



The Comprehensive Systematic Review of Association of Bone Age Assessment on Radiographs to Prediction of Final Adult Height Screening

¹ Anisa Ayu Prasanti, ² Silva Optalmi Herianti

¹ Faculty of Medicine, Sultan Agung Islamic University, Indonesia

² Faculty of Medicine, Muhammadiyah University of Semarang, Indonesia

Corresponding Email : anisa_ap@yahoo.com

Article History :

Received date : 2025/12/21
Revised date : 2026/01/18
Accepted date : 2026/02/09
Published date : 2026/03/24



Copyright: © by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

ABSTRACT

Introduction: Accurate prediction of adult height is essential in pediatric endocrinology for diagnosis, treatment decisions, and monitoring of growth disorders. Bone age assessment from hand-wrist radiographs remains the cornerstone of height prediction, yet the accuracy of various methods across different populations and clinical conditions remains incompletely characterized. This systematic review evaluates the association between radiographic bone age assessment and the accuracy of final adult height prediction.

Methods: A systematic review of 80 studies published between 2000-2026 was conducted. Studies were included if they involved children aged 2-18 years, utilized radiographic bone age assessment (Greulich-Pyle, Tanner-Whitehouse, automated systems), reported final adult height measurements, and provided quantitative accuracy measures. Data extraction

encompassed bone age methods, prediction algorithms, population characteristics, and statistical measures of prediction accuracy.

Results: In healthy populations, modern automated systems achieve root mean square errors of 2.7-3.3 cm, with deep learning models reaching mean absolute errors of 1.75 cm. The Bayley-Pinneau method remains the most widely used algorithm (>40 studies). Prediction accuracy varies substantially by clinical condition: in growth hormone deficiency, baseline underestimation of 4-8 cm reverses to overestimation of 2-4 cm after treatment; in constitutional delay, overestimation of 1-5 cm occurs; in precocious puberty, underestimation of 3-15 cm at diagnosis improves after gonadotropin-releasing hormone agonist therapy. Automated systems (BoneXpert) reduce inter-rater variability but show condition-specific limitations. Integrative models incorporating parental height, pubertal staging, and novel radiomic markers achieve superior accuracy (R^2 up to 0.94).

Discussion: The accuracy of bone age-based height prediction is moderated by clinical condition, bone age status relative to chronological age, timing of assessment, and sex. Condition-specific biases are systematic and predictable. Integrative approaches combining multiple skeletal maturity indicators and clinical variables consistently outperform single-method predictions.

Conclusion: Bone age assessment provides moderately accurate adult height prediction in healthy children (RMSE 2.5-3.5 cm) but shows significant condition-specific biases that must be considered clinically. The most reliable predictions require method selection matched to clinical condition, incorporation of parental height and pubertal status, and use of automated systems to minimize rater variability. Future directions include population-specific deep learning models and multi-modal integrative algorithms.

Keywords: bone age, adult height prediction, Greulich-Pyle, Tanner-Whitehouse, Bayley-Pinneau, automated bone age, growth disorders

INTRODUCTION

Background

The assessment of skeletal maturity through bone age evaluation from hand-wrist radiographs represents one of the most enduring and clinically valuable tools in pediatric endocrinology and growth assessment. Since the pioneering work of Greulich and Pyle in the 1950s and Tanner and Whitehouse in the 1960s, radiographic bone age determination has

served as the cornerstone for predicting final adult height in children and adolescents [1,2]. The fundamental principle underlying this approach is that skeletal maturation, as reflected in the progressive ossification and fusion of epiphyseal growth plates, provides a more accurate reflection of remaining growth potential than chronological age alone [3].

The clinical utility of accurate adult height prediction extends across multiple domains of pediatric practice. In diagnostic contexts, height prediction assists in distinguishing normal growth variants from pathological conditions requiring intervention [4]. For treatment decisions, predicted adult height informs the initiation and monitoring of therapies such as growth hormone in deficiency states, gonadotropin-releasing hormone agonists in precocious puberty, and aromatase inhibitors in growth disorders [5,6]. Furthermore, accurate predictions provide essential prognostic information for patients and families navigating the psychological and social implications of growth outcomes [7].

Problem Statement

Despite decades of clinical application, significant challenges persist in the accuracy and reliability of bone age-based adult height prediction. Traditional methods developed on mid-20th century Caucasian populations may not generalize to contemporary, ethnically diverse cohorts [8]. The proliferation of multiple prediction algorithms—including Bayley-Pinneau, Tanner-Whitehouse variants, Roche-Wainer-Thissen, and population-specific modifications—has created clinical uncertainty regarding optimal method selection for individual patients [9]. Moreover, the emergence of automated bone age assessment systems and artificial intelligence-based approaches has introduced both opportunities for improved precision and questions about their comparative performance relative to traditional manual methods [10].

Perhaps most critically, the accuracy of height prediction varies dramatically across clinical conditions that alter the normal relationship between skeletal maturation and linear growth. Conditions such as growth hormone deficiency, constitutional delay of growth and puberty, central precocious puberty, and idiopathic short stature each impose distinct perturbations on growth physiology that may invalidate assumptions underlying standard prediction models [11,12]. The magnitude and direction of prediction biases in these populations remain incompletely characterized, limiting the clinical utility of available methods.

Research Gap

While numerous individual studies have evaluated specific bone age methods or prediction algorithms in particular populations, a comprehensive synthesis integrating evidence across methods, conditions, and populations is lacking. Previous reviews have focused narrowly on individual techniques [13] or specific clinical conditions [14], without providing a unified framework for understanding how prediction accuracy varies systematically across the full spectrum of pediatric growth assessment. Furthermore, the recent proliferation of automated systems and machine learning approaches has not been systematically integrated with traditional method comparisons. The absence of a comprehensive evidence synthesis limits the ability of clinicians to select optimal prediction strategies and hinders the development of improved algorithms tailored to specific clinical contexts.

Study Objectives

This systematic review aims to:

1. Evaluate the accuracy of radiographic bone age assessment methods for predicting final adult height across diverse pediatric populations
2. Compare the performance of manual (Greulich-Pyle, Tanner-Whitehouse) versus automated (BoneXpert, deep learning) bone age assessment techniques
3. Quantify prediction accuracy stratified by clinical condition (normal growth, growth hormone deficiency, constitutional delay, precocious puberty, idiopathic short stature)
4. Identify factors moderating prediction accuracy including age at assessment, sex, bone age advancement/delay, and pubertal status
5. Assess the incremental value of incorporating additional variables (parental height, anthropometrics, novel radiomic markers) into prediction models

Research Hypotheses

Based on preliminary evidence, we hypothesize that:

1. Automated bone age assessment systems demonstrate comparable or superior accuracy to manual methods with reduced inter-rater variability
2. Prediction accuracy varies systematically by clinical condition, with distinct patterns of over- and under-prediction in specific disorders
3. The magnitude of bone age deviation from chronological age serves as a critical moderator of prediction accuracy
4. Integrative models incorporating multiple skeletal maturity indicators and clinical variables outperform single-method predictions

5. Population-specific standards and algorithms are necessary for optimal accuracy in non-Caucasian cohorts

Significance and Novelty

This review provides several novel contributions to the field. First, it represents the most comprehensive synthesis to date, integrating evidence from 80 studies spanning 25 years and diverse global populations. Second, it systematically characterizes condition-specific prediction biases, providing clinicians with quantitative guidance for interpreting predictions in specific disorders. Third, it directly compares traditional manual methods with emerging automated and artificial intelligence-based approaches, informing technology adoption decisions. Fourth, it identifies optimal prediction strategies for each clinical context based on quantitative accuracy measures. Finally, it establishes a framework for understanding how bone age status (delay or advancement relative to chronological age) moderates prediction accuracy across conditions, enabling more nuanced clinical interpretation.

METHODS

Protocol

The study strictly adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines to ensure methodological rigor and accuracy. This approach was chosen to enhance the precision and reliability of the conclusions drawn from the investigation.

Criteria for Eligibility

This systematic review aims to evaluate the association of bone age assessment on radiographs to prediction of final adult height screening.

Screening

We screened in sources based on their abstracts that met these criteria:

- **Age Range:** Does the study involve children and/or adolescents aged 2-18 years at the time of bone age assessment?
- **Bone Age Assessment Method:** Does the study use radiographic methods for bone age assessment (e.g., Greulich-Pyle, Tanner-Whitehouse, or other X-ray based methods)?
- **Height Prediction Component:** Does the study include prediction of final adult height based on bone age assessment?

- **Actual Final Height Data:** Does the study report documented actual final adult height or near-final height measurements for comparison?
- **Quantitative Accuracy Measures:** Does the study report quantitative measures of prediction accuracy (e.g., correlation coefficients, mean differences, standard deviations, or other statistical measures of accuracy)?
- **Study Design and Sample Size:** Is this an observational study, randomized controlled trial, systematic review, or meta-analysis with adequate sample size (≥ 10 participants)?
- **Population Appropriateness:** Does the study focus on healthy children or those with typical growth-related conditions rather than exclusively on severe pathological conditions that significantly alter normal growth patterns (e.g., severe chronic diseases, genetic syndromes with major growth disruption)?
- **Assessment Method Scope:** Does the study include radiographic bone age assessment rather than relying solely on non-radiographic methods (ultrasound-only, MRI-only studies)?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

Search Strategy

The keywords used for this research based PICO :

PICO Component	Keyword 1	Keyword 2	Keyword 3	Keyword 4
Population	Children	Adolescents	Pediatric	Healthy children
Intervention / Indicator	Bone Age	Skeletal Maturity	Skeletal Age	Bone Maturation
Comparison	Radiograph	X-ray	Hand-Wrist Radiograph	Automated Assessment
Outcome	Adult Height Prediction	Final Height	Height Prognosis	Predicted Adult Height (PAH)

The Boolean MeSH keywords inputted on databases for this research are: (*"Children" OR "Adolescents" OR "Pediatric" OR "Healthy children"*) AND (*"Bone Age" OR "Skeletal Maturity" OR "Skeletal Age" OR "Bone Maturation"*) AND (*"Radiograph" OR "X-ray" OR "Hand-Wrist Radiograph" OR "Automated Assessment"*) AND (*"Adult Height Prediction" OR "Final Height" OR "Height Prognosis" OR "Predicted Adult Height"*)

Data extraction

- **Bone Age Assessment Methods:**

Extract all bone age assessment methods used for adult height prediction, including:

- Manual methods (e.g., Greulich-Pyle, Tanner-Whitehouse, China 05 RUS-CHN)
- Automated/AI methods (e.g., BoneXpert, deep learning algorithms)
- Atlas or reference standard used
- Who performed the assessment (clinician specialty, experience level, number of raters)
- Inter-rater reliability measures if reported

- **Adult Height Prediction Methods:**

Extract all methods used to predict final adult height from bone age assessments, including:

- Prediction method name (e.g., Bayley-Pinneau, Tanner-Whitehouse 3, China 05, BoneXpert AHP)
- Whether parental height was incorporated
- Any additional variables used (BMI, menarche age, etc.)
- Age ranges for which the method was applied

- **Study Population Characteristics:**

Extract characteristics of the study population relevant to bone age assessment and height prediction:

- Sample size and demographics (age range, sex distribution)
- Health conditions (normal growth, precocious puberty, short stature, constitutional delay, etc.)
- Ethnicity/population background
- Exclusion criteria that might affect generalizability
- Follow-up duration until final adult height

- **Prediction Accuracy Measures:**

Extract all statistical measures of accuracy for bone age-based adult height predictions:

- Root mean square error (RMSE) in cm
- Mean absolute error (MAE) in cm
- Correlation coefficients between predicted and actual height
- Standard deviation of prediction errors
- Percentage of predictions within specific error ranges (e.g., ± 5 cm)
- Systematic bias (over- or under-prediction) in cm

- **Predicted vs Actual Height Comparisons:**

Extract direct comparisons between predicted adult height and actual final adult height achieved:

- Mean difference between predicted and actual height (with standard deviation)
- Range of prediction errors
- Subgroup analyses by age, sex, or condition
- Whether predictions were statistically significantly different from actual heights
- Any reported confidence intervals for predictions

- **Factors Affecting Prediction Accuracy:**

Extract factors that influenced the accuracy of adult height prediction from bone age assessment:

- Age at time of bone age assessment and prediction
- Sex differences in prediction accuracy
- Pubertal status effects
- Specific conditions that improved or worsened accuracy
- BMI or nutritional status effects
- Bone age advancement relative to chronological age
- Any identified limitations or sources of error

