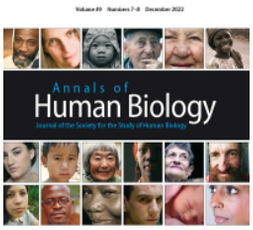
Taylor & Francis
Taylor & Francis Group



Annals of Human Biology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/iahb20

Deep phenotyping of pubertal development in Norwegian children: the Bergen Growth Study 2

Petur B. Juliusson, Ingvild S. Bruserud, Ninnie Helen Bakken Oehme, Andre Madsen, Ingvild H. Forthun, Melissa Balthasar, Karen Rosendahl, Kristin Viste, Astanand Jugessur, Lawrence M. Schell, Robert Bjerknes & Mathieu Roelants

To cite this article: Petur B. Juliusson, Ingvild S. Bruserud, Ninnie Helen Bakken Oehme, Andre Madsen, Ingvild H. Forthun, Melissa Balthasar, Karen Rosendahl, Kristin Viste, Astanand Jugessur, Lawrence M. Schell, Robert Bjerknes & Mathieu Roelants (2023) Deep phenotyping of pubertal development in Norwegian children: the Bergen Growth Study 2, Annals of Human Biology, 50:1, 226-235, DOI: 10.1080/03014460.2023.2174272

To link to this article: https://doi.org/10.1080/03014460.2023.2174272

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 26 Jun 2023.

Submit your article to this journal

Article views: 1152

View related articles

View Crossmark data

Citing articles: 4 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=iahb20

ANNALS OF HUMAN BIOLOGY
2023, VOL. 50, NO. 1, 226–235
https://doi.org/10.1080/03014460.2023.2174272

HUMAN NUTRITION AND LIFESTYLE

OPEN ACCESS Check for updates

Deep phenotyping of pubertal development in Norwegian children: the Bergen Growth Study 2

Petur B. Juliusson[a,b]*, Ingvild S. Bruserud[b,c], Ninnie Helen Bakken Oehme[b], Andre Madsen[d], Ingvild H. Forthun[a,b], Melissa Balthasar[a,e], Karen Rosendahl[f,g], Kristin Viste[d], Astanand Jugessur[h,i], Lawrence M. Schell[j,k], Robert Bjerknes[a,b] and Mathieu Roelants[l]

[a]Department of Clinical Science, University of Bergen, Bergen, Norway; [b]Children and Youth Clinic, Haukeland University Hospital, Bergen, Norway; [c]VID Specialized University, Bergen, Norway; [d]Hormone Laboratory, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway; [e]Department of Paediatrics and Adolescents, Stavanger University Hospital, Stavanger, Norway; [f]Department of Radiology, University Hospital of North Norway, Tromsø, Norway; [g]Department of Clinical Medicine, the Artic University of Norway, Tromsø, Norway; [h]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway; [i]Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway; [j]Department of Epidemiology and Biostatistics, University at Albany, Albany, New York, USA; [k]Center for the Elimination of Minority Health Disparities, University at Albany, Albany, New York, USA; [l]Department of Public Health and Primary Care, University of Leuven, Belgium

ABSTRACT

Background: The Bergen Growth Study 2 (BGS2) aims to characterise somatic and endocrine changes in healthy Norwegian children using a novel methodology.

Subjects and methods: A cross-sectional sample of 1285 children aged 6–16 years was examined in 2016 using novel objective ultrasound assessments of breast developmental stages and testicular volume in addition to the traditional Tanner pubertal stages. Blood samples allowed for measurements of pubertal hormones, endocrine disruptive chemicals, and genetic analyses.

Results: Ultrasound staging of breast development in girls showed a high degree of agreement within and between observers, and ultrasound measurement of testicular volume in boys also showed small intra- and interobserver differences. The median age was 10.4 years for Tanner B2 (pubertal onset) and 12.7 years for menarche. Norwegian boys reached a pubertal testicular volume at a mean age of 11.7 years. Continuous reference curves for testicular volume and sex hormones were constructed using the LMS method.

Conclusions: Ultrasound-based assessments of puberty provided novel references for breast developmental stages and enabled the measurement of testicular volume on a continuous scale. Endocrine *z*-scores allowed for an intuitive interpretation of changing hormonal levels during puberty on a quantitative scale, which, in turn, provides opportunities for further analysis of pubertal development using machine-learning approaches.

ARTICLE HISTORY
Received 3 November 2022
Revised 17 January 2023
Accepted 18 January 2023

KEYWORDS
Puberty; development; ultrasound; hormones; references

Introduction

Puberty is a period of dramatic somatic changes that leads to adult reproductive function. Alterations in the timing of puberty have been associated with a wide range of adverse health outcomes, placing a significant personal and economic burden on families and society (Day et al. 2015; Golub et al. 2008). In girls, for example, early pubertal timing is associated with an earlier sexual debut, a higher risk for sexual abuse and psychosocial maladjustment, and an increased lifetime susceptibility to reproductive cancers (Golub et al., 2008; Michaud et al., 2006). In boys, several studies have reported an association between early puberty and testicular cancer, whereas late pubertal onset has been associated with reduced semen quality (Jensen et al., 2016). Both early and late puberty in boys have been linked to psychosocial difficulties (Golub et al., 2008; Michaud et al., 2006). Furthermore, early puberty in both sexes has been associated with an increased risk of cardiovascular disease and type 2 diabetes in adulthood (Golub et al., 2008; Day et al., 2015).

The Tanner scales and the timing of menarche are routinely used to assess pubertal development. The British paediatrician James Tanner introduced his eponymous scoring system for assessing the development of secondary sex characteristics in the late 1960s, and it is still widely used today (Tanner & Whitehouse 1976). In girls, the Tanner scale includes five distinct stages of the breast (B1–B5) and pubic hair (PH1–PH5) development. In boys, it includes five stages of genital (G1–G5) and pubic hair (PH1–PH5) development. Testicular volume measured with a Prader orchidometer is also a part of the routine pubertal assessment in boys. Puberty onset is commonly defined by the Tanner stage B2

CONTACT Petur B. Juliusson petur.juliusson@uib.no Department of Clinical Science, University of Bergen, Bergen, Norway
*Additional affiliation: Department of Health Registry Research and Development, Norwegian Institute of Public Health, Bergen, Norway

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

of breast development in girls, and a testicular volume (TV) larger than 3 mL in boys. Assessment of the Tanner breast stage is prone to subjectivity, as it is mainly based on visual inspection. Overweight and obesity in the paediatric population have also led to increased uncertainty regarding the reliability of Tanner B staging since the presence of fat tissue could be misinterpreted as actual breast development, although this concern is more based on clinical experience than on scientific data (Euling et al., 2008). In boys, several studies have shown that the Prader orchidometer systematically overestimates small TVs, probably due to the inability of the instrument to differentiate the actual testicle from its surrounding tissues, e.g. epididymis, scrotal skin, and tunica vaginalis (Al Salim et al., 1995). Although the Tanner scale and the Prader orchidometer are easy to implement in a clinical setting, the need for a more objective classification system for pubertal development is justified. Furthermore, because such an assessment of breast development and testicular volume is based partially on palpation, it may be perceived as being psychologically invasive.

