] [PubMed]
48. Zhu, S.; Pang, Y.; Xu, J.; Chen, X.; Zhang, C.; Wu, B.; Gao, J. Endocrine Regulation on Bone by Thyroid. *Front. Endocrinol.* **2022**, *13*, 873820. [CrossRef]
49. Leung, A.M.; Brent, G.A. The Influence of Thyroid Hormone on Growth Hormone Secretion and Action. In *Growth Hormone Deficiency*; Cohen, L., Ed.; Springer: Cham, Switzerland, 2016; pp. 221–237. [CrossRef]
50. Petrie, K.A.; Burbank, K.; Sizer, P.S.; James, C.R.; Zumwalt, M. Considerations of Sex Differences in Musculoskeletal Anatomy Between Males and Females. In *The Active Female*; Robert-McComb, J.J., Zumwalt, M., Fernandez-del-Valle, M., Eds.; Springer: Cham, Switzerland, 2023; pp. 45–61. [CrossRef]
51. Papalia, D.E.; Martorell, G. *Desarrollo Humano*, 13th ed.; McGraw Hill: Ciudad de México, Mexico, 2017.
52. Satoh, M.; Hasegawa, Y. Factors Affecting Prepubertal and Pubertal Bone Age Progression. *Front. Endocrinol.* **2022**, *13*, 967711. [CrossRef]
53. Nieves, J.W. Sex-Differences in Skeletal Growth and Aging. *Curr. Osteoporos. Rep.* **2017**, *15*, 70–75. [CrossRef]
54. Gavela-Pérez, T.; Garcés, C.; Navarro-Sánchez, P.; López Villanueva, L.; Soriano-Guillén, L. Earlier Menarcheal Age in Spanish Girls Is Related with an Increase in Body Mass Index Between Pre-Pubertal School Age and Adolescence. *Pediatr. Obes.* **2015**, *10*, 410–415. [CrossRef] [PubMed]
55. De Bont, J.; Díaz, Y.; Casas, M.; García-Gil, M.; Vrijheid, M.; Duarte-Salles, T. Time Trends and Sociodemographic Factors Associated with Overweight and Obesity in Children and Adolescents in Spain. *JAMA Netw. Open* **2020**, *3*, e201171. [CrossRef] [PubMed]
56. Itriyeva, K. The Effects of Obesity on the Menstrual Cycle. *Curr. Probl. Pediatr. Adolesc. Health Care* **2022**, *52*, 101241. [CrossRef]
57. Calcaterra, V.; Verduci, E.; Magenes, V.C.; Pascuzzi, M.C.; Rossi, V.; Sangiorgio, A.; Bosetti, A.; Zuccotti, G.; Mameli, C. The Role of Pediatric Nutrition as a Modifiable Risk Factor for Precocious Puberty. *Life* **2021**, *11*, 1353. [CrossRef] [PubMed]
58. Du, Y.; Yan, W.; Bigambo, F.M.; Zhou, Q.; Ma, C.; Gu, W.; Wang, X. Association Between Dietary Behavior and Puberty in Girls. *BMC Pediatr.* **2024**, *24*, 349. [CrossRef]
59. Frenkel, B.; Hong, A.; Baniwal, S.K.; Coetzee, G.A.; Ohlsson, C.; Khalid, O.; Gabet, Y. Regulation of Adult Bone Turnover by Sex Steroids. *J. Cell. Physiol.* **2010**, *224*, 305–310. [CrossRef]
60. Khosla, S.; Monroe, D.G. Regulation of Bone Metabolism by Sex Steroids. *Cold Spring Harb. Perspect. Med.* **2018**, *8*, a031211. [CrossRef]

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.