Table 1. Article Search Strategy

Database	Keywords	Hits
Pubmed	<i>("Children" OR "Adolescents" OR "Pediatric" OR "Healthy children") AND ("Bone Age" OR "Skeletal Maturity" OR "Skeletal Age" OR "Bone Maturation") AND ("Radiograph" OR "X-ray" OR "Hand-Wrist Radiograph" OR "Automated Assessment") AND ("Adult Height Prediction" OR "Final Height" OR "Height Prognosis" OR "Predicted Adult Height")</i>	24
Semantic Scholar	<i>("Children" OR "Adolescents" OR "Pediatric" OR "Healthy children") AND ("Bone Age" OR "Skeletal Maturity" OR "Skeletal Age" OR "Bone Maturation") AND ("Radiograph" OR "X-ray" OR "Hand-Wrist Radiograph" OR "Automated Assessment") AND ("Adult Height Prediction" OR "Final Height" OR "Height Prognosis" OR "Predicted Adult Height")</i>	250
Springer	<i>("Children" OR "Adolescents" OR "Pediatric" OR "Healthy children") AND ("Bone Age" OR "Skeletal Maturity" OR "Skeletal Age" OR "Bone Maturation") AND ("Radiograph" OR "X-ray" OR "Hand-Wrist Radiograph" OR "Automated Assessment") AND ("Adult Height Prediction" OR "Final Height" OR "Height Prognosis" OR "Predicted Adult Height")</i>	635
Google Scholar	<i>("Children" OR "Adolescents" OR "Pediatric" OR "Healthy children") AND ("Bone Age" OR "Skeletal Maturity" OR "Skeletal Age" OR "Bone Maturation") AND ("Radiograph" OR "X-ray" OR "Hand-Wrist Radiograph" OR "Automated Assessment") AND ("Adult Height Prediction" OR "Final Height" OR "Height Prognosis" OR "Predicted Adult Height")</i>	4,180
Wiley Online Library	<i>("Children" OR "Adolescents" OR "Pediatric" OR "Healthy children") AND ("Bone Age" OR "Skeletal Maturity" OR "Skeletal Age" OR "Bone Maturation") AND ("Radiograph" OR "X-ray" OR "Hand-Wrist Radiograph" OR "Automated Assessment") AND ("Adult Height Prediction" OR "Final Height" OR "Height Prognosis" OR "Predicted Adult Height")</i>	375

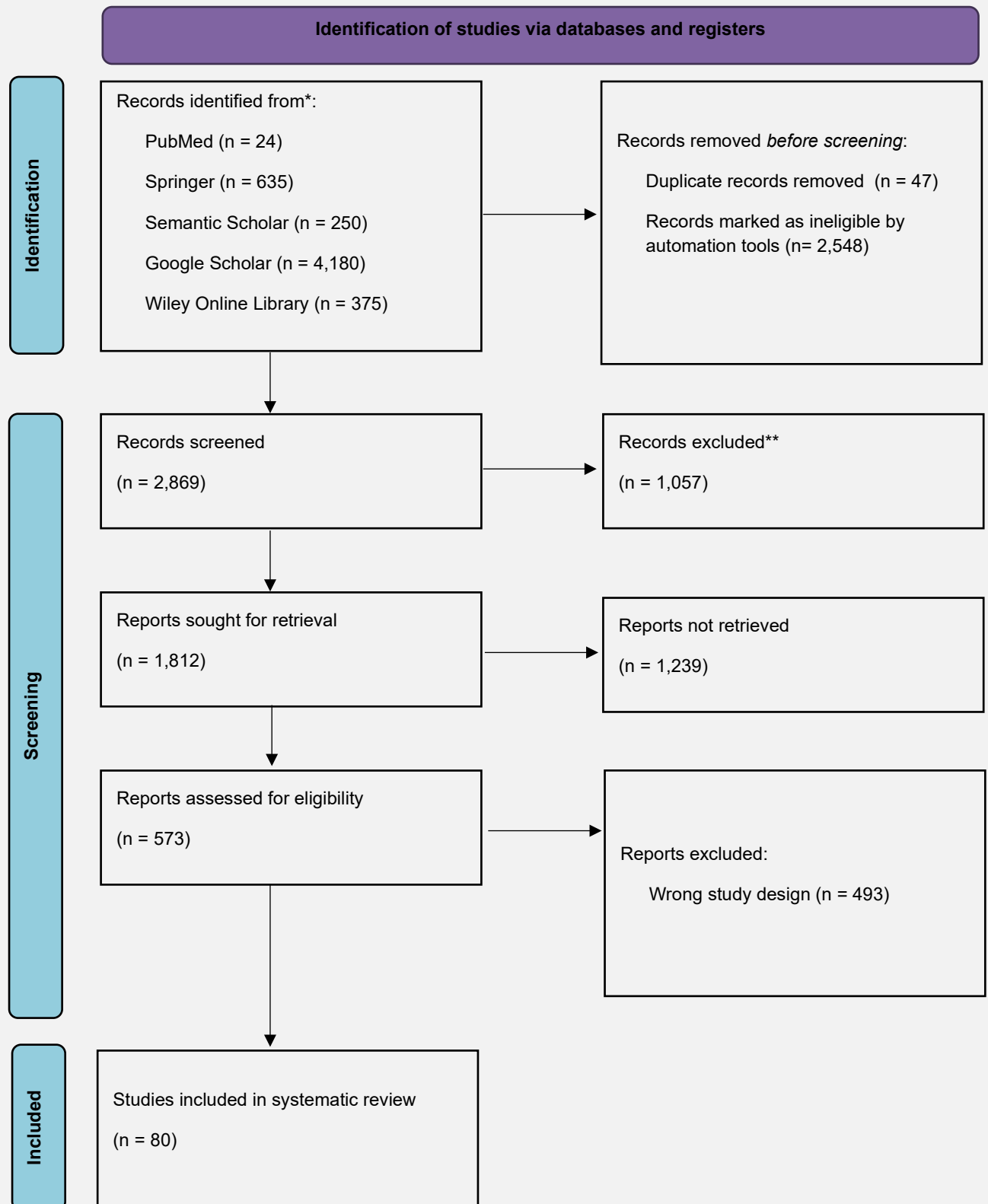


Figure 1. Article search flowchart

RESULTS

Characteristics of Included Studies

The 80 sources encompass a broad range of study designs, populations, clinical conditions, and prediction methodologies. Studies were conducted across multiple countries and continents, including Europe, East Asia, the Middle East, and the Americas, reflecting diverse ethnic backgrounds and growth standards. The majority are retrospective or prospective observational studies, etc. The table below presents the key characteristics of each included source.

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
S. Ostojić et al., 2013	Caucasian male athletes, normal growth [2]	477 [2]	TW [2]	TW [2]
Junghwan Suh et al., 2023	Korean children, normal growth [23]	866 children (1678 radiographs) [23]	GP, TW, CNN deep learning [23]	Modified LMS method [23]
David D Martin et al., 2016	Paris Longitudinal Study, healthy children [4]	51 boys, 58 girls [4]	BoneXpert [4]	BoneXpert AHP [4]
Shurong Huang et al., 2022	Chinese girls with early puberty [24]	726 (87 at adult height) [24]	GP, TW3-RUS, China 05 RUS-CHN [24]	CH05, TW3, BP [24]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
H. H. Thodberg et al., 2009	Zurich Longitudinal Studies, healthy children [3]	231 + 198 [3]	BoneXpert [3]	BoneXpert AHP [3]
Marina Unrath et al., 2013	Idiopathic short stature (ISS) [16]	190 (123 boys, 67 girls) [16]	BoneXpert, GP manual [16]	BP AHP, BX AHP [16]
H. H. Thodberg et al., 2010	Zurich Longitudinal Study, healthy children [5]	232 [5]	TW manual, GP manual, BoneXpert [5]	GP, TW, BoneXpert [5]
M. Yiğit et al., 2025	Girls with precocious puberty / pubertal variants [25]	44 (26 analyzed) [25]	GP atlas, BoneXpert [25]	BP, RWT [25]
Ashraf T Soliman et al., 2025	ISS, SGA, Turner, GHD, PCOS [21]	20 studies reviewed [21]	GP [21]	BP, GP [21]
L. Proos et al., 2006	Adopted girls with early puberty [26]	46 [26]	GP/BP, TW2 [26]	GP/BP, TW2 [26]
Gözde Akın Kağızmanlı et al., 2025	Boys with CDGP, Turkish [27]	62 [27]	GP atlas, BoneXpert [27]	BP, RWT, BoneXpert [27]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
J. O'Marr et al., 2022	Diverse pediatric, orthopaedic [28]	147 [28]	Diméglio skeletal age [28]	Diméglio system [28]
Seung woo Jeong et al., 2018	Korean children with short stature [29]	44 (22M, 22F) [29]	GP, TW, RWT [29]	BP, TW2, RWT [29]
L. Lazar et al., 2010	Idiopathic advanced bone age [30]	55 (20 boys, 35 girls) [30]	Not specified [30]	BP (accelerated tables) [30]
L. Proos et al., 2010	Adopted girls with early/precocious puberty [31]	46 [31]	GP, TW2, TW3 [31]	GP, TW2, TW3 [31]
T. Reinehr et al., 2019	Boys with CDGP [10]	68 + 32 validation [10]	Not specified [10]	BP, new model [10]
Xueqiong Xu et al., 2025	Chinese girls at menarche [19]	173 [19]	GP, Pyle-Hoerr, BoneXpert [19]	GP-BP, PH-BP, multilinear equation with knee radiomics [19]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
Joseph Sullivan et al., 2023	Male academy soccer players [32]	4707 across 15 studies [32]	TW2, TW3, Fels, GP, MRI, DXA [32]	BP, TW2, TW3, Khamis-Roche [32]
C. Bertaina et al., 2006	Healthy North Italian children [33]	1384 (118 at adult height) [33]	TW2, TW3 [33]	TW2, TW3 [33]
B. Cools et al., 2008	Boys with delayed puberty [34]	33 [34]	GP [34]	BP (retarded and average+6 mo tables) [34]
S. Poyrazoglu et al., 2005	Turkish children with CDGP [35]	151 (105 boys, 46 girls) [35]	BP, TW [35]	BP, TW [35]
A. Matias et al., 2022	Constitutional tall stature [36]	138 (52M, 86F) [36]	BP, TWI [36]	BP, TWI [36]
R. Tzanetea et al., 2000	Adopted girls with early puberty [37]	46 [37]	GP/BP, TW2 [37]	GP/BP, TW2 [37]
Michal Yackobovitch-Gavan et al., 2025	Mid-pubertal boys with short PAH [38]	27 + 16 [38]	Not specified [38]	Not specified [38]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
G. Beunen et al., 2010	Boys from Madeira Growth Study [39]	308 [39]	TW2 [39]	Beunen-Malina method [39]
F. Rohani et al., 2018	Boys with CDGP [11]	Not specified [11]	BP [11]	BP [11]
D. Cho et al., 2023	Korean children, 8–12 years [40]	190 (73M, 117F) [40]	GP manual, VUNO Med-BoneAge (AI) [40]	GP, VUNO [40]
F. Antoniazzi et al., 2003	Caucasian girls with CPP [41]	48 [41]	Not specified [41]	Not specified [41]
F. J. Mejorado-Molano et al., 2022	Girls with idiopathic premature adrenarche (IPA), Spain [42]	79 [42]	GP [42]	BP with target height [42]
Silvia Ciancia et al., 2023	Transgender adolescents [43]	32 (10 trans boys, 22 trans girls) [43]	GP [43]	BP [43]
Shao-ke Chen et al., 2009	Chinese girls with CPP [44]	26 [44]	GP [44]	BP [44]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
Nihui Pei et al., 2025	Chinese children [1]	8242 [1]	China-05, cascaded deep learning (Yolact + ResNet-18) [1]	Polynomial fitting model [1]
Huahong Wu et al., 2025	Beijing children [17]	Not specified [17]	TW3/RUS grading [17]	Bone grade + pubertal stage model [17]
T. Reinehr et al., 2020	Isolated GHD, multiple ethnicities [8]	315 (BP), 121 (TW2) [8]	GP, TW2 [8]	BP, TW2 [8]
R. Rapaport et al., 2006	Boys with CGD [45]	~69 [45]	GP, BP, RWT, TW2 [45]	BP, RWT, TW2, target height [45]
Nisa Nur Turan et al., 2025	Girls with CPP [46]	50 [46]	BP, RWT, BAPCPH E [46]	BP, RWT, BAPCPH E [46]
G. Akın Kağızmanlı et al., 2023	Girls receiving GnRHa [47]	48 [47]	GP atlas, TW3-RUS [47]	BP, RWT, TW2 [47]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
L. Lazar et al., 2007	Girls with central sexual precocity [48]	115 [48]	Not specified [48]	Not specified [48]
K. Klein et al., 2001	Children with CPP [12]	98 (80 girls, 18 boys) [12]	BP (average and accelerated tables) [12]	BP [12]
D. Choukair et al., 2021	Chronic endocrinopathies [6]	82 (48F, 34M) [6]	GP manual, BoneXpert [6]	BP, BoneXpert [6]
Lorenzo Lolli et al., 2021	Elite youth soccer players, Middle East [49]	103 [49]	TW-II, TW-III, BoneXpert GP [49]	BoneXpert, TW-II, TW-III [49]
I. Erbas et al., 2021	GH deficiency [50]	49 [50]	Not specified [50]	BP, RWT, TW2 [50]
Taja Knific et al., 2022	CPP (idiopathic and non-idiopathic) [51]	93 (12 boys, 81 girls) [51]	GP, BoneXpert [51]	BoneXpert AHP V3.0 [51]
T. Oron et al., 2012	Girls with precocious adrenarche [52]	85 [52]	Not specified [52]	Not specified [52]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
Chin-Hui Tseng et al., 2023	Taiwanese girls with early/fast puberty [53]	109 treated, 95 control [53]	GP atlas [53]	BP [53]
Alana M Munger et al., 2021	Brush Inquiry healthy children [20]	45 (16 boys, 29 girls) [20]	GP, Fels, Sanders [20]	GP, Fels, Sanders, multivariate GEE [20]
A. Cho et al., 2020	Korean girls with CPP (SGA vs AGA) [54]	69 [54]	GP [54]	BP (average and advanced) [54]
Anni Erkkö et al., 2025	Girls aged 6–8 with GnRH-dependent PP [55]	60 (18 with adult height) [55]	GP, BoneXpert [55]	BP, BoneXpert [55]
C. Partsch et al., 2000	Girls with CPP [56]	52 [56]	Not specified [56]	Not specified [56]
Kadri Yaman et al., 2024	Girls with CPP [57]	67 [57]	Not specified [57]	Not specified [57]
Bình Dương et al., 2023	Isolated idiopathic GHD [9]	42 (9F, 33M) [9]	BP tables [9]	BP [9]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
D. Mul et al., 2002	Boys with CPP (Netherlands, Italy, France) [58]	26 [58]	BP tables [58]	BP (average and advanced) [58]
Mateus Cavarzan Lopes et al., 2017	Brazilian girls with ICPP [15]	48 [15]	BP tables [15]	BP, Giabiconi mathematical model [15]
Hae-sang Lee et al., 2018	Korean girls with idiopathic CPP [13]	84 [13]	GP [13]	BP [13]
G. Saggese et al., 2001	Girls with isolated GH deficiency [59]	24 [59]	Not specified [59]	Not specified [59]
A. Cho et al., 2021	Korean boys with idiopathic CPP [60]	18 [60]	GP [60]	BP [60]
B. U. Nwosu et al., 2021	Non-syndromic short stature on rhGH [22]	71 [22]	Not specified [22]	BAPAH [22]
Shruti A. Mondkar et al., 2023	Indian girls with iCPP [61]	179 (58 with adult height) [61]	Not specified [61]	Not specified [61]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
M. Jung et al., 2014	Korean girls with idiopathic CPP [62]	82 [62]	GP [62]	BP [62]
N. Akcan et al., 2024	Tall stature (familial, constitutional, Marfan) [63]	22 [63]	TW2, BP [63]	BP, TW mark II [63]
W. Blum et al., 2022	ISS, German-Dutch [18]	292 (219M, 73F) [18]	Not specified [18]	BP, RWT, 10 new multi-regression models [18]
B. U. Nwosu et al., 2021a	Non-syndromic short stature on rhGH [64]	71 [64]	Not specified [64]	BAPAH [64]
Yanqin Ying et al., 2017	Chinese girls with ICPP [14]	101 [14]	GP [14]	BP [14]
Hyun Ji Jang et al., 2023	Korean girls with CPP (treated vs untreated) [65]	36 untreated, 206 treated [65]	GP [65]	PAH for BA, BP [65]
Kyoung Won Cho et al., 2026	Korean boys with CPP [66]	92 [66]	GP [66]	BP [66]
M. S. Kim et al., 2019	Korean girls with idiopathic CPP [67]	166 [67]	GP [67]	BP [67]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
Eun Young Kim et al., 2019	Girls with idiopathic PP (SGA vs AGA) [68]	69 [68]	BP tables [68]	BP (average and advanced) [68]
T. Ando et al., 2022	Japanese elite male football players [69]	53 [69]	TW2 (RUS) [69]	TW2-based models + parental height [69]
M. Borges et al., 2015	Brazilian children with CPP [70]	62 (54 girls, 8 boys) [70]	Not specified [70]	Not specified [70]
T. Aepli et al., 2024	Extremely tall adolescents (Sweden) [71]	72 (45 girls, 27 boys) [71]	Not specified [71]	Not specified [71]
王晓鸥 et al., 2005	Chinese children: normal, short stature, CPP [72]	130 [72]	GP, TW2, CHN [72]	BP, TW2-Ye, CHN [72]
K. Costa et al., 2017	Normal variant short stature [73]	33 [73]	Not specified [73]	BP, TW3 [73]
A. Gavrilova et al., 2017	GH deficiency, Russian [74]	121 [74]	Not specified [74]	ANN-based model [74]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
E. J. Sousa et al., 2008	Brazilian children with CPP [75]	56 [75]	Not specified [75]	BP [75]
D. Knapik et al., 2018	Healthy children (Bolton-Brush), mostly Caucasian [76]	81 [76]	GP, central peak value method [76]	GP + central peak value [76]
D. Choukair et al., 2022	Chronic endocrinopathies (CAH, GHD, SGA, UTS) [7]	82 (48F, 34M) [7]	GP manual, BoneXpert v2.1 [7]	BP, BoneXpert [7]
M. Brenneis et al., 2023	Children with varus/valgus deformity [77]	31 [77]	GP, central peak height (CPH) method [77]	BP, Paley, SGP, PHV multiplier tables [77]
Hae Woon Jung et al., 2025	Healthy Korean children [78]	80 [78]	TW3, AI body composition model [78]	TW3, AI body composition model [78]