Age at menarche has decreased significantly since the nineteenth century (Parent et al., 2015). More recent data from Denmark and the United States indicate that the onset of puberty is continuing to decline (Aksglaede et al., 2009; Eckert-Lind et al., 2020; Herman-Giddens et al., 1997, 2001). Herman-Giddens et al. (1997) reported an earlier onset of breast development (thelarche) in girls, especially among Afro-Americans and Hispanics in the US. Aksglaede et al. (2009) compared pubertal development between 1991 and 2006 in Copenhagen and found similar evidence of early pubertal development, especially thelarche, that could not be explained by changes in hormonal levels or BMI. Trends towards accelerated maturation have also been reported in boys, but the magnitude of these changes appears to be less than in girls (Herman-Giddens et al., 2001).

Knowledge about puberty in Norwegian children before the Bergen Growth Study 2 (BGS2) was relatively limited. Brudevoll et al. (1979) compiled data on age at menarche collected in Oslo between 1840 and the 1970s, showing a decrease in mean age from 15.6 years in 1850 to 13.3 years in 1940. The mean age at menarche was 13.2 years in the first Bergen Growth Study conducted by our research group in 2003–2006 (BGS1), indicating only a small change in the space of over six decades (Juliusson et al., 2009). Finally, Per Erik Waaler conducted a small study on pubertal development in boys in the 1970s (Waaler et al., 1974). Because no normative data on puberty were available in Norway at the time, the growth reference charts for Norwegian children released in 2009 included age percentiles for the Tanner stages based on data collected from 1991 to 1993 in Copenhagen (Juliusson et al., 2009).

The main objective of BGS2 was to describe pubertal development in healthy Norwegian girls and boys using ultrasound as a novel method for objective assessment of breast maturation stages and testicular volume – compare with traditional Tanner stages and estimate normative clinical references. Other objectives were to construct endocrine references partitioned by stage of pubertal development and to investigate the association of timing of puberty with early growth, weight-related anthropometric measurements and body composition, endocrine-disrupting chemicals (EDCs), and genetic markers.

The Bergen Growth Study 2 (BGS2)

The BGS2 was conducted by our research group in 2016 and included a cross-sectional sample of 1285 children (735 girls)

Table 1. Data collected as a part of the Bergen Growth Study 2.

Children included in the reference sample	Girls (*n*)	Boys (*n*)
6–7 years	120	105
8–9 years	174	101
10–11 years	148	119
12–13 years	123	74
14–16 years	138	115
Total	703	514
Anthropometric measurements		
	Height	
	Weight	
	Waist circumference	
	Subscapular skinfold	
	Bio-impedance	
Assessment of pubertal development	Girls	
	Ultrasound of left breast[a]	
	Tanner breast (B) stage	
	Tanner pubic hair (PH) stage	
	Menarche (yes/no, (if yes, month/year)	
	Boys	
	Ultrasound of right testicle[b]	
	Tanner PH stage	
Blood samples	For measurement of pubertal hormones and endocrine disruptive chemicals, and for genetical analyses	
Parental questionnaire	General health and lifestyle factors, parental ethnicity, and educational level	
Routine measurements from primary health care	Length/height, weight, head circumference measurements from birth	

[a]If the right breast seemed larger or more mature on visual inspection, the right breast was also assessed.
[b]If deemed to be larger upon inspection, then the left testicle was also assessed.

aged 6–16 years (Table 1). Study participants were recruited from seven randomly selected public schools in the municipality of Bergen, which is the second-largest city in Norway with a population of approximately 290,000. All children in the selected schools were invited to participate, and all the participants who agreed to take part in the study were included in the analyses, regardless of their ethnic background. Examinations were conducted during school hours. The clinical assessment of puberty by the stage of breast development in girls and testicular volume in boys was done using ultrasound. The results of this novel approach were then compared to measurements conducted using the traditional Tanner scale and Prader orchidometry. Further "deep phenotyping" was achieved by compiling results with data on anthropometry (height and weight, subscapular skinfold, and waist circumference) and body composition (measured by bioelectrical impedance). In addition, blood samples were collected for endocrine profiling of gonadotropins, androgens, oestrogens and adipokines, and the quantification of endocrine disrupting chemicals (EDCs). Furthermore, DNA was extracted from the blood samples to enable genetic analyses of single-nucleotide polymorphisms (SNPs) (Table 1). The work on EDCs and SNPs is ongoing.

The BGS2 was approved by the Norwegian Regional Committee for Medical and Health Research Ethics West (REC-WEST 2015/128). Written informed consent was obtained from a parent or legal guardian of each participant in the study, as well as assent from the participants themselves. A movie voucher was given as an incentive to participate.

Pubertal development in Norwegian girls

Participants

The descriptive references for pubertal development are based on data collected between January and June 2016 ($n = 673$) and in February 2017 ($n = 57$ girls who participated in a test-retest study) (Bruserud et al. 2018, 2020). All seven schools were in more urbanised areas of the Bergen municipality. We conducted a test-retest study of the ultrasound assessment of puberty in one of the seven participating schools in February 2017 to determine the extent of intra- and inter-observer error when ultrasound was used to assess breast developmental stages in puberty (Bruserud et al., 2018). A random sample of 116 girls were invited to the study. Of these, 76 (65.5%) agreed to participate. However, due to time constraints, only 57 girls aged between 6.1 and 15.9 years were included.

All the girls attending the selected schools ($n = 1349$) were invited to participate in the main study, and parental consent was obtained for 673 girls (the participation rate was 49.4%). Of these, 27 with a chronic illness that could affect growth were excluded from the analyses (e.g. coeliac disease, diabetes type 1, heart disease, epilepsy, hypothyroidism, anorexia, cancer, or a kidney disorder). Based on the criteria for weight status defined by the International Obesity Task Force (IOTF) (Cole et al., 2000), 47 (7.2%) of the girls were considered underweight, 504 (78.1%) normal weight, 80 (12.4%) overweight, and 14 (2.2%) obese. Data on ethnicity was obtained from 466 girls, of whom 381 (81.2%) had parents of Norwegian/Nordic origin, 27 (5.8%) of European origin, and 51 (11.1%) of non-European origin (Bruserud et al., 2020). The highest parental educational level was no secondary educational degree in 16 (3.3%) girls, secondary education in 82 (17.1%), and higher education in 382 (79.6%). The proportion of parents with higher education was above the Norwegian mean.

Methods

The ultrasound-based scoring system used to characterise the maturation of glandular breast tissue (US B) was primarily based on a description by García et al. (2000) but was adapted to reflect relevant details and characteristic features highlighted by Bruni et al. (1990). For observer training and quality control, the ultrasound method was piloted in healthy girls. All ultrasound examinations during the first three days of data collection in BGS2 were performed jointly by the study nurse (I.S.B.) and an experienced paediatric radiologist (K.R) to evaluate observer agreement. These training and calibration sessions were used to standardise the ultrasound procedure, the result of which led to an adjustment in the ultrasound protocol, i.e. the addition of a distinct second prepubertal stage: US B0 (Figure 1) (Bruserud et al., 2018).