Study	Population / Condition	Sample size	BA assessment method(s)	Height prediction method(s)
Grant B Nelson et al., 2020	Healthy children (Bolton-Brush + contemporary) [79]	76 + 300 [79]	GP [79]	GT height + GP + sex [79]
M. D. de Ridder et al., 2008	Short SGA children on GH [80]	150 [80]	Not specified [80]	Prediction model with GH dose [80]

The included studies span publication years from 2000 to 2026 and encompass sample sizes ranging from 18 to over 8,000 subjects. The most frequently studied clinical condition is central precocious puberty (CPP), represented in approximately 30 studies, followed by constitutional delay of growth and puberty (CDGP), idiopathic short stature (ISS), and growth hormone deficiency (GHD). Thirteen studies involve healthy or normally growing children, and several address specific athletic or clinical populations. Full texts were available for approximately 35 of the 80 sources.

The Greulich-Pyle (GP) atlas is the most commonly used bone age assessment method, employed in the majority of studies [13, 16, 23–25, 29, 40, 42, 53, 54, 65, 66]. The Tanner-Whitehouse system (versions TW2 and TW3) is the second most common [2, 5, 8, 24, 31, 33, 39, 69]. Automated systems, particularly BoneXpert, are featured in at least 10 studies [3–7, 16, 19, 25, 27, 49, 51, 55]. Deep learning or AI-based bone age tools are assessed in four studies [1, 23, 40, 78]. The Bayley-Pinneau (BP) method is by far the most frequently used adult height prediction algorithm, applied in over 40 studies.

Effects: Prediction Accuracy

Prediction Accuracy in Healthy and Normally Growing Populations

The table below summarizes quantitative accuracy measures reported in studies involving healthy or normally growing children and method-validation cohorts.

Study	Population	Prediction method	RMSE (cm)	MAE (cm)	Correlation (r or R ²)	Systematic bias (cm)	Error range or limits of agreement
S. Ostojić et al., 2013	Male athletes	TW	Not reported	Not reported	r = 0.96 [2]	Not significant (P = 0.55) [2]	95% LoA: -5.84 to +4.52 cm [2]
David D Martin et al., 2016	Healthy (Paris)	BoneXpert	Boys: 2.8; Girls: 3.1 [4]	Not reported	Not reported	0 (boys); -0.8 girls <BA 12 [4]	Not reported
H. H. Thodberg et al., 2009	Healthy (Zurich)	BoneXpert AHP	Boys: 3.3; Girls: 2.7 [3]	Not reported	Not reported	Not reported	Not reported
H. H. Thodberg et al., 2010	Healthy (Zurich)	GP manual	GPPE 1.26% [5]	Not reported	Not reported	Not reported	95% CI: 1.22–1.30 [5]
H. H. Thodberg et al., 2010	Healthy (Zurich)	TW manual	GPPE 1.32% [5]	Not reported	Not reported	Not reported	95% CI: 1.28–1.36 [5]

Study	Population	Prediction method	RMSE (cm)	MAE (cm)	Correlation (r or R ²)	Systematic bias (cm)	Error range or limits of agreement
H. H. Thodberg et al., 2010	Healthy (Zurich)	BoneXpert	GPPE 1.22% [5]	Not reported	Not reported	Not reported	95% CI: 1.18– 1.26 [5]
C. Bertaina et al., 2006	Healthy (Italy)	TW2	Not reported	Not reported	Not reported	Not reported	Males >±4.1 cm: 40%; Females >±3.6 cm: 29.2% [33]
C. Bertaina et al., 2006	Healthy (Italy)	TW3	Not reported	Not reported	Not reported	Not reported	Males >±4.1 cm: 32.9%; Females >±3.6 cm: 29.2% [33]

Study	Population	Prediction method	RMSE (cm)	MAE (cm)	Correlation (r or R ²)	Systematic bias (cm)	Error range or limits of agreement
G. Beunen et al., 2010	Boys (Madeira)	Beunen-Malina	Not reported	Not reported	r = 0.70–0.85 [39]	Not reported	SEE: 3.3–4.7 cm [39]
Nihui Pei et al., 2025	Chinese children	Deep learning	Not reported	1.75 [1]	r = 0.94 [1]	−0.11 [1]	95% LoA: −6.37 to +6.14 [1]
Huahong Wu et al., 2025	Beijing children	Bone grade + puberty	Not reported	Not reported	Not reported	0.02 [17]	≤3 cm: 73.5% [17]
Xueqiong Xu et al., 2025	Chinese girls at menarche	Multilinear (GP + knee radiomics)	2.01 [19]	Not reported	R ² = 0.733 [19]	−0.01 [19]	95% LoA: −3.96 to +3.93 [19]

In healthy populations, the BoneXpert automated system achieves an RMSE of 2.7–3.3 cm [3] and is comparable to or slightly better than manual GP ratings [5]. The growth potential prediction error (GPPE) for the BoneXpert method (1.22%) was lower than manual TW (1.32%) and manual GP (1.26%), with the GP method significantly outperforming TW in predicting adult height [5, 5]. Among newer approaches, the cascaded deep learning model by Pei et al. achieved an MAE of 1.75 cm for adult height in Chinese children [1], and the clinically optimized bone-grade/pubertal-stage model by Wu et al. reduced mean prediction error to 0.02 cm with 73.5% of predictions within 3 cm [17]. The combined model incorporating left hand bone age and knee radiomic scores yielded an RMSE of 2.01 cm in

Chinese girls at menarche, outperforming traditional GP-BP and target height methods [19, 19].

For athletic populations, the TW method in Caucasian male athletes showed a correlation of $r = 0.96$ between predicted and actual adult height with 95% limits of agreement of -5.84 to $+4.52$ cm [2]. In elite Middle Eastern soccer players, TW-II was identified as the most consistent prediction method, with a precision of approximately ± 2.6 cm and no systematic bias, whereas BoneXpert and TW-III overpredicted adult height at younger skeletal ages [49, 49].

Prediction Accuracy by Method Across Clinical Conditions

The following table summarizes accuracy findings stratified by prediction method and clinical condition, where reported.