We performed all the ultrasound examinations on the girls in the supine position, with their arms rested on the side. The left breast was examined in all participants, in addition to the right breast when it appeared to be visually more mature (three girls only). The left breast was chosen over the right because this allowed the observer to rest her arm (to keep it steady) during the examination. The ultrasound device was a SonoSite Edge (Fujifilm SonoSite, USA) machine with a 15–6 MHz (5-cm) linear transducer. The probe was placed perpendicular to the skin and centred on the nipple to produce a sagittal standard section that was used for all measurements and staging procedures. Based on this standard section, the depth and diameter of the breast were measured, followed by morphological staging on a scale from US B0 to US B5 (Figure 1) (Bruserud et al., 2018).

For this study specifically, we used a 5-cm long linear transducer to measure the longest diameter of the fibro glandular area. For breast diameters larger than 5 cm but less than 10 cm, we combined the measurements from two scans in the same plane. The depth was measured from the nipple and then vertically down toward the pectoral muscle and/or the end of the glandular tissue. The degree of compression was kept to a minimum, as determined during the training and standardisation sessions. Direct measurements of the depth and diameter were chosen for the purpose of calculating the glandular volumes using the formula for a conical shape (volume $= (\pi/3) * \text{radius}^2 * \text{depth}$), as previously described (Calcaterra et al., 2009; Fugl et al., 2016).

For 166 girls, a preliminary US B stage was recorded during the examination, but a final decision regarding the stage was made afterwards based on saved standardised ultrasound images. The main reasons for this post-hoc assessment were time constraints and the need to avoid prolonged unnecessary exposure of the girls, with regard to the intimate nature

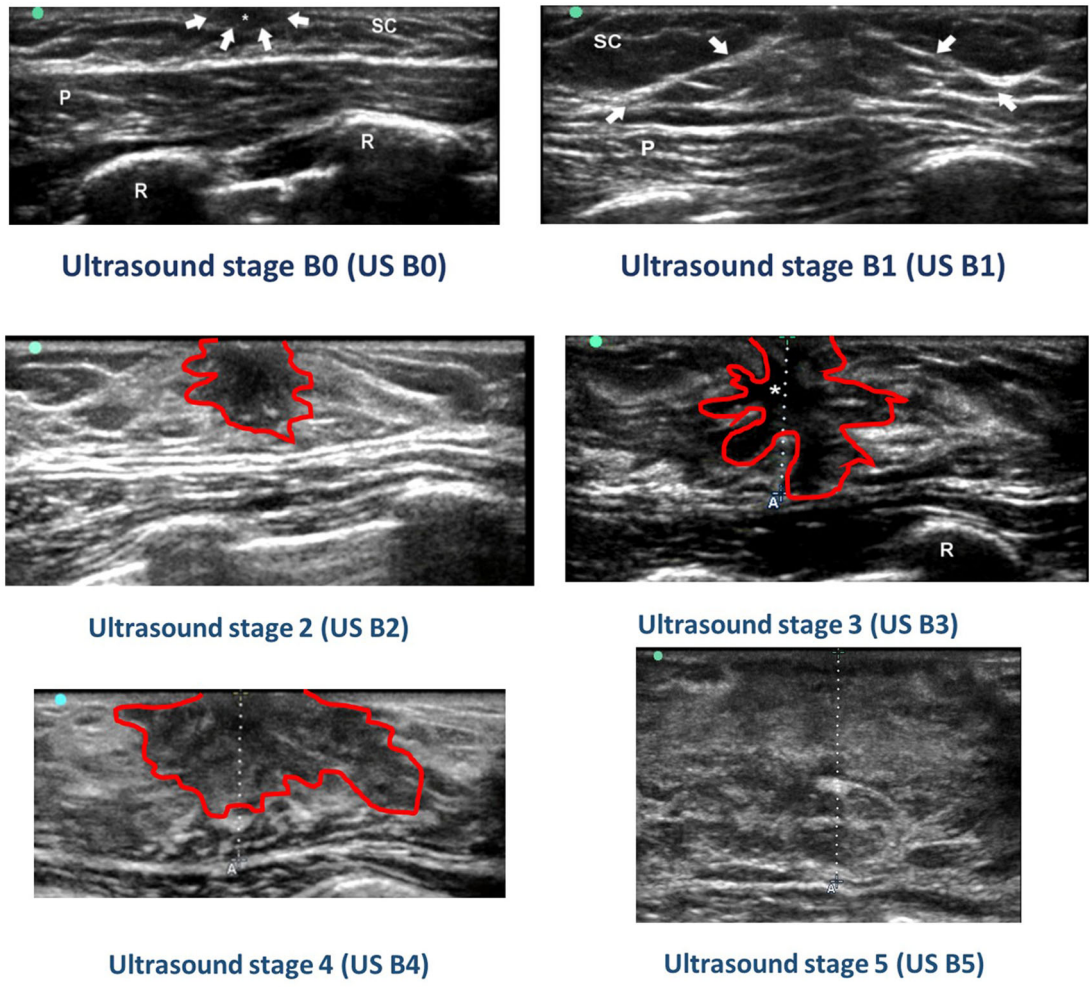


Figure 1. The ultrasound breast developmental stages. Ultrasound stage (US) B0 was defined as immature glandular breast tissue beneath the papilla, recognised as a small dark (hypoechoic) area. In US B1, the breast tissue is triangle-shaped and hyperechoic (light) compared to the surrounding tissue, but not compared to the pectoral muscle, with or without a small dark centre. In stage US B2, there is a hypoechoic centre that appears roundish. The surrounding breast tissue appears hyperechoic. In stage US B3, the hypoechoic centre is "spider-shaped", although the breast tissue appears hyperechoic. US B4 was defined when the hypoechoic centre, (also observed in US B2 and B3), had a rounder shape. In US B5, mature breast tissue was observed as a heterogeneous mass without any hypoechoic centre. One or more ribs (R) are observed in most images, and the pectoral muscle (P) is observed on all images (Bruni et al., 1990; Bruserud, 2018; García, 2000).

of the examination. Agreement between live scoring and scoring based on saved ultrasound images was estimated by re-examination of the images from 122 girls (>10 from each age year) by the same observer after a period of two years. Agreement between the original and rescored stage had a Cohen's kappa with linear weights of 0.76 (95% confidence interval [CI]: 0.698–0.812), which shows good agreement.

Age references for discrete stages of pubertal development (e.g. B2) were estimated from the cumulative incidence of pubertal status (e.g. B2 or higher vs not yet reached B2) by age using probit regression. The curves were estimated with a generalised linear model (GLM) when the distribution was Gaussian or with a nonparametric generalised additive model (GAM) otherwise.

Results

The intra-observer comparison of breast staging by the trained study nurse had a linear-weighted kappa coefficient of 0.84 (95% CI: 0.78–0.91) and a concordance of 70.2% (40/70; 95% CI: 56.4–81.2%). The inter-observer (nurse vs. paediatric radiologist) comparison had a kappa coefficient of 0.71 (0.62–0.80) and a concordance of 51.8% (29/56; 95% CI: 38.1–65.2%) when using all six stages of breast development. When the two prepubertal stages (US B0 and US B1) and the pubertal stages (US B2 and higher stages) were combined, we found a perfect agreement for one observer (i.e. 100% concordance) and a concordance of 96.4% (95% CI: 86.6%–99.4%) for the inter-observer assessments. For the measurement of depth and diameter of the mammary gland, the mean difference between measurements by the same observer was not significantly different from zero (one sample t-test, $p = .86$ and $p = .070$ for diameter and depth, respectively), indicating minimal systematic bias. For two different observers, the mean difference was not significantly different from zero for the diameter ($p = .86$), but the depth differed on average 0.1 cm ($p < .01$). However, the limits of agreement were wide for both depth (29% of the sample mean) and diameter (45.0% of the sample mean) (Bruserud et al., 2018). A constant variance across the range of measurements was observed in the Bland-Altman plots (data not shown).

The pubertal references were based on 696 girls for ultrasound breast staging (US B), 700 girls for Tanner B, 372 girls for Tanner PH, and 643 girls for menarche (Bruserud et al.,

Table 2. Age percentiles of pubertal developmental stages in girls and boys.

Pubertal marker	P_3	P_{50}	P_{97}
Girls			
Ultrasound Breast stage ($n = 696$)			
US B2	7.8	10.2	12.6
US B3	9.2	11.5	13.7
US B4	10.5	12.3	NA
US B5	11.5	15.9	NA
Tanner Breast (B) stage ($n = 700$)			
Tanner B2	8.2	10.4	12.6
Tanner B3	9.7	11.4	13.7
Tanner B4	10.8	12.5	16.4
Tanner B5	11.2	14.7	NA
Tanner Pubic Hair (PH) stage ($n = 372$)			
Tanner PH2	8.6	10.9	13.1
Tanner PH3	9.8	11.9	13.63
Tanner PH4	10.6	13.1	15.5
Tanner PH5	11.6	15.1	NA
Menarche ($n = 643$)	11.0	12.7	15.9
Boys			
US TV 2.7ml ($n = 514$)	9.7	11.7	13.7
Tanner PH2 ($n = 452$)	9.5	11.8	14.1
Tanner PH3	10.6	12.7	14.8
Tanner PH4	11.8	13.5	15.1
Tanner PH5	12.7	14.4	16.1

Age percentiles for breast developmental stage 2 to 5 based on ultrasound (US B) and Tanner (B) stages, menarche, Tanner stages of pubic hair (PH) development in girls. Age percentiles for ultrasound determined testicular volume corresponding to pubertal onset (US testicular volume (TV) of ≥2.7 ml) and for Tanner PH in boys. Median of the age (in years) at attainment of the specific stage in 6–16-year-old girls and boys in Norway.

2020). The median age at the onset of breast development was 10.2 years according to ultrasound staging (US B2) and 10.4 years according to Tanner staging (B2). The median age at Tanner PH2 was 10.9 years, while that of menarche was 12.7 years (Table 2). Pubertal onset occurred at a slightly earlier age (0.2 years) when using ultrasound staging compared to the Tanner method, while the opposite was found for the higher maturational stages (Tanner B4 and B5), where the age at transition with Tanner staging was ahead of the ultrasound assessment. The ultrasound and Tanner methods had a good overall level of agreement (kappa = 0.87 (95% CI: 0.85–0.89)) and were concordant in 551 of 695 (79.3%) assessments. When dichotomising the breast developmental stage into thelarche (B2 or higher) or no thelarche (US B0/B1 or Tanner B1), the agreement was very good (kappa = 0.94 (95% CI: 0.91–0.96)). The kappa coefficients were comparable in girls with average weight (kappa = 0.88 (95% CI: 0.86–0.91)) and overweight/obesity (kappa = 0.85 (95% CI: 0.79–0.90). The onset of all pubertal markers occurred earlier in girls with a non-Norwegian ancestry ($n = 92$) compared to girls of Norwegian ancestry ($n = 374$). A comparison with data from BGS1 demonstrated that age at menarche had significantly decreased from 13.3 (SD 1.7) years in 2006 to 13.1 (SD 1.2) years in 2016 ($p < .05$) in girls of Norwegian ancestry only. This difference remained statistically significant (odds ratio (OR): 2.0; 95% CI: 1.1–3.6; $p = .016$) when adjusted for the BMI *z*-score and parental educational level.

Pubertal development in Norwegian boys

Participants

Pubertal references for boys living in Norway were estimated regardless of ancestry. The curve and corresponding age references were based on 514 boys with a mean age of 11.0 (range, 6.1–16.4) years, of whom 57 participated in a test-retest study (Oehme et al., 2018, 2020). A random sample of 130 boys aged 6 to 16 years were invited to the test-retest study conducted in 2017, of whom 34 from the selected school and 24 from a sports club agreed to participate (Oehme et al., 2018). The mean age of the participants was 12.0 (range, 6.5–16.4) years. One boy with a history of undescended testis (cryptorchidism) was excluded, with the remaining 57 boys eligible for examination.

In the main reference study, all 1329 boys aged 6–16 years from seven selected schools were invited to participate (Oehme et al., 2020). Parental informed consent was obtained for 493 (37%) of the boys. On the day of the examination, two boys refused to give their assent, six did not attend, and eight were excluded as their medical history included a condition that could affect growth and development (e.g. coeliac disease, cancer, benign glioma, Down's syndrome, di George syndrome, ulcerative colitis, rheumatoid arthritis, and epilepsy with ongoing antiepileptic drug therapy). In addition, 20 boys were excluded due to scrotal pathology which was either known or newly discovered during the examination. Specifically, 4 had bilateral cryptorchidism; 11 unilateral cryptorchidism; 2 retractile testes (inguinal canal); 1 hydrocele; 1 operated retractile testis; and 1 microlithiasis. Combined with the 57 boys from the test–retest study, the reference sample thus included a total of 514 boys.

Based on data from 328 (71.8%) boys with known ancestry, 77.4% had both parents from Norway, 10.1% had one or both parents from another European country, and 12.5% had either one or two non-European parents, mostly from Asia, South America, or Africa. Of the 336 boys with information about parental education, the highest educational level attained by either parent was classified as: no secondary education (2.7%); secondary education (high school: 15.8%); and higher education (college or university degree: 81.6%. According to the IOTF BMI cut-off points (Cole et al., 2000), 7.7% of the participating boys were classified as underweight, 80.5% as normal weight, 11.8% as overweight, and 1.9% as obese.