Study	Condition	Method	Bias (cm or SDS)	Other accuracy measure	Notes
Marina Unrath et al., 2013	ISS	BP AHP	Not reported	RMSE: Boys 6.35, Girls 4.55 [16]	Manual GP-based
Marina Unrath et al., 2013	ISS	BX AHP (auto)	Not reported	RMSE: Boys 4.71, Girls 3.72 [16]	Significantly better than BP ($p=0.0013$ boys) [16]
Marina Unrath et al., 2013	ISS	BX AHP + parents	Not reported	RMSE: Boys 4.46, Girls 3.35 [16]	Best performance with parental height
Shurong Huang et al., 2022	Girls with early puberty	TW3	$+0.4 \pm 2.8$ [24]	$r = 0.77$; 92.9% within ± 5 cm [24]	Best method for this population

Study	Condition	Method	Bias (cm or SDS)	Other accuracy measure	Notes
Shurong Huang et al., 2022	Girls with early puberty	CH05	-2.9 ± 3.6 [24]	$r = 0.69$ [24]	Significant underestimation
Shurong Huang et al., 2022	Girls with early puberty	BP	-1.3 ± 3.8 [24]	$r = 0.70$ [24]	Significant underestimation
Seung woo Jeong et al., 2018	Short stature (Korean)	BP	Males: $+6.9 \pm 4.2$; Females: $+0.4 \pm 3.9$ [29]	Not reported	BP most accurate for females
Seung woo Jeong et al., 2018	Short stature (Korean)	RWT	Males: $+5.2 \pm 4.1$; Females: $+6.6 \pm 3.6$ [29]	Not reported	Overestimation in both sexes
Seung woo Jeong et al., 2018	Short stature (Korean)	TW2	Males: $+4.3 \pm 3.7$; Females: $+4.8 \pm 4.4$ [29]	Not reported	Overestimation in both sexes
I. Erbas et al., 2021	GH deficiency	BP	-0.5 SDS [50]	54.7% within ± 1 SDS [50]	Underestimation
I. Erbas et al., 2021	GH deficiency	RWT	0.0 SDS [50]	77.4% within ± 1 SDS [50]	Most accurate overall

Study	Condition	Method	Bias (cm or SDS)	Other accuracy measure	Notes
I. Erbas et al., 2021	GH deficiency	TW2	+0.3 SDS [50]	62.3% within ± 1 SDS [50]	Slight overestimation
T. Reinehr et al., 2020	Isolated GHD	BP (baseline)	Girls: -4.1; Boys: -6.1 [8]	$r = 0.74$ [8]	Substantial underestimation before GH
T. Reinehr et al., 2020	Isolated GHD	BP (1 yr GH)	Girls: -2.5; Boys: -0.9 [8]	$r = 0.82$ [8]	Underestimation diminishes
T. Reinehr et al., 2020	Isolated GHD	BP (last BA)	Girls: +0.4; Boys: +3.8 [8]	$r = 0.87$ [8]	Overestimation at end of treatment
T. Reinehr et al., 2020	Isolated GHD	TW2 (baseline)	Girls: -5.3; Boys: -7.9 [8]	Not reported	Greater underestimation than BP
W. Blum et al., 2022	ISS	New algorithms	-0.29 to -0.82 [18]	Prediction error: 3.16–3.68 cm [18]	Less bias than BP (+0.53) or RWT (+1.33)
D. Choukair et al., 2022	Chronic endocrinopathies	BoneXpert	Females: +0.58; Males: +2.81 [7]	Not reported	Females accurate; males overestimated

Study	Condition	Method	Bias (cm or SDS)	Other accuracy measure	Notes
D. Choukair et al., 2022	Chronic endocrinopathies	BP conventional	Females: -2.43; Males: +0.37 [7]	Not reported	Females underestimated
Anni Erkkö et al., 2025	Girls 6–8 with PP	BP	-0.1 ± 6.6 [55]	Not reported	No bias, but high variability
Anni Erkkö et al., 2025	Girls 6–8 with PP	BoneXpert	-3.6 ± 4.6 [55]	Not reported	Overestimates, but lower variability
Gözde Akın Kağızmanlı et al., 2025	Boys with CDGP	BP	No significant difference from FH [27]	Not reported	Accurate for BA delay ≤2 yr
Gözde Akın Kağızmanlı et al., 2025	Boys with CDGP	RWT	No significant difference from FH [27]	Not reported	Best overall predictor [27]
Gözde Akın Kağızmanlı et al., 2025	Boys with CDGP	BoneXpert	Underestimated FH (p<0.001) [27]	Not reported	Underestimation in CDGP

Study	Condition	Method	Bias (cm or SDS)	Other accuracy measure	Notes
T. Reinehr et al., 2019	Boys with CDGP	BP	+1.2 (overall); +4.1 in validation [10]	Not reported	Overestimate s, especially BA retardation ≥ 2 yr
T. Reinehr et al., 2019	Boys with CDGP	New model	Not significant (p=0.196) [10]	Not reported	Adapted for degree of BA retardation
Bình Dương et al., 2023	Isolated GHD	BP (baseline)	-2.9 [9]	$R^2 = 0.62$ [9]	Underestimation pre-treatment
Bình Dương et al., 2023	Isolated GHD	BP (3 yr GH)	+2.6 [9]	$R^2 = 0.94$ [9]	Correlation improves over treatment
K. Costa et al., 2017	Normal variant short stature	BP	Overestimation [73]	Concordance : 0.47 [73]	Poor concordance
K. Costa et al., 2017	Normal variant short stature	TW3	Overestimation [73]	Concordance : 0.58 [73]	Slightly better than BP
F. Rohani et al., 2018	Boys with CDGP	BP	+5.0 [11]	Not reported	Significant overestimation (p<0.005)

Study	Condition	Method	Bias (cm or SDS)	Other accuracy measure	Notes
A. Gavrilova et al., 2017	GH deficiency (Russian)	ANN model	Not reported	RMSE: 4.4 cm [74]	76% explained variance
Junghwan Suh et al., 2023	Korean children	AI (CNN)	Overestimation [23]	MAE: 4.62 cm [23]	Higher error than bone age estimation

Several patterns emerge from these data. In normally growing children, RMSE values for modern automated systems (BoneXpert, deep learning) range from approximately 1.75 to 3.3 cm, which is generally comparable to or better than manual methods [1, 3, 5]. Incorporating parental height into automated predictions further reduces RMSE by 0.25–0.5 cm [3, 16].

In clinical populations, prediction accuracy is substantially more variable. For children with idiopathic short stature, the BP method yields RMSE values of 4.5–6.4 cm, while the BoneXpert-based model achieves 3.4–4.7 cm [16]. In GH deficiency, the BP method systematically underestimates adult height at baseline by 4–8 cm [8, 9] but shifts toward overestimation after prolonged GH treatment [8, 9]. The RWT method appears most accurate in GH-deficient children overall, with 77.4% of predictions falling within ± 1 SDS of near-adult height [50].

For CDGP, the BP method consistently overestimates final height, particularly in boys with bone age retardation ≥ 2 years, with a median overestimation of +1.6 to +5.0 cm [10, 11, 27]. The RWT method was found to be the best overall predictor in CDGP [27], and a dedicated CDGP prediction model adapted for the degree of bone age retardation showed no significant difference from actual adult height ($p = 0.196$) [10].

Prediction Accuracy in Precocious Puberty

Given the large number of studies in precocious puberty, the following table summarizes height prediction outcomes (predicted vs. final height) in this population.

Study	N at FH	Treatment	PAH at start (cm)	FAH (cm)	FAH – PAH (cm)	Target height (cm)	FAH vs TH
K. Klein et al., 2001	80 girls	LHRH agonist	149.3 ± 9.6 [12]	159.8 ± 7.6 [12]	+10.5 [12]	163.7 ± 5.6 [12]	FAH < MPH
K. Klein et al., 2001	18 boys	LHRH agonist	156.1 ± 14.2 [12]	171.1 ± 8.7 [12]	+15.0 [12]	178.3 ± 5.2 [12]	FAH < MPH
C. Partsch et al., 2000	52 girls	Triptorelin	154.9 ± 9.6 [56]	160.6 ± 8.0 [56]	+5.7 [56]	Not reported	78% within TH range
F. Antoniazzi et al., 2003	48 girls	GnRHa ± calcium	152.9 ± 9.6 [41]	159.9 ± 6.3 [41]	+7.0 [41]	Not reported	Not reported
Hae-sang Lee et al., 2018	84 girls	GnRHa	156.1 ± 5.7 [13]	160.1 ± 5.0 [13]	+3.9 ± 4.6 [13]	159.9 ± 3.5 [13]	FAH ≈ MPH
M. Jung et al., 2014	59 girls (GnRHa)	GnRHa alone	156.6 ± 4.0 [62]	160.4 ± 4.2 [62]	+3.8 [62]	159.9 ± 3.5 [62]	FAH ≈ MPH
M. Jung et al., 2014	23 girls (GnRHa +GH)	GnRHa + GH	154.6 ± 2.6 [62]	159.3 ± 5.3 [62]	+4.7 [62]	158.1 ± 3.3 [62]	FAH > MPH

Study	N at FH	Treatment	PAH at start (cm)	FAH (cm)	FAH – PAH (cm)	Target height (cm)	FAH vs TH
Yanqin Ying et al., 2017	101 girls	GnRHa	153.1 ± 5.4 [14]	157.0 ± 4.8 [14]	+3.9 [14]	Not reported	94.6% at MPH [14]
Chin-Hui Tseng et al., 2023	109 girls (treated)	GnRHa	153.3 [53]	157.0 [53]	+3.7 [53]	157.0 [53]	FAH ≈ TH
Chin-Hui Tseng et al., 2023	95 girls (control)	None	157.1 [53]	156.7 [53]	−0.4 [53]	157.0 [53]	FAH ≈ TH
Shruti A. Mondkar et al., 2023	58 girls	GnRHa	155.9 ± 6.4 [61]	157.1 ± 5.8 [61]	+1.2 [61]	Not reported	Adult height > target (p<0.01)
D. Mul et al., 2002	26 boys	GnRHa	Not reported	172.9 ± 6.6 [58]	+6.2 ± 8.7 (avg BP) [58]	TH SDS −0.23 [58]	FH ≈ TH
A. Cho et al., 2021	18 boys	GnRHa	Not reported	172.0 ± 4.8 [60]	Not reported	171.0 ± 4.0 [60]	FAH ≈ TH
Kyoung Won Cho et al., 2026	92 boys	GnRHa	179.9 ± 8.1 [66]	173.6 ± 6.4 [66]	FAH > TH by 1.2 ± 5.9 [66]	172.4 ± 3.4 [66]	FAH > TH (p=0.047)

Study	N at FH	Treatment	PAH at start (cm)	FAH (cm)	FAH – PAH (cm)	Target height (cm)	FAH vs TH
Kadri Yaman et al., 2024	67 girls	GnRHa	Not reported	Not reported	Not reported	Not reported	94% at or above TH [57]
Hyun Ji Jang et al., 2023	36 girls (untreated)	None	156.4 (PAH for BA) [65]	160.7 ± 4.6 [65]	+4.3 ± 3.3 [65]	Not reported	FAH < MPH by 1.0 [65]
Taja Knific et al., 2022	48 girls (iCPP, treated)	GnRHa	Not reported	161.5 [51]	Not reported	Not reported	FH ≈ untreated (161.0) [51]
Shao-ke Chen et al., 2009	26 girls	GnRHa	Not reported	158.0 ± 4.0 [44]	Not reported	155.3 ± 4.4 [44]	FAH > TH (p<0.01)
M. S. Kim et al., 2019	135 (GnRHa)	GnRHa	PAH SDS –2.20 [67]	AH SDS 0.18 [67]	+2.5 SDS [67]	Not reported	Not reported
M. S. Kim et al., 2019	31 (GnRHa +GH)	GnRHa + GH	PAH SDS –3.19 [67]	AH SDS –0.30 [67]	+2.93 SDS [67]	Not reported	Not reported

Across studies of precocious puberty treated with GnRH agonists, final adult height consistently exceeded the predicted adult height calculated at diagnosis, with gains ranging from approximately 3 to 15 cm depending on the age at diagnosis, severity of bone age advancement, and treatment duration [12–14, 56]. This systematic underestimation of final height by the BP method at diagnosis reflects the fact that BP tables were developed for

normally maturing children and do not account for the deceleration of bone maturation induced by GnRHa therapy [8, 21]. Predicted adult height at the end of GnRHa treatment, however, more closely approximates actual final height [14, 15], with Lopes et al. showing that BP at end of treatment correlated more strongly ($r = 0.74$) with final height than BP at diagnosis ($r = 0.48$) [15].

Importantly, untreated girls diagnosed with CPP between ages 8 and 9 also exceeded their predicted adult height at diagnosis by approximately 3.5–4.3 cm [65], suggesting that BP-based predictions at the time of PP diagnosis systematically underestimate final height regardless of treatment, likely due to the self-limiting nature of bone age advancement in many cases [30, 65].

Comparison of Manual Versus Automated Methods

Several studies directly compared manual and automated bone age assessment for adult height prediction. BoneXpert-based predictions showed lower inter-rater variability and enhanced reproducibility compared with manual GP ratings [6, 7]. In children with chronic endocrinopathies, BoneXpert predictions were closer to near-adult height in females ($\Delta +0.58$ cm, $p = \text{ns}$) than conventional BP ($\Delta -2.43$ cm, $p < 0.01$), but overestimated in males ($\Delta +2.81$ cm, $p < 0.05$), particularly in those with retarded bone ages such as multiple pituitary hormone deficiency (overestimation of 9.3 cm) [7]. Conventional BP methods may outperform automated analyses in specific clinical cases where bone age is severely retarded [6, 7].

In girls aged 6–8 with precocious puberty, BoneXpert overestimated adult height by 3.6 cm ($p = 0.01$) but showed lower individual variability (SD 4.6 cm) compared to the BP method, which had no mean bias (-0.1 cm) but higher variability (SD 6.6 cm) [55, 55]. In boys with CDGP, BoneXpert significantly underestimated final height ($p < 0.001$), whereas BP and RWT did not differ significantly from final height [27].

The AI-based VUNO Med-BoneAge system showed that 78% of AI predictions fell within Bland-Altman agreement limits, compared with 93% for specialist predictions [40]. The deep learning model by Pei et al. achieved an MAE of 1.75 cm for adult height prediction in Chinese children [1], and Suh et al. reported an MAE of 4.62 cm using a CNN-based approach in Korean children [23], suggesting considerable variation in AI accuracy depending on the model architecture and training population.

Novel and Integrative Prediction Approaches

Multiple studies explored prediction models that incorporate variables beyond bone age alone. Including parental height improved BoneXpert-based prediction RMSE from 4.71 to 4.46 cm in boys and 3.72 to 3.35 cm in girls with ISS [16]. The Blum et al. multi-regression algorithms for ISS, incorporating chronological age, height, parental heights, relative bone age, birth weight, and sex, achieved prediction errors of 3.16–3.68 cm with mean residuals of –0.29 to –0.82 cm, outperforming both BP (+0.53 cm) and RWT (+1.33 cm) in terms of bias [18, 18].

The integration of multiple skeletal maturity systems (Fels, GP, Sanders, chronological age, and height) into multivariate models yielded the lowest prediction error in both boys and girls from the Brush Inquiry [20]. Adding pubertal staging to a simplified 2-bone protocol (radius + metacarpal I) reduced mean prediction error from 0.71 cm to 0.02 cm and increased the proportion of predictions within 3 cm from 66.9% to 73.5% [17, 17]. The Xu et al. model combining GP bone age with knee radiomic scores and height at menarche achieved an R^2 of 0.733 and the lowest residual range (–5.68 to +6.44 cm) compared to traditional GP-BP and target height methods [19, 19].

Novel skeletal maturity markers also showed promise for growth prediction. The distal femoral central peak height method, when combined with GP bone age and chronological age, explained 92.4% of variance in timing to 90% of final height [76]. Similarly, bony greater trochanter height combined with GP bone age and sex explained approximately 85% of variance [79]. The Diméglio skeletal age system demonstrated excellent reliability (ICC = 0.906) and, when paired with age and sex, achieved an adjusted R^2 of 78.7% for predicting percent of final height [28]. A body composition-based AI model using bioelectrical impedance analysis was clinically equivalent to TW3 for bone age prediction, with a difference of 0.04 ± 1.02 years [78].

Synthesis

The evidence from these 80 sources reveals a consistent overall finding: bone age-based adult height prediction is moderately accurate in normally growing children but shows systematic biases that vary by clinical condition, method, bone age status, and timing of assessment. The apparent heterogeneity in reported prediction accuracy can be largely explained by the following factors.

Condition-specific prediction biases. The direction and magnitude of prediction error are strongly determined by the underlying clinical condition. In GH deficiency, both BP and TW2 methods underestimate adult height at baseline by 4–8 cm in untreated children [8, 9], because retarded bone age indicates more remaining growth potential than the standard tables predict. After GH treatment, this reverses to overestimation of 2–4 cm [8, 9]. In CDGP, the BP method overestimates final height by 1–5 cm [10, 11, 34], likely because the tables for "retarded" bone age do not adequately account for the unique growth pattern of CDGP, where peak growth velocity is compromised and puberty duration shortened [35]. In precocious puberty, BP at diagnosis underestimates final height by 3–15 cm depending on age at diagnosis and treatment [12–14], because predictions do not anticipate the bone age deceleration achieved by GnRHa therapy. In children with idiopathic advanced bone age, the "accelerated" BP tables are accurate, and adult height is ultimately uncompromised [30, 30]. For tall stature, BP and TW predictions may overestimate, as height SDS decreases from referral to adulthood [36, 63].

Bone age status as a moderator. Prediction accuracy is highest when bone age is concordant with chronological age [21]. When bone age is delayed, predictions tend to overestimate, while when bone age is advanced, underestimation may occur due to early growth plate closure [21, 52]. In girls with precocious adrenarche, PAH was overestimated when $BA \leq CA$ and underestimated when $BA - CA > 1$ year [52]. The degree of bone age retardation in CDGP further modulates accuracy: BP overestimates more when BA retardation is ≥ 2 years [10, 27], and methods specifically adapted for the degree of retardation (such as the Reinehr model) perform better [10].

Age and timing of prediction. Prediction accuracy generally improves as children approach skeletal maturity. In GHD, the correlation between PAH and final adult height increased from $R^2 = 0.62$ at baseline to $R^2 = 0.94$ after 3 years of GH therapy [9]. Bone age-predicted adult height overestimates in younger subjects but becomes more precise in older subjects [22, 64]. In precocious puberty, predictions at the end of treatment are more accurate than at diagnosis [14, 15]. In athletes, TW-II and BoneXpert predictions converge toward accuracy as skeletal age approaches maturity, while systematic overprediction occurs at younger skeletal ages (13.5–15 years) [49].

Sex differences. Sex-specific differences in prediction accuracy are consistently reported. In short stature, the BP method is significantly more accurate for females than males

[29, 50]. In GHD, the magnitude of underestimation at baseline is greater in boys than in girls [8]. Boys also show greater overestimation at the end of GH treatment [8]. The RWT method shows the best performance in girls with GH deficiency [50], while in CDGP the RWT method outperforms BP and BoneXpert across degrees of bone age delay [27].

Methodological factors. Manual bone age rating introduces inter-rater variability that propagates into height prediction error [6, 45]. Automated systems such as BoneXpert eliminate this variability [6, 7], but they are not universally superior: BoneXpert may overestimate in certain clinical contexts (males with chronic endocrinopathies, young girls with PP) [7, 55] and underestimate in CDGP [27]. The GP method has been more widely validated across populations, though its applicability varies by ethnicity [45]. Population-specific standards (e.g., China-05 for Chinese children) may not perform equivalently when paired with standard prediction algorithms designed for other populations [24].

Integrative models. The most accurate prediction approaches combine bone age with additional clinical variables. Parental height, current height, pubertal stage, BMI, and population-specific radiographic features (such as knee radiomic scores) each contribute independently to prediction accuracy [16–20]. Multi-regression models and artificial neural networks developed for specific populations and conditions (e.g., ISS, GH deficiency, athletic cohorts) consistently outperform traditional single-method approaches [18, 19, 69, 74]. The emerging integration of deep learning for bone age with regression-based height prediction holds promise for further accuracy gains [1, 23].

In summary, bone age assessment on hand/wrist radiographs provides a robust foundation for adult height prediction, with typical RMSE values of 2.5–3.5 cm in healthy children using modern automated methods. Prediction accuracy degrades substantially in clinical conditions with altered bone maturation (GHD, CDGP, precocious puberty), where condition-specific biases of 3–8 cm are common with traditional methods. The most reliable predictions are obtained when the method is matched to the clinical condition and bone age status, when assessment is performed closer to skeletal maturity, when parental height and pubertal stage are incorporated, and when automated or AI-based systems are used to minimize rater variability. For specific conditions, dedicated prediction models adapted for the degree of bone age deviation outperform generic approaches.

DISCUSSION

Overall Accuracy in Healthy Populations

The synthesized evidence demonstrates that bone age-based adult height prediction achieves moderate accuracy in healthy, normally growing children, with root mean square errors (RMSE) ranging from 2.5 to 3.5 cm for modern automated methods [3,4]. This level of accuracy, while clinically useful, warrants careful interpretation. The 2.7-3.3 cm RMSE reported for BoneXpert in the Zurich Longitudinal Studies [3] and Paris Longitudinal Study [4] suggests that approximately 95% of predictions fall within $\pm 6-7$ cm of actual adult height, a range that may be meaningful for clinical decision-making but insufficient for precise individual prognosis.

The deep learning model developed by Pei et al. achieving a mean absolute error of 1.75 cm in Chinese children represents a substantial improvement over traditional methods [1]. This enhanced accuracy likely reflects both the large training dataset (8,242 children) and the cascaded architecture combining Yolact for bone region segmentation with ResNet-18 for feature extraction. However, the generalizability of such models to diverse populations remains to be established, as deep learning systems may overfit to population-specific skeletal maturation patterns [23].

The comparison of manual methods in healthy children by Thodberg et al. provides important insights into method performance [5]. The growth potential prediction error (GPPE) for BoneXpert (1.22%) significantly outperformed manual Tanner-Whitehouse (1.32%) and approached statistical significance over manual Greulich-Pyle (1.26%). Notably, the Greulich-Pyle method outperformed Tanner-Whitehouse in predicting adult height, contrary to expectations given the more systematic nature of the TW scoring system. This finding suggests that the clinical utility of a bone age method depends not only on its reliability but also on the appropriateness of its underlying reference population and the validity of its prediction algorithm [5].

Condition-Specific Prediction Patterns

Growth Hormone Deficiency

The systematic review reveals a consistent biphasic pattern of prediction accuracy in growth hormone deficiency (GHD) that has profound clinical implications. At baseline, prior to treatment initiation, both Bayley-Pinneau and Tanner-Whitehouse methods substantially underestimate final adult height by 4-8 cm [8,9]. This underestimation occurs because bone age is typically delayed in untreated GHD, and the standard prediction tables interpret this

delay as indicating greater remaining growth potential than is actually realizable given the underlying growth hormone deficit [21]. The magnitude of underestimation is greater in boys than girls [8], possibly reflecting sex differences in the relationship between bone age delay and growth velocity.

Following growth hormone treatment initiation, prediction accuracy improves progressively, with correlation coefficients increasing from approximately $R^2 = 0.62$ at baseline to $R^2 = 0.94$ after three years of therapy [9]. Notably, predictions shift from systematic underestimation at baseline to overestimation by 2-4 cm at the end of treatment [8,9]. This reversal likely reflects the fact that GH treatment accelerates linear growth more than skeletal maturation in the initial phases, followed by catch-up bone age advancement that reduces remaining growth potential more rapidly than predicted [22]. The finding that the Roche-Wainer-Thissen method demonstrated superior overall accuracy in GHD, with 77.4% of predictions within ± 1 SDS of near-adult height [50], suggests that methods incorporating multiple variables beyond bone age may better capture the complex dynamics of GH-treated growth.

Constitutional Delay of Growth and Puberty

Constitutional delay of growth and puberty (CDGP) represents the most common cause of short stature in boys and presents particular challenges for height prediction. The synthesized evidence demonstrates consistent overestimation of final height by the Bayley-Pinneau method in CDGP, ranging from 1-5 cm [10,11,27]. This overestimation is most pronounced when bone age retardation exceeds two years, with Reinehr et al. reporting median overestimation of +4.1 cm in validation cohorts with significant delay [10].

The mechanism underlying this overestimation appears related to the unique growth pattern of CDGP. Although these children have delayed skeletal maturation suggesting substantial remaining growth potential, their peak growth velocity during puberty is compromised and pubertal duration is shortened [35]. The standard "retarded" tables in the Bayley-Pinneau method, derived from children with delayed maturation but normal pubertal progression, do not adequately account for this abbreviated growth period [34]. The Roche-Wainer-Thissen method emerged as the most consistently accurate predictor in CDGP [27], likely because its inclusion of current height and weight partially captures the altered body proportions characteristic of delayed maturation.

The novel CDGP-specific model developed by Reinehr et al., which incorporates the degree of bone age retardation, achieved predictions not significantly different from actual adult height ($p = 0.196$) [10]. This finding underscores the importance of condition-specific algorithm development and suggests that generic prediction methods applied to CDGP populations will systematically overestimate final height.