Methods

All ultrasound examinations were performed by an experienced male radiographer, who was trained for this specific measurement protocol by an experienced paediatric radiologist (K.R.) before the study start. Further, the first 30 ultrasound examinations were supervised by K.R. A SonoSite Edge Ultrasound machine (Fujifilm SonoSite, USA) was used for examinations performed in the schools, and a SonoSite M-Turbo® HFL50 machine (Fujifilm SonoSite, USA) for examinations carried out on the boys from the local sports club; both devices were equipped with the same 15-6 MHz linear probe. With the boy in the supine position, the length (L), width (W), and depth (D) of the right testicle were measured according to a standardised protocol (Figure 2). The left testicle was also measured if deemed larger on visual inspection ($n = 3$), and the volume of the largest testicle was recorded. First, the ultrasound probe was placed in the mid-sagittal testicular plane, perpendicular to the skin surface. Second, the

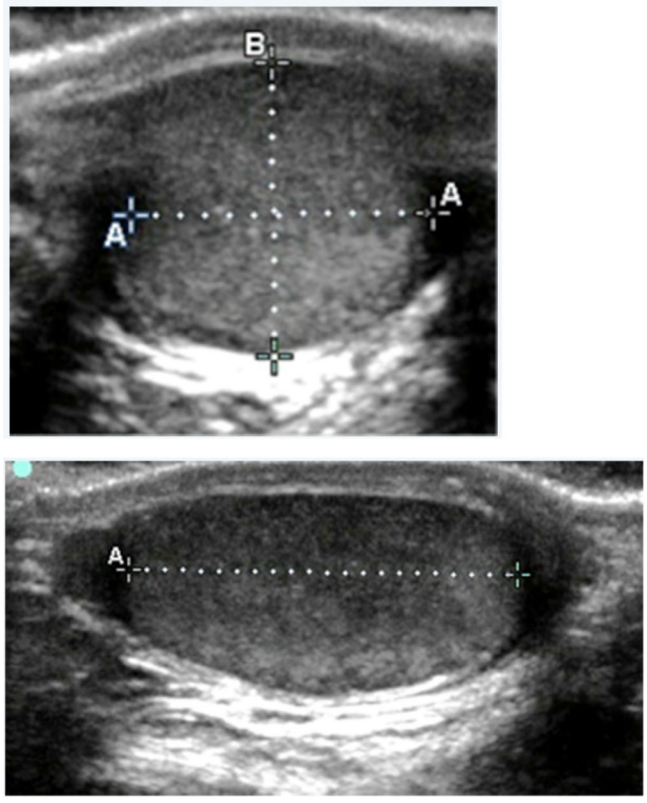


Figure 2. The ultrasound determined testicular volume (TV). Measurements of width and depth (above). Measurement of length (below).

examiner gently moved the ultrasound probe slightly back and forth until the largest diameter was recorded – namely the length. Third, the probe was rotated 90° and the width and depth measured in the mid-transverse plane (Figure 2). Testicular volume (TV) was then calculated later using the empirical Lambert formula ($TV = L \times W \times D \times 0.71$) (Lambert, 1951).

In the test–retest study, TV was measured twice by the main observer, with a time interval of at least 20 min between two measurements, during which at least three other participants were examined (Oehme et al., 2018). This was done to minimise the risk of recall of the first measurement. The participating boys were examined once by the second observer, who was blinded to the results obtained by the first observer. TV measurements of the right testicle were also performed using a Prader orchidometer by a paediatric endocrinologist (P.B.J.). The boys were examined in a standing position. The volume was that of the best matching bead of a Prader orchidometer as determined by comparative palpation. If the testicular size was perceived to be in between two consecutive beads, the mean volume of the beads was recorded.

References for the continuous ultrasound TV were estimated with the LMS method (Cole and Green, 1992). LMS allows the calculation of the distribution of a measurement at a given age and to convert any measurement into a *z*-score or a percentile. Age-references for discrete testicular volumes and stages of pubic hair (PH) were estimated from the cumulative incidence of pubertal status (e.g. PH2 or higher vs not yet reached PH2) by age using probit regression. The curves were estimated with a generalised linear model (GLM) when the distribution was Gaussian or with a nonparametric generalised additive model (GAM) otherwise.

Results

The comparison of TV measurements using ultrasound versus Prader orchidometer in the test-retest study revealed that the overall mean and standard deviation (SD) were highly comparable (Oehme et al., 2018). As the variation in measurement increased with mean TV, the differences between measurements, observers and methods were expressed as relative differences. Intra-observer agreement, which is the measure of *repeatability*, showed a mean difference (bias) of −2.2% ($p = .08$), indicating minimal systematic bias. The corresponding 95% limits of agreement (LOA) ranged from −20.3% to 15.9%, with a variability of 9.2% and a technical error of measurement (TEM) of 6.5%. Interobserver agreement, a measure of *reproducibility*, showed a small bias of 4.8% ($p = .052$), and the 95% LOA were somewhat wider, ranging from −35.7% to 45.3%, with a variability of 20.7% and a TEM of 14.6%.

Pubertal onset was defined as an ultrasound measured TV (USTV) of ≥2.7 mL in at least one testicle, which corresponds to a TV of ≥4 mL when measured using a Prader orchidometer. Tabulated values of L, M, and S for age are presented in the original publication, providing the information needed to calculate percentiles or to convert the measurements into *z*-scores. The mean age for attainment of a USTV of 2.7 mL was 11.7 (SD = 1.1) years, and the 3rd and 97th percentiles were 9.7 and 13.7 years, respectively. In addition, cumulative incidence curves for reaching selected discrete Prader orchidometer volumes are also presented (Table 2).

The pubertal reference for pubic hair development was based on 452 boys with a mean age of 10.9 (range, 6.1–16.3) years. The mean age (SD) of the development of pubic hair (pubarche; Tanner stage PH2) was 11.8 (1.2) years, with the 3rd and 97th percentiles of 9.5 and 14.1 years, respectively. Further, more boys achieved pubertal TV (≥2.7 mL) before pubarche (Tanner stage PH2), compared to boys who developed pubic hair as the first sign of puberty (14% versus 8.1%, respectively). Further, there was no indication that Norwegian boys entered puberty earlier than boys from comparable European countries.

References for pubertal hormones

Statistically robust hormone references in girls, in relation to chronological ages, ultrasound breast stages, and traditional Tanner B stages, were extrapolated from serum levels of oestrogens (estrone and oestradiol), gonadotropins (LH and FSH), and other biomarkers (SHBG and IGF1) (Madsen, Bruserud, et al., 2020). Although the breast stages determined by ultrasound and Tanner stages were highly concordant in terms of clinical stage occurrence and levels of oestrogens and gonadotrophins, ultrasound evaluations revealed nonpalpable glandular tissue in a subset of clinically prepubertal girls (Tanner B1 stratified into ultrasound stages B0 or B1). This ultrasound dichotomy was also corroborated

by distinct endocrine profiles, and the ultrasonographic presence of glandular tissue was associated with significantly increased levels of circulating oestradiol (Madsen, Bruserud, et al., 2020).

In boys, references were constructed for testosterone, LH, FSH, and SHBG (Madsen, Oehme, et al., 2020). Our finding that TV accounted for more variation in testosterone levels than age in pubertal boys emphasises the biological relevance of TV during puberty. Accordingly, we established an additional set of references for hormone levels in relation to TV. Reference intervals stratified by sex and age are essential for interpreting results from paediatric blood tests, and our findings suggest that the addition of TV as a covariate may provide more appropriate reference intervals for precision medicine. We also provided nonparametric continuous reference intervals in relation to age and USTV. Results showed that the studied hormones varied both with age and puberty progression, and that TV was significantly correlated with circulating testosterone levels in pubertal boys (Madsen, Oehme, et al., 2020).

With new blood sample data and methodological approach, we later remodelled the biomarker references using the established LMS growth curve algorithm (Madsen et al., 2022). The conventional practice of assigning age-adjusted percentile *z*-scores to paediatric patients is readily applicable to endocrine parameters as well and may be useful for clinical classifications. Clinically adoptable reference curves detailing the sex-specific and age-dependent levels of androgens, glucocorticoids and adrenal precursors (testosterone, androstenedione, 17-hydroxyprogesterone, 11-deoxycortisol and cortisol), oestrogens (estrone and oestradiol), gonadotropins (LH and FSH), adipokines (leptin and adiponectin) and other biomarkers (SHBG and IGF1) were recently published (Madsen et al., 2022) and received positive reviews in a subsequent editorial (Koskenniemi & Toppari, 2022). By leveraging the obtained biomarker *z*-scores as independent feature variables, we devised a proof-of-concept machine learning (ML) model that was successful at detecting obesity from blood sample data alone (Madsen et al., 2022). Configuring ML prediction models to classify certain paediatric conditions based on anthropometric and endocrine feature variables may provide clinical utility and improve patient care.

Age-stratified hormone reference intervals applicable for routine laboratory information systems were generated from BGS2, built on the framework proposed in a white paper by the Canadian Laboratory Initiative for Paediatric Reference Intervals (CALLIPER) (Adeli et al., 2017), conforming to the guidelines outlined by the Clinical Laboratory Standards Institute (CLSI) (CLSI, 2016). Steroid hormones were analysed by liquid chromatography coupled to mass spectrometry (LC-MS/MS) which is considered the gold standard.

Discussion

Our findings in BGS2 demonstrate that ultrasound-derived references are reliable for the assessment of pubertal development in girls and boys. In girls, the staging of breast development assessed by ultrasound was found robust with a high degree of agreement within and between observers. Contrary to expectations, the onset of breast development was detected earlier when using ultrasound compared to the Tanner method, and pubertal development thus started earlier than the current non-ultrasound based pubertal references imply. Norwegian girls do not seem to enter puberty significantly earlier than their peers in neighbouring countries. Our data show a decline in the age at menarche between BGS1 and BGS2, which remained significant after adjusting for BMI and ancestry. In boys, we found ultrasound to be a reliable method for assessing TV, with high intra-observer agreement and little bias, which makes it suitable for quantification and constructing continuous references. However, a slightly smaller interobserver agreement warrants a better standardisation of the measurements and training of the observers. As expected, we observed a slight tendency for the Prader orchidometer to overestimate smaller TVs than ultrasound. The age distribution for reaching pubertal milestones in boys was consistent with that observed in other Northern European countries.

Our report of a decrease in age at menarche between BGS1 and BGS2 was not explained by differences in age-adjusted BMI between the studies and was still significant when comparing the Norwegian girls only. The design and interpretation of data in the BGS1 and BGS2 were similar, and the samples were comparable. All girls were asked if they had experienced menarche at examination, avoiding the risk of recall bias. Earlier pubertal onset is a probable explanation for earlier menarche. However, no previous studies have investigated pubertal onset (i.e. the onset of breast development) in Norway. A similar trend of a significant decline in age at menarche when adjusting for BMI was also reported in the Netherlands from 1997 to 2009 (Talma et al., 2013).

Current paediatric endocrine references are often based on small sample sizes or clinical populations and may not be representative of the healthy paediatric population or attain common sample size requirements for reference ranges. Further, reference interval studies typically account for sex and chronological age only (Elmlinger et al., 2005). However, studies have shown complex changes in hypothalamic-pituitary-gonadal (HPG) axis hormones both during the first year of life and especially throughout adolescence (Busch et al., 2022; Konforte et al., 2013). This highlights the importance of stratifying reference intervals by age and pubertal stages. The CALLIPER project sets a new standard for presenting sex- and age-specific reference intervals, with their white paper article covering over 100 biomarkers for paediatric diseases, and presents reference intervals for HPG axis hormones partitioned based on self-reported Tanner stages (Adeli et al., 2017). In BGS2, we have addressed the variability by age and sex using continuous LMS based reference ranges, and further partitioned references by clinically determined Tanner stages, US-measured breast developmental stages, and USTV. The LMS methods offer the possibility of calculating age-adjusted *z*-scores that can be useful in clinical settings and research. Endocrine parameter *z*-scores can be added to the clinical report, giving intuitive information that, again, can be used in clinical decision-making.

Furthermore, we used endocrine *z*-scores along with other blood and anthropometric biomarkers, to perform machine learning analysis and demonstrate their usefulness in research (Madsen et al., 2022).

Another objective of our study was to investigate the relationship between weight-related anthropometric measurements and the onset of puberty. Previously, we observed a stronger association of low values of weight-related anthropometric measurements with later onset of puberty than high values with early puberty (Bratke et al., 2017; Oehme et al., 2021). The association of low weight-related anthropometric values in boys with later pubertal development received less attention in the literature, as previous studies often merged normal and underweight children in a single group (Busch et al., 2020). In girls, our findings so far were limited to menarche, but further analysis including the development of breast tissue is ongoing. The BGS2 includes few children with obesity or severe obesity because the sample is representative, and the prevalence of obesity in general is relatively low in Norwegian children. The BGS2 sample is therefore not well suited to address the issue of pubertal timing and severe excess weight. The increased prevalence of overweight and obesity with erroneous classification of adipose tissue as pubertal breast development when applying Tanner B staging has been proposed as a contributing factor to the observed advancement in age at onset of breast development (Euling et al., 2008). We, therefore, hypothesised that the use of ultrasound could detect if Tanner B staging erroneously classified girls who were overweight or obese as being pubertal due to having more subcutaneous fat tissue. We did not find any evidence for this in our current study, but again, our study may have been limited due to the low number of girls with overweight/obesity.

Finally, we aimed to investigate the association between early postnatal and childhood growth and later pubertal development, which is also ongoing. Growth is routinely monitored in primary health care from birth through primary school, and early growth data from birth to six years of age have been obtained from most of the participating children.

The analysis of the EDCs in BGS2 has been funded and is currently being carried out. EDCs are exogenous chemicals or mixtures of chemicals that interfere with hormonal action. EDCs are either synthetic (the majority) or naturally occurring chemicals that have been demonstrated to exhibit endocrine properties including oestrogenic, antiandrogenic and thyroid actions (Diamanti-Kandarakis et al., 2009). Our modern society is significantly more exposed to synthetic EDCs than in any known period of human history. Implications of this exposure have yet to be determined. Naturally occurring chemicals that have been linked to earlier menarche or breast development include phytoestrogens, lavender oil, tea tree oil and fennel (Fisher & Eugster, 2014). Synthetic EDCs which have been linked to pubertal timing include, among others, polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs), organochlorine pesticides 3 (DDE, HCB) and perfluoroalkyl substances (PFASs) (Rappazzo et al., 2017; Schell & Gallo, 2010; Schell et al., 2014). All these synthetic EDCs can cross the placenta and are present in breast milk. An important benefit of the ultrasound-based references is that they are well-tailored to an assessment of EDCs and their subsequent influence on pubertal development. Investigations into this question often involve small samples because of the difficulty in engaging children at a sensitive age into a study of a sensitive character (pubertal development). This often limits sample sizes, while the cost of measuring a wide panel of EDCs further limits sample sizes. In studies with small samples, it is more important to obtain the most accurate assessment of maturation milestones and reduce variation due to inaccurate measurements. BSG2 is characterised by a combination of factors that add precision to the measures of pubertal development: a relatively large sample, a large panel of EDCs assessed, and the most accurate assessments of maturational stages.