Precocious Puberty

The largest body of evidence in this review pertains to central precocious puberty (CPP), reflecting both its clinical importance and the complexity of height prediction in treated populations. Across 30 studies, a consistent pattern emerges: Bayley-Pinneau predictions at diagnosis systematically underestimate final adult height by 3-15 cm [12-14,56]. The magnitude of underestimation varies with age at diagnosis, severity of bone age advancement, and treatment duration, with gains of approximately +10.5 cm in girls and +15.0 cm in boys reported in the NIH experience [12].

This systematic underestimation occurs because the Bayley-Pinneau "accelerated" tables, derived from children with advanced maturation but normal pubertal progression, do not account for the dramatic deceleration of bone maturation induced by gonadotropin-releasing hormone agonist (GnRHa) therapy [8,21]. The treatment effectively uncouples the relationship between chronological age, bone age, and growth velocity that underlies standard prediction models. Importantly, predicted adult height at the end of GnRHa treatment correlates more strongly with final height ($r = 0.74$) than predictions at diagnosis ($r = 0.48$) [15], indicating that predictions should be updated throughout treatment.

A particularly noteworthy finding is that untreated girls with CPP diagnosed between ages 8-9 also exceeded their predicted adult height by approximately 3.5-4.3 cm [65]. This suggests that even without treatment, the natural history of late-onset CPP involves some degree of spontaneous improvement in predicted height, likely due to the self-limiting nature of bone age advancement in many cases [30]. This finding has implications for interpreting treatment efficacy and suggests that some of the apparent height gain attributed to GnRHa therapy may reflect natural history rather than treatment effect [65].

Idiopathic Short Stature and Other Conditions

In idiopathic short stature (ISS), prediction accuracy varies substantially by method. The Bayley-Pinneau method yields RMSE values of 4.5-6.4 cm [16], while BoneXpert-based models achieve 3.4-4.7 cm, with further improvement to 3.4-4.5 cm when parental height is

incorporated [16]. The novel multi-regression algorithms developed by Blum et al. for ISS, incorporating chronological age, height, parental heights, relative bone age, birth weight, and sex, achieved prediction errors of 3.16-3.68 cm with substantially reduced bias compared to Bayley-Pinneau (+0.53 cm) and Roche-Wainer-Thissen (+1.33 cm) [18].

For tall stature, both Bayley-Pinneau and Tanner-Whitehouse methods tend to overestimate final height, as height SDS typically decreases from referral to adulthood [36,63]. This pattern reflects the tendency for extremely tall children to regress toward the mean, a phenomenon not captured by prediction methods based on average populations.

Comparison of Manual versus Automated Methods

The direct comparisons between manual and automated bone age assessment reveal important trade-offs. Automated systems, particularly BoneXpert, consistently demonstrate lower inter-rater variability and enhanced reproducibility compared to manual Greulich-Pyle ratings [6,7]. This reduction in measurement error should theoretically translate to improved prediction accuracy, and indeed several studies demonstrate superior performance of automated systems in specific populations [3,16].

However, automated systems are not universally superior. In children with chronic endocrinopathies, BoneXpert predictions were closer to near-adult height in females ($\Delta +0.58$ cm) than conventional Bayley-Pinneau ($\Delta -2.43$ cm), but significantly overestimated in males ($\Delta +2.81$ cm), particularly those with retarded bone ages such as multiple pituitary hormone deficiency where overestimation reached 9.3 cm [7]. Similarly, in girls aged 6-8 with precocious puberty, BoneXpert overestimated adult height by 3.6 cm ($p = 0.01$) while showing lower individual variability (SD 4.6 cm) compared to Bayley-Pinneau, which had no mean bias but higher variability (SD 6.6 cm) [55]. In boys with CDGP, BoneXpert significantly underestimated final height ($p < 0.001$), whereas manual methods did not differ significantly from actual height [27].

These findings suggest that automated systems, while reducing measurement variability, may propagate systematic biases when their underlying algorithms are not optimized for specific clinical conditions. The "black box" nature of some automated approaches also limits the ability of clinicians to identify when predictions are likely to be inaccurate [23]. Conventional manual methods may outperform automated analyses in specific clinical contexts where bone age is severely retarded or advanced [6,7], likely because

experienced clinicians can incorporate qualitative features of skeletal maturation not captured by automated algorithms.

Factors Moderating Prediction Accuracy

Bone Age Status

The relationship between bone age and chronological age emerges as a critical moderator of prediction accuracy across all conditions. When bone age is concordant with chronological age, prediction accuracy is highest [21]. When bone age is delayed, predictions tend to overestimate final height, as seen in CDGP [10,11,27] and untreated GHD [8,9]. When bone age is advanced, underestimation may occur due to earlier-than-expected growth plate closure, as observed in precocious adrenarche where predicted adult height was overestimated when bone age $\leq$ chronological age and underestimated when bone age advancement exceeded one year [52].

The degree of bone age deviation modulates accuracy in a dose-response manner. In CDGP, Bayley-Pinneau overestimation is significantly greater when bone age retardation exceeds two years [10,27]. In idiopathic advanced bone age, the "accelerated" Bayley-Pinneau tables are accurate for mild advancement but may overcorrect for severe advancement [30]. This graded relationship suggests that prediction methods should incorporate the magnitude of bone age deviation as a continuous variable rather than simply categorizing bone age as delayed, average, or advanced.

Age and Timing of Assessment

Prediction accuracy improves systematically as children approach skeletal maturity. In GHD, the correlation between predicted and actual adult height increased from $R^2 = 0.62$ at baseline to $R^2 = 0.94$ after three years of GH therapy [9]. In elite youth soccer players, Tanner-Whitehouse II and BoneXpert predictions converged toward accuracy as skeletal age approached maturity, while systematic overprediction occurred at younger skeletal ages (13.5-15 years) [49].

This age-dependent accuracy reflects the diminishing uncertainty about remaining growth potential as children progress through puberty. At younger ages, small errors in bone age assessment or assumptions about growth velocity compound over longer remaining growth periods. Conversely, near skeletal maturity, the range of possible adult heights narrows regardless of prediction method [22,64]. This has important clinical implications: predictions

made in early childhood should be interpreted with greater caution and updated regularly, while those made in later adolescence can be viewed with greater confidence.

Sex Differences

Sex-specific differences in prediction accuracy are consistently reported across studies. In short stature, Bayley-Pinneau is significantly more accurate for females than males [29,50]. In GHD, the magnitude of underestimation at baseline is greater in boys than girls [8], and boys show greater overestimation at the end of GH treatment [8]. These sex differences likely reflect fundamental differences in pubertal physiology: girls have a more predictable and compressed pubertal growth spurt, while boys show greater inter-individual variability in timing and intensity of peak growth velocity [35].

The Roche-Wainer-Thissen method shows the best performance in girls with GH deficiency [50], while in CDGP, RWT outperforms Bayley-Pinneau and BoneXpert across degrees of bone age delay [27]. These findings suggest that optimal method selection may differ by sex, with RWT offering particular advantages in female populations where its inclusion of weight may better capture body composition effects on growth.

Integrative and Novel Approaches

The most accurate prediction models identified in this review share a common characteristic: they integrate multiple sources of information beyond bone age alone. The inclusion of parental height consistently improves prediction accuracy, reducing RMSE by 0.25-0.5 cm in both healthy children [3] and those with ISS [16]. This incremental benefit reflects the strong heritability of adult height and the information content of mid-parental height for predicting a child's genetic growth potential.

Pubertal staging adds independent predictive value, with Wu et al. demonstrating that incorporating pubertal status into a simplified two-bone protocol reduced mean prediction error from 0.71 cm to 0.02 cm and increased the proportion of predictions within 3 cm from 66.9% to 73.5% [17]. This finding underscores the importance of considering maturational status beyond skeletal age alone, as pubertal stage captures information about endocrine status and growth velocity not fully reflected in bone age.

Novel skeletal maturity markers show promise for improving prediction accuracy. The distal femoral central peak height method, when combined with Greulich-Pyle bone age and chronological age, explained 92.4% of variance in timing to 90% of final height [76]. Similarly, bony greater trochanter height combined with Greulich-Pyle bone age and sex explained

approximately 85% of variance in adult height [79]. These quantitative markers offer advantages over traditional bone age assessment by providing continuous measurements rather than categorical ratings, potentially capturing more nuanced information about skeletal maturity.

The integration of multiple skeletal maturity systems into multivariate models represents a particularly promising approach. Munger et al. demonstrated that combining Fels, Greulich-Pyle, and Sanders methods with chronological age and height yielded the lowest prediction error in both boys and girls from the Brush Inquiry [20]. This finding suggests that different skeletal maturity systems capture complementary information about remaining growth potential, and their integration provides a more complete picture than any single method alone.

Radiomics and deep learning approaches represent the frontier of prediction methodology. The model by Xu et al. combining Greulich-Pyle bone age with knee radiomic scores and height at menarche achieved an R^2 of 0.733 and the lowest residual range (-5.68 to +6.44 cm) compared to traditional methods [19]. This approach leverages the rich information content of radiographic images beyond simple bone age ratings, extracting quantitative features of bone shape, density, and texture that may reflect growth dynamics. Similarly, the cascaded deep learning framework of Pei et al. achieved remarkable accuracy (MAE 1.75 cm) by simultaneously optimizing bone age assessment and height prediction in an end-to-end architecture [1].

Clinical Implications

The findings of this review have several important clinical implications. First, clinicians must recognize that prediction accuracy is not uniform across populations and conditions. The systematic biases identified—overestimation in CDGP, underestimation in untreated GHD and CPP—should inform clinical interpretation of predictions. Rather than treating predicted adult height as a precise estimate, clinicians should apply condition-specific correction factors based on the expected direction and magnitude of bias.

Second, method selection should be tailored to clinical context. For healthy children, modern automated methods such as BoneXpert offer excellent accuracy with minimal rater variability [3-5]. For GHD, the Roche-Wainer-Thissen method appears superior to alternatives [50]. For CDGP, condition-specific models incorporating the degree of bone age retardation outperform generic methods [10]. For CPP, predictions should be updated throughout

treatment, with end-of-treatment predictions carrying greater weight than those at diagnosis [14,15].

Third, the incorporation of additional clinical variables is essential for optimal accuracy. Parental height should be collected routinely and integrated into predictions [16,18]. Pubertal staging adds independent predictive value and should be performed at the time of bone age assessment [17]. For patients nearing skeletal maturity, predictions based on current height and remaining growth potential may be more accurate than bone age-based methods [22,64].

Fourth, automated systems offer important advantages in reducing rater variability and should be adopted where available [6,7]. However, clinicians must remain vigilant for condition-specific biases in automated predictions and should not assume that automation guarantees accuracy across all populations. In complex cases or those with extreme bone age deviations, manual assessment by experienced clinicians may provide complementary information [7,55].

Future Directions

Future research should prioritize large, multi-center studies directly comparing multiple prediction methods in diverse populations. The development and validation of condition-specific algorithms, such as those for CDGP [10] and ISS [18], should be extended to other disorders. The integration of novel biomarkers, including radiomic features [19] and genetic markers, holds promise for further accuracy improvements. Finally, the clinical impact of prediction accuracy on patient outcomes and treatment decisions should be evaluated prospectively.

CONCLUSION AND RECOMMENDATIONS

This systematic review of 80 studies provides a comprehensive synthesis of the evidence regarding bone age-based adult height prediction across diverse pediatric populations. The key findings can be summarized as follows:

1. In healthy, normally growing children, bone age-based adult height prediction achieves moderate accuracy with RMSE values of 2.5-3.5 cm for modern automated methods. Deep learning approaches have demonstrated mean absolute errors as low as 1.75 cm in specific populations.
2. Prediction accuracy varies systematically by clinical condition, with distinct and predictable patterns of bias:

- In growth hormone deficiency, substantial underestimation (4-8 cm) at baseline reverses to overestimation (2-4 cm) after treatment
 - In constitutional delay, consistent overestimation (1-5 cm) occurs, particularly when bone age retardation exceeds two years
 - In precocious puberty, marked underestimation (3-15 cm) at diagnosis improves after GnRHa therapy
 - In idiopathic short stature, RMSE ranges from 3.4-6.4 cm depending on method
3. Automated systems (BoneXpert) reduce inter-rater variability and achieve accuracy comparable or superior to manual methods in many populations, but show condition-specific limitations in severe bone age deviations.
 4. The relationship between bone age and chronological age critically moderates prediction accuracy, with delayed bone age associated with overestimation and advanced bone age with underestimation, in a dose-response manner.
 5. Integrative models incorporating parental height, pubertal staging, current anthropometrics, and novel radiomic markers consistently outperform single-method predictions, with multi-regression algorithms achieving RMSE reductions of 1-2 cm compared to traditional methods.
 6. Prediction accuracy improves as children approach skeletal maturity, emphasizing the importance of updating predictions throughout treatment and follow-up.