Genome-wide association study (GWAS) data on BGS2 samples have been generated, and polygenic risk scores constructed using the index variants are ready to be used in a multitude of genetic association analyses. The timing of puberty is a highly polygenic trait, and more than 400 significant SNPs have been identified through large GWASs (Day et al., 2017). As with virtually all other complex traits, most of the identified variants confer small effects individually, but in aggregate they explain approximately 7.5% of the total variance in the timing of menarche, corresponding to approximately 25% of the estimated heritability (Day et al., 2017). Heritability was previously demonstrated, as age at menarche of mothers is associated with the timing of puberty in both their daughters and sons (Sorensen et al., 2018). Furthermore, a GWAS of age at voice breaking (commonly used as proxy to assess male puberty) indicated that many of the abovementioned puberty-associated genetic variants are shared among boys and girls (Day, Bulik-Sullivan, et al., 2015). This is expected because several of the same genes are needed to restart the HPG axis and drive pubertal development in both sexes.

Many of the participating children in this study also took part in the Norwegian Mother, Father, and Child Cohort Study (MoBa), which included the collection of blood samples from mothers and fathers during week 18 of gestation, and from newborns after delivery (cord blood). Together with the data generated on pubertal development, we plan to perform a family-based analysis using the genetic material collected in BGS2 (during puberty) and MoBa (gestation). We will use an integrative omics approach where data from GWAS and epigenome-wide association studies (EWAS) are combined to elucidate the genetic and epigenetic underpinnings of puberty. Parallel to this, we will also focus on mapping the impact of environmental factors on puberty timing. EDCs will also be analysed in the MoBa samples, giving us the possibility to compare levels at birth with those in later childhood, and the relationship between pubertal development and body composition.

As the study design of the BGS2 is cross-sectional, it allowed us to estimate the distribution of ages when children reach certain pubertal milestones. However, longitudinal aspects of pubertal development, such as the time needed to progress from one stage to the next or individual variation in

the sequence of events during puberty, cannot be estimated. Although we have limited data on the non-participants in the BGS2, key characteristics such as ethnicity and the prevalence of overweight and obesity of the included children were comparable with the childhood population of Bergen (Bruserud et al., 2018). There are several potential advantages of using ultrasound when assessing pubertal development. The approach might be perceived as being less intrusive/invasive for the participant due to the more technical nature of the measurement. Indeed, participants reported less intrusion with the ultrasound approach than palpation (data not published). Furthermore, ultrasound provides a potentially more objective approach, with the possibility to avoid misclassification of adipose tissue as pubertal development, to detect scrotal pathology and to discard surrounding tissue from the testicular measurement. Finally, ultrasound images can be saved for later comparisons and follow-up.

To conclude, BGS2 is the first pubertal reference study conducted in Norway. The knowledge and definition of normal puberty timing in contemporary girls and boys are crucial to assess normal versus aberrant pubertal development. Identifying alterations in the timing of growth and sexual maturation at a population level is also important, particularly in relation to potential later adverse health outcomes. This study has demonstrated that ultrasound is a suitable method for evaluating pubertal development. In girls, breast developmental stages were found to be robust, and we have defined a new prepubertal stage with corresponding endocrine profile. In boys, a continuous testicular reference has been constructed, making calculation of *z*-scores possible. Hormonal reference curves and the use of age-dependent *z*-scores open for a more intuitive presentation of the hormones involved in the pubertal development and can be used for further analysis using machine learning approaches. Investigating the impact of EDCs and genetic variation on the timing of pubertal development by leveraging the wide scope of data collected in BGS2 offer unprecedented opportunities for such analyses.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The study was funded by the Western Norway Regional Health Authority (grants no. 911975, 912131 and 91221), and also internal funding from Laboratory Medicine and Pathology, Haukeland University Hospital.

ORCID

Petur B. Juliusson http://orcid.org/0000-0002-7064-1407

Data availability statement

Restrictions apply to the availability of data generated or analysed during this study to preserve patient confidentiality or because they were used under licence. The corresponding author can be contacted regarding these restrictions and any conditions under which access to some of the data may be provided to other investigators.

References

Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. 2017. The Canadian laboratory initiative on pediatric reference intervals: a CALIPER white paper. Crit Rev Clin Lab Sci. 54(6):358–413.

Aksglaede L, Sorensen K, Petersen JH, Skakkebaek NE, Juul A. 2009. Recent decline in age at breast development: the Copenhagen Puberty Study. Pediatrics. 123(5):e932–e939.

Al Salim A, Murchison PJ, Rana A, Elton RA, Hargreave TB. 1995. Evaluation of testicular volume by three orchidometers compared with ultrasonographic measurements. Br J Urol. 76(5):632–635.

Bratke H, Bruserud IS, Brannsether B, Aßmus J, Bjerknes R, Roelants M, Júlíusson PB. 2017. Timing of menarche in Norwegian girls: associations with body mass index, waist circumference and skinfold thickness. BMC Pediatr. 17(1):138.

Brudevoll JE, Liestøl K, Walløe L. 1979. Menarcheal age in Oslo during the last 140 years. Ann Hum Biol. 6(5):407–416.

Bruni V, Dei M, Deligeoroglou E, Innocenti P, Pandimiglio AM, Magini A, Bassi F. 1990. Breast development in adolescent girls. Adolescent and Pediatric Gynecology. 3(4):201–205.

Bruserud IS, Roelants M, Oehme NHB, Eide GE, Bjerknes R, Rosendahl K, Juliusson PB. 2018. Ultrasound assessment of pubertal breast development in girls: intra- and interobserver agreement. Pediatr Radiol. 48(11):1576–1583.

Bruserud IS, Roelants M, Oehme NHB, Madsen A, Eide GE, Bjerknes R, Rosendahl K, Juliusson PB. 2020. References for ultrasound staging of breast maturation, tanner breast staging, pubic hair, and menarche in Norwegian girls. J Clin Endocrinol Metab. 105(5):1599–1607.

Busch AS, Højgaard B, Hagen CP, Teilmann G. 2020. Obesity is associated with earlier pubertal onset in boys. J Clin Endocrinol Metab. 105(4): dgz222.

Busch AS, Ljubicic ML, Upners EN, Fischer MB, Raket LL, Frederiksen H, Albrethsen J, Johannsen TH, Hagen CP, Juul A. 2022. Dynamic changes of reproductive hormones in male minipuberty: temporal dissociation of leydig and sertoli cell activity. J Clin Endocrinol Metab. 107(6):1560–1568.

Calcaterra V, Sampaolo P, Klersy C, Larizza D, Alfei A, Brizzi V, Beneventi F, Cisternino M. 2009. Utility of breast ultrasonography in the diagnostic work-up of precocious puberty and proposal of a prognostic index for identifying girls with rapidly progressive central precocious puberty. Ultrasound Obstet Gynecol. 33(1):85–91.