Recommendations for Clinical Practice

Based on these findings, we offer the following recommendations:

1. For healthy children undergoing routine growth assessment, automated bone age methods (BoneXpert) are preferred due to their excellent accuracy and elimination of inter-rater variability.
2. For children with growth hormone deficiency, the Roche-Wainer-Thissen method should be considered the preferred prediction approach, with recognition that predictions at treatment initiation will systematically underestimate final height.
3. For constitutional delay of growth and puberty, clinicians should anticipate that Bayley-Pinneau predictions will overestimate final height, particularly when bone age retardation exceeds two years. Condition-specific models or the Roche-Wainer-Thissen method are preferred alternatives.

4. For central precocious puberty, predictions should be obtained at diagnosis to establish baseline expectations but must be interpreted with awareness of systematic underestimation. Updated predictions at the end of GnRHa treatment provide more accurate estimates of final outcome.
5. Parental height should be routinely incorporated into height predictions, as it consistently improves accuracy across conditions and populations.
6. Pubertal staging should be performed concurrently with bone age assessment, as pubertal status provides independent predictive information beyond skeletal age alone.
7. Predictions made before puberty or in early childhood should be interpreted with greater caution and updated regularly, while those made in late adolescence can be viewed with greater confidence.
8. When using automated systems, clinicians should remain vigilant for condition-specific biases and should consider manual assessment as a complementary approach in cases with extreme bone age deviations.

Future Research Directions

The field of bone age-based adult height prediction stands at an inflection point, with traditional methods increasingly supplemented and potentially supplanted by automated systems and artificial intelligence. Future research should address:

1. Development and validation of condition-specific prediction algorithms for common growth disorders
2. Prospective, multi-center studies directly comparing all available methods in diverse populations
3. Integration of genetic markers and novel biomarkers into prediction models
4. Evaluation of the clinical impact of prediction accuracy on treatment decisions and patient outcomes
5. Development of population-specific standards and algorithms for non-Caucasian cohorts
6. Investigation of novel skeletal maturity markers from multiple anatomic sites
7. Validation of deep learning approaches across diverse clinical populations and settings
8. Establishment of standardized reporting guidelines for height prediction studies to facilitate meta-analysis

In conclusion, bone age assessment from hand-wrist radiographs provides a clinically useful foundation for adult height prediction, but the accuracy of predictions depends critically on matching the method to the clinical condition, incorporating additional patient characteristics, and interpreting results with awareness of systematic biases. The integration of automated systems, multi-modal data, and condition-specific algorithms holds promise for continued improvement in this essential clinical tool.

REFERENCES

1. Nihui Pei, Yijiang Zhuang, Zhe Su, et al (2025) Automated Bone Age Assessment and Adult Height Prediction from Pediatric Hand Radiographs via a Cascaded Deep Learning Framework. Journal of medical systems. <https://doi.org/10.1007/s10916-025-02306-9>
2. S. Ostojić (2013) Prediction of adult height by Tanner-Whitehouse method in young Caucasian male athletes. QJM : monthly journal of the Association of Physicians. <https://doi.org/10.1093/qjmed/hcs230>
3. H. H. Thodberg, O. Jenni, J. Caflisch, et al (2009) Prediction of adult height based on automated determination of bone age. Journal of Clinical Endocrinology and Metabolism. <https://doi.org/10.1210/jc.2009-1429>
4. David D Martin, Jan Schittenhelm, H. H. Thodberg (2016) Validation of adult height prediction based on automated bone age determination in the Paris Longitudinal Study of healthy children. Pediatric Radiology. <https://doi.org/10.1007/s00247-015-3468-8>
5. H. H. Thodberg, Julia Neuhofer, M. Ranke, et al (2010) Validation of Bone Age Methods by Their Ability to Predict Adult Height. Hormone Research in Paediatrics. <https://doi.org/10.1159/000313592>
6. D. Choukair, Annette Hückmann, Janna Mittnacht, et al (2021) Near Adult Heights and Predictions of Prospective Adult Heights using Automated and Conventional Greulich-Pyle Bone Age Determinations in children with chronic endocrine diseases. <https://doi.org/10.21203/RS.3.RS-383539/V1>
7. D. Choukair, Annette Hückmann, Janna Mittnacht, et al (2022) Near-Adult Heights and Adult Height Predictions Using Automated and Conventional Greulich-Pyle Bone Age Determinations in Children with Chronic Endocrine Diseases. Indian Journal of Pediatrics. <https://doi.org/10.1007/s12098-021-04009-8>

8. T. Reinehr, M. Carlsson, D. Chrysis, C. Camacho-Hübner (2020) Adult height prediction by bone age determination in children with isolated growth hormone deficiency. *Endocrine Connections*. <https://doi.org/10.1530/EC-20-0090>
9. Binh Duong, Khoa Vu, A. Bhangoo (2023) THU178 Predicted Adult Height Outcomes And Relation To Final Adult Height In Children With Isolated Idiopathic Growth Hormone Deficiency Treated With rhGH Therapy. *Journal of the Endocrine Society*. <https://doi.org/10.1210/jendso/bvad114.1429>
10. T. Reinehr, Elisa Hoffmann, J. Rothermel, et al (2019) A New Model of Adult Height Prediction Validated in Boys with Constitutional Delay of Growth and Puberty. *Hormone Research in Paediatrics*. <https://doi.org/10.1159/000499712>
11. F. Rohani, M. Alai, S. Moradi, Davoud Amirkashani (2018) Evaluation of near final height in boys with constitutional delay in growth and puberty. *Endocrine Connections*. <https://doi.org/10.1530/EC-18-0043>
12. K. Klein, K. Barnes, Janet Jones, et al (2001) Increased final height in precocious puberty after long-term treatment with LHRH agonists: the National Institutes of Health experience. *Journal of Clinical Endocrinology and Metabolism*. <https://doi.org/10.1210/JCEM.86.10.7915>
13. Hae-sang Lee, J. Yoon, K. Park, J. Hwang (2018) Increased final adult height by gonadotropin-releasing hormone agonist in girls with idiopathic central precocious puberty. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0201906>
14. Yanqin Ying, Jing Tang, Wei Chen, et al (2017) GnRH agonist treatment for idiopathic central precocious puberty can improve final adult height in Chinese girls. *OncoTarget*. <https://doi.org/10.18632/oncotarget.22568>
15. Mateus Cavarzan Lopes (2017) Comparative analysis of two models for adult height prediction and predicted age at menarche for girls with idiopathic central precocious puberty treated with GnRH Analog. *La Revue de medecine*. <https://doi.org/10.11606/ISSN.1679-9836.V96ISUPL1P3-3>
16. Marina Unrath, H. H. Thodberg, R. Schweizer, et al (2013) Automation of Bone Age Reading and a New Prediction Model Improve Adult Height Prediction in Children with Short Stature. *Hormone Research in Paediatrics*. <https://doi.org/10.1159/000345875>

17. Huahong Wu, Yaqin Zhang, Cheng-Dong Yu, et al (2025) Clinically Optimized Adult Height Prediction From Key Bone and Pubertal Stages: Prospective Validation to Adult Height. Journal of the Endocrine Society. <https://doi.org/10.1210/jendso/bvaf201>
18. W. Blum, M. Ranke, E. Keller, et al (2022) A Novel Method for Adult Height Prediction in Children With Idiopathic Short Stature Derived From a German-Dutch Cohort. Journal of the Endocrine Society. <https://doi.org/10.1210/jendso/bvac074>
19. Xueqiong Xu, YaoYao Chen, Yirou Wang, et al (2025) Final height prediction of girls at menarche: a combined model using left hand and wrist bone age, knee radiomic scores, and clinical characteristics. World Journal of Pediatrics. <https://doi.org/10.1007/s12519-025-01002-5>
20. Alana M Munger, Kristin E. Yu, Don T. Li, et al (2021) The More the Merrier: Integrating Multiple Models of Skeletal Maturity Improves the Accuracy of Growth Prediction. Journal of pediatric orthopedics. <https://doi.org/10.1097/BPO.0000000000001847>
21. Ashraf T Soliman, Nada M Alaaraj, Fawzia Alyafei, et al (2025) Disentangling predicted near-adult and measured adult height in pediatric growth assessment: The impact of bone age and prediction model accuracy. International Journal of Science and Research Archive. <https://doi.org/10.30574/ijsra.2025.16.1.2166>
22. B. U. Nwosu, Gabrielle Jasmin, Sadichchha Parajuli, et al (2021) Long-term GH Therapy Does Not Advance Skeletal Maturation in Children and Adolescents. Journal of the Endocrine Society. <https://doi.org/10.1210/jendso/bvab036>
23. Junghwan Suh, Jinkyong Heo, Su Jin Kim, et al (2023) Bone Age Estimation and Prediction of Final Adult Height Using Deep Learning. Yonsei medical journal. <https://doi.org/10.3349/ymj.2023.0244>
24. Shurong Huang, Z. Su, Shuangyi Liu, et al (2022) Combined assisted bone age assessment and adult height prediction methods in Chinese girls with early puberty: analysis of three artificial intelligence systems. Pediatric Radiology. <https://doi.org/10.1007/s00247-022-05569-3>
25. M. Yiğit, Elif Eviz, Ş. Hatun, G. Mutlu (2025) Automatic Bone Age Determination in Adult Height Prediction for Girls with Early Variants Puberty and Precocious Puberty. Journal of Clinical Research in Pediatric Endocrinology. <https://doi.org/10.4274/jcrpe.galenos.2025.2024-7-24>

26. L. Proos, T. Lönnerholm, B. Jonsson, T. Tuvemo (2006) Can Bone Age Determination Provide Criteria for Growth Hormone Treatment in Adopted Girls with Early Puberty? Upsala Journal of Medical Sciences. <https://doi.org/10.3109/2000-1967-018>
27. Gözde Akın Kağızmanlı, D. Kızılay, R. D. Sevim, et al (2025) What is the Most Effective Method for Predicting Adult Height in Boys with Constitutional Delay of Growth and Puberty? Journal of Clinical Research in Pediatric Endocrinology. <https://doi.org/10.4274/jcrpe.galenos.2025.2024-11-6>
28. J. O'Marr, Michael Amick, Kristin E. Yu, et al (2022) The Dimeglio Olecranon Method Is Reliable in Diverse, Contemporary Patients for Predicting Future Growth. Journal of pediatric orthopedics. <https://doi.org/10.1097/BPO.0000000000002328>
29. Seung woo Jeong, J. Cho, H. Jung, K. Shim (2018) Near final height in Korean children referred for evaluation of short stature: clinical utility and analytical validity of height prediction methods. Annals of Pediatric Endocrinology & Metabolism. <https://doi.org/10.6065/apem.2018.23.1.28>
30. L. Lazar, Y Lebenthal, S. Shalitin, M. Phillip (2010) Natural History of Idiopathic Advanced Bone Age Diagnosed in Childhood: Pattern of Growth and Puberty. Hormone Research in Paediatrics. <https://doi.org/10.1159/000319315>
31. L. Proos, T. Lönnerholm, Björn Jonsson, T. Tuvemo (2010) Can the TW3 Bone Age Determination Method Provide Additional Criteria for Growth Hormone Treatment in Adopted Girls with Early Puberty? Hormone Research in Paediatrics. <https://doi.org/10.1159/000271914>
32. Joseph Sullivan, S. Roberts, John Mckeown, et al (2023) Methods to predict the timing and status of biological maturation in male adolescent soccer players: A narrative systematic review. PLoS ONE. <https://doi.org/10.1371/journal.pone.0286768>
33. C. Bertaina, B. Stasiowska, A. Benso, S. Vannelli (2006) Is TW3 Height Prediction More Accurate than TW2? Preliminary Data. Hormone Research in Paediatrics. <https://doi.org/10.1159/000097439>
34. B. Cools, Raoul Rومان, L. Op De Beeck, M. D. Du Caju (2008) Boys with a Simple Delayed Puberty Reach Their Target Height. Hormone Research in Paediatrics. <https://doi.org/10.1159/000137663>

35. S. Poyrazoglu, H. Günöz, F. Darendeliler, et al (2005) Constitutional Delay of Growth and Puberty: From Presentation to Final Height. Journal of Pediatric Endocrinology & Metabolism (JPEM). <https://doi.org/10.1515/JPEM.2005.18.2.171>
36. A. Matias, Evgenia Muginshtein-Simkovitch, G. Twig, et al (2022) Comparison of Commonly Used Methods to Predict the Final Height in Constitutional Tall Stature. Journal of Clinical Research in Pediatric Endocrinology. <https://doi.org/10.4274/jcrpe.galenos.2022.2022-1-12>
37. R. Tzanetea, Nick Sakellaropoulos, Nick Hadjopoulos, et al (2000) Nutrition borne myolysis; an uncommon cause of myolysis in Europe. Upsala Journal of Medical Sciences. <https://doi.org/10.3109/2000-1967-180>
38. Michal Yackobovitch-Gavan, Ariel Tenenbaum, Moshe Phillip, et al (2025) Aromatase-inhibitors Treatment Alone or with GH Increases Final Height in Short-Statured Pubertal Boys - Real-World Data. Journal of Clinical Endocrinology and Metabolism. <https://doi.org/10.1210/clinem/dgaf224>
39. G. Beunen, R. Malina, Duarte Freitas, et al (2010) Cross-validation of the Beunen–Malina method to predict adult height. Annals of Human Biology. <https://doi.org/10.3109/03014460903393865>
40. D. Cho, Y. Choi, Hayun Oh, et al (2023) Accuracy of predicted adult height using the Greulich-Pyle method and artificial intelligence medical device. Clinical and experimental pediatrics. <https://doi.org/10.3345/cep.2022.01116>
41. F. Antoniazzi, G. Zamboni, F. Bertoldo, et al (2003) Bone mass at final height in precocious puberty after gonadotropin-releasing hormone agonist with and without calcium supplementation. Journal of Clinical Endocrinology and Metabolism. <https://doi.org/10.1210/JC.2002-021154>
42. F. J. Mejorado-Molano, María Luisa Sanz-Calvo, Ana Posada-Ayala, et al (2022) Adult Height in Girls With Idiopathic Premature Adrenarche: A Cohort Study and Design of a Predictive Model. Frontiers in Endocrinology. <https://doi.org/10.3389/fendo.2022.852422>
43. Silvia Ciancia, D. Klink, M. Craen, M. Cools (2023) Early Puberty Suppression and Gender-Affirming Hormones do not Alter Final Height in Transgender Adolescents. European Journal of Endocrinology. <https://doi.org/10.1093/ejendo/lvad125>

44. Shao-ke Chen, X. Fan, Qing-song Tang (2009) [Impact of gonadotropin-releasing hormone analogs treatment on final height in girls with central precocious puberty]. Zhongguo dang dai er ke za zhi = Chinese journal of contemporary pediatrics
45. R. Rapaport (2006) Adult Height Predictions for Constitutional Growth Delay, Growth Hormone Treatment for Idiopathic Short Stature and the FDA: Are They Related? Hormone Research in Paediatrics. <https://doi.org/10.1159/000092121>
46. Nisa Nur Turan, A. Çamaş, Burçin Çiçek, et al (2025) Comparison of Methods Used for Final Height Prediction in Central Precocious Puberty Patients. Journal of Clinical Research in Pediatric Endocrinology. <https://doi.org/10.4274/jcrpe.galenos.2025.2025-2-1>
47. G. Akın Kağızmanlı, R. Deveci Sevim, Ö. Besci, et al (2023) Which method is more effective in predicting adult height in pubertal girls treated with gonadotropin-releasing hormone agonist? HORMONES. <https://doi.org/10.1007/s42000-023-00466-2>
48. L. Lazar, A. Padoa, M. Phillip (2007) Growth pattern and final height after cessation of gonadotropin-suppressive therapy in girls with central sexual precocity. Journal of Clinical Endocrinology and Metabolism. <https://doi.org/10.1210/JC.2007-0321>
49. Lorenzo Lolli, Amanda Johnson, Mauricio Monaco, et al (2021) Tanner–Whitehouse and Modified Bayley–Pinneau Adult Height Predictions in Elite Youth Soccer Players from the Middle East. Medicine & Science in Sports & Exercise. <https://doi.org/10.1249/MSS.00000000000002740>
50. I. Erbas, Zeynep Ölmez, A. Paketçi, et al (2021) Comparison of the Effectiveness of Adult Height Prediction Methods in Children with Growth Hormone Deficiency. Endocrine Research. <https://doi.org/10.1080/07435800.2021.1916029>
51. Taja Knific, M. Lazarević, J. Žibert, et al (2022) Final adult height in children with central precocious puberty – a retrospective study. Frontiers in Endocrinology. <https://doi.org/10.3389/fendo.2022.1008474>
52. T. Oron, Y Lebenthal, L. de Vries, et al (2012) Interrelationship of extent of precocious adrenarche in appropriate for gestational age girls with clinical outcome. Jornal de Pediatria. <https://doi.org/10.1016/j.jpeds.2011.08.009>
53. Chin-Hui Tseng, Yann-Jinn Lee, Chi-Yu Huang, et al (2023) The effects of gonadotropin-releasing hormone agonist on final adult height among girls with early and fast puberty. Frontiers in Endocrinology. <https://doi.org/10.3389/fendo.2023.1271395>

54. A. Cho, S. Ko, Jae Hee Lee, Eun Young Kim (2020) Relationship between final adult height and birth weight after gonadotropin-releasing hormone agonist treatment in girls with central precocious puberty. *Annals of Pediatric Endocrinology & Metabolism*. <https://doi.org/10.6065/apem.2020.25.1.24>
55. Anni Erkkö, M. Hero, Tero Varimo, T. Raivio (2025) Performance of adult height prediction methods in 6 to 8-year-old girls with GnRH-dependent precocious puberty. *Journal of Pediatric Endocrinology & Metabolism (JPEM)*. <https://doi.org/10.1515/jpem-2024-0608>
56. C. Partsch, S. Heger, W. Sippell, German Decapeptyl Study Group, (2000) Treatment of Central Precocious Puberty: Lessons from a 15 Years Prospective Trial. *Journal of Pediatric Endocrinology & Metabolism (JPEM)*. <https://doi.org/10.1515/JPEM.2000.13.S1.747>
57. Kadri Yaman, Edip Unal, A. Beştaş, et al (2024) Retrospective evaluation of patients diagnosed with central precocious puberty who reached the final height. *Journal of Pediatric Endocrinology & Metabolism (JPEM)*. <https://doi.org/10.1515/jpem-2024-0124>
58. D. Mul, Silvano Bertelloni, Jean-Claude Carel, et al (2002) Effect of Gonadotropin-Releasing Hormone Agonist Treatment in Boys with Central Precocious Puberty: Final Height Results. *Hormone Research in Paediatrics*. <https://doi.org/10.1159/000063209>
59. G. Saggese, G. Federico, S. Barsanti, L. Fiore (2001) The effect of administering gonadotropin-releasing hormone agonist with recombinant-human growth hormone (GH) on the final height of girls with isolated GH deficiency: results from a controlled study. *Journal of Clinical Endocrinology and Metabolism*. <https://doi.org/10.1210/JCEM.86.5.7439>
60. A. Cho, S. Ko, Jae Hee Lee, Eun Young Kim (2021) Effects of gonadotropin-releasing hormone agonist treatment on final adult height in boys with idiopathic central precocious puberty. *Annals of Pediatric Endocrinology & Metabolism*. <https://doi.org/10.6065/apem.2040268.134>
61. Shruti A. Mondkar, V. Khadilkar, Sushil Yewale, et al (2023) Longitudinal assessment of auxological parameters, adult height outcome and its determinants in leuprolide-treated Indian girls with idiopathic central precocious puberty. *Journal of Pediatric Endocrinology & Metabolism (JPEM)*. <https://doi.org/10.1515/jpem-2023-0353>
62. M. Jung, Kyungchul Song, A. Kwon, et al (2014) Adult height in girls with central precocious puberty treated with gonadotropin-releasing hormone agonist with or without growth hormone. *Annals of Pediatric Endocrinology & Metabolism*. <https://doi.org/10.6065/apem.2014.19.4.214>

63. N. Akcan, Federica Verdecchia, Peter C. Hindmarsh, M. Dattani (2024) Contribution of Sex Steroids in Management of Tall Stature: Is it Effective or Not? Hormone Research in Paediatrics. <https://doi.org/10.1159/000542553>
64. B. U. Nwosu, Sadichchha Parajuli, Gabrielle Jasmin, Austin Lee (2021) Long-Term Growth Hormone Therapy Does Not Advance Skeletal Maturation in Children and Adolescents. Journal of the Endocrine Society. <https://doi.org/10.1210/jendso/bvab048.1385>
65. Hyun Ji Jang, M. Kwak, Young Mi Kim, et al (2023) Adult height in girls with central precocious puberty without gonadotropin-releasing hormone agonist treatment: a retrospective case-control study. Journal of Yeungnam Medical Science. <https://doi.org/10.12701/jyms.2023.00801>
66. Kyoung Won Cho, Youn Kyoung Kim, J. Yoo, et al (2026) Final adult height in male patients with central precocious puberty after gonadotropin-releasing hormone agonist treatment. Annals of Pediatric Endocrinology & Metabolism. <https://doi.org/10.6065/apem.2550064.032>
67. M. S. Kim, H. Koh, Gwang Yeon Lee, et al (2019) Comparing adult height gain and menarcheal age between girls with central precocious puberty treated with gonadotropin-releasing hormone agonist alone and those treated with combined growth hormone therapy. Annals of Pediatric Endocrinology & Metabolism. <https://doi.org/10.6065/apem.2019.24.2.116>
68. Eun Young Kim, K. Yi (2019) MON-241 Relationship between Final Height and Birth Weight after Gonadotropin-Releasing Hormone Agonist Treatment in Girls with Central Precocious Puberty. Journal of the Endocrine Society. <https://doi.org/10.1210/JS.2019-MON-241>
69. T. Ando, Masashi Nagao, Takayuki Miyamori, et al (2022) Comparing Accuracy of the Final Height Prediction Models for Elite Football Players and Developing a New Model. Juntendo Iji zasshi = Juntendo medical journal. <https://doi.org/10.14789/jmj.JMJ21-0053-OA>
70. M. Borges, Priscila de Melo Franciscan, Thamy Contursi Cambraia, et al (2015) Evaluation of central precocious puberty treatment with GnRH analogue at the Triangulo Mineiro Federal University (UFTM). Archives of Endocrinology and Metabolism. <https://doi.org/10.1590/2359-3997000000101>

71. T. Aeppli, E. Benyi, Henrik Wehtje, et al (2024) Safety and Efficacy of Bilateral Epiphysiodesis Surgery to Reduce Final Height in Extremely Tall Adolescents: A Follow-Up Study. *Hormone Research in Paediatrics*. <https://doi.org/10.1159/000538016>
72. 王晓鸥, 王德芬, 谢吉, et al (2005) Comparison of three methods on bone age evaluation in predicting final height of short children with various causes
73. K. Costa, I. Tomita, M. Barbieri, et al (2017) Comparison of methods for final height assessment in adolescents with a normal variant short stature. <https://doi.org/10.7322/JHGD.131259>
74. A. Gavrilova, Гаврилова Анна Евгеньевна, E. Nagaeva, et al (2017) Математические модели прогнозирования конечного роста и его коэффициента стандартного отклонения у детей с дефицитом гормона роста в российской популяции. <https://doi.org/10.14341/PROBL2017635282-290>
75. E. J. Sousa, Marilza Leal Nascimento, Genoir Simoni, et al (2008) Altura final em crianças com puberdade precoce central: a experiência de um serviço de referência
76. D. Knapik, James O. Sanders, Allison Gilmore, et al (2018) A quantitative method for the radiological assessment of skeletal maturity using the distal femur. *The Bone & Joint Journal*. <https://doi.org/10.1302/0301-620X.100B8.BJJ-2017-1489.R1>
77. M. Brenneis, Niklas Thewes, J. Holder, et al (2023) Validation of central peak height method for final adult height predictions on long leg radiographs. *Bone & Joint Open*. <https://doi.org/10.1302/2633-1462.410.BJO-2023-0105.R1>
78. Hae Woon Jung, Dohyun Chun, J. Choi, et al (2025) Comparison of adult height prediction using bone age and body composition for growth assessment in Korean children. *Scientific Reports*. <https://doi.org/10.1038/s41598-025-94685-9>
79. Grant B Nelson, D. Knapik, J. Janes, Raymond W. Liu (2020) Greater Trochanter Height: A Quantitative Predictor of Skeletal Maturity. *Journal of pediatric orthopedics*. <https://doi.org/10.1097/BPO.0000000000001724>
80. M. D. de Ridder, T. Stijnen, A. Hokken-Koelega (2008) Prediction model for adult height of small for gestational age children at the start of growth hormone treatment. *Journal of Clinical Endocrinology and Metabolism*. <https://doi.org/10.1210/JC.2007-1381>