CLSI. 2016. Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline—3rd edition. CLSI document EP28-A3c. Wayne (PA): Clinical and Laboratory Standards Institute.

Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. 2000. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ. 320(7244):1240–1243.

Cole TJ, Green PJ. 1992. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med. 11(10):1305–1319.

Day FR, Bulik-Sullivan B, Hinds DA, Finucane HK, Murabito JM, Tung JY, Ong KK, Perry JR. 2015. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. Nat Commun. 6:8842.

Day FR, Elks CE, Murray A, Ong KK, Perry JR. 2015. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. Sci Rep. 5:11208.

Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, Ruth KS, Whalen S, Sarkar AK, Albrecht E, et al. 2017. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. Nat Genet. 49(6): 834–841.

Diamanti-Kandarakis E, Bourguignon JP, Giudice LC, Hauser R, Prins GS, Soto AM, Zoeller RT, Gore AC. 2009. Endocrine-disrupting chemicals: an Endocrine Society scientific statement. Endocr Rev. 30(4):293–342.

Eckert-Lind C, Busch AS, Petersen JH, Biro FM, Butler G, Brauner EV, Juul A. 2020. Worldwide secular trends in age at pubertal onset assessed

by breast development among girls: a systematic review and meta-analysis. JAMA Pediatr. 174(4):e195881.
Elmlinger MW, Kühnel W, Wormstall H, Döller PC. 2005. Reference intervals for testosterone, androstenedione and SHBG levels in healthy females and males from birth until old age. Clin Lab. 51(11–12):625–632.
Euling SY, Herman-Giddens ME, Lee PA, Selevan SG, Juul A, Sørensen TI, Dunkel L, Himes JH, Teilmann G, Swan SH. 2008. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. Pediatrics. 121(Suppl 3):S172–S191.
Fisher MM, Eugster EA. 2014. What is in our environment that effects puberty? Reprod Toxicol. 44:7–14.
Fugl L, Hagen CP, Mieritz MG, Tinggaard J, Fallentin E, Main KM, Juul A. 2016. Glandular breast tissue volume by magnetic resonance imaging in 100 healthy peripubertal girls: evaluation of clinical Tanner staging. Pediatr Res. 80(4):526–530.
García CJ, Espinoza A, Dinamarca V, Navarro O, Daneman A, García H, Cattani A. 2000. Breast US in children and adolescents. Radiographics. 20(6):1605–1612.
Golub MS, Collman GW, Foster PM, Kimmel CA, Rajpert-De Meyts E, Reiter EO, Sharpe RM, Skakkebaek NE, Toppari J. 2008. Public health implications of altered puberty timing. Pediatrics. 121(Suppl 3):S218–S230.
Herman-Giddens ME, Slora EJ, Wasserman RC, Bourdony CJ, Bhapkar MV, Koch GG, Hasemeier CM. 1997. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in office settings network. Pediatrics. 99(4):505–512.
Herman-Giddens ME, Wang L, Koch G. 2001. Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988-1994. Arch Pediatr Adolesc Med. 155(9):1022–1028.
Jensen TK, Finne KF, Skakkebæk NE, Andersson AM, Olesen IA, Joensen UN, Bang AK, Nordkap L, Priskorn L, Krause M, et al. 2016. Self-reported onset of puberty and subsequent semen quality and reproductive hormones in healthy young men. Hum Reprod. 31(8):1886–1894.
Juliusson PB, Roelants M, Eide GE, Moster D, Juul A, Hauspie R, Waaler PE, Bjerknes R. 2009. Growth references for Norwegian children. Tidsskr Nor Laegeforen. 129(4):281–286.
Konforte D, Shea JL, Kyriakopoulou L, Colantonio D, Cohen AH, Shaw J, Bailey D, Chan MK, Armbruster D, Adeli K. 2013. Complex biological pattern of fertility hormones in children and adolescents: a study of healthy children from the CALIPER cohort and establishment of pediatric reference intervals. Clin Chem. 59(8):1215–1227.
Koskenniemi JJ, Toppari J. 2022. The beauty of age-dependent standardization in pediatric endocrine research and practice. J Clin Endocrinol Metab. 107(8):e3528–e3529.
Lambert B. 1951. The frequency of mumps and of mumps orchitis and the consequences for sexuality and fertility. Acta Genet Stat Med. 2(Suppl 1):1–166.
Madsen A, Almås B, Bruserud IS, Oehme NHB, Nielsen CS, Roelants M, Hundhausen T, Ljubicic ML, Bjerknes R, Mellgren G, et al. 2022. Reference curves for pediatric endocrinology: leveraging biomarker *Z*-scores for clinical classifications. J Clin Endocrinol Metab. 107(7): 2004–2015.
Madsen A, Bruserud IS, Bertelsen BE, Roelants M, Oehme NHB, Viste K, Bjerknes R, Almås B, Rosendahl K, Mellgren G, et al. 2020. Hormone references for ultrasound breast staging and endocrine profiling to detect female onset of puberty. J Clin Endocrinol Metab. 105(12): e4886–e4895.
Madsen A, Oehme NB, Roelants M, Bruserud IS, Eide GE, Viste K, Bjerknes R, Almas B, Rosendahl K, Sagen JV, et al. 2020. Testicular ultrasound to stratify hormone references in a cross-sectional Norwegian study of male puberty. J Clin Endocrinol Metab. 105(6):dgz094.
Michaud PA, Suris JC, Deppen A. 2006. Gender-related psychological and behavioural correlates of pubertal timing in a national sample of Swiss adolescents. Mol Cell Endocrinol. 254–255:172–178.
Oehme NHB, Roelants M, Bruserud IS, Eide GE, Bjerknes R, Rosendahl K, Juliusson PB. 2018. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys – intra- and interobserver agreement and comparison with Prader orchidometry. Pediatr Radiol. 48(12):1771–1778.
Oehme NHB, Roelants M, Bruserud IS, Madsen A, Bjerknes R, Rosendahl K, Juliusson PB. 2021. Low BMI, but not high BMI, influences the timing of puberty in boys. Andrology. 9(3):837–845.
Oehme NHB, Roelants M, Særvold Bruserud I, Madsen A, Eide GE, Bjerknes R, Rosendahl K, Juliusson PB. 2020. Reference data for testicular volume measured with ultrasound and pubic hair in Norwegian boys are comparable with Northern European populations. Acta Paediatr. 109(8):1612–1619.
Parent AS, Franssen D, Fudvoye J, Gérard A, Bourguignon JP. 2015. Developmental variations in environmental influences including endocrine disruptors on pubertal timing and neuroendocrine control: revision of human observations and mechanistic insight from rodents. Front Neuroendocrinol. 38:12–36.
Rappazzo KM, Coffman E, Hines EP. 2017. Exposure to perfluorinated alkyl substances and health outcomes in children: a systematic review of the epidemiologic literature. Int J Environ Res Public Health. 14(7):691.
Schell LM, Gallo MV. 2010. Relationships of putative endocrine disruptors to human sexual maturation and thyroid activity in youth. Physiol Behav. 99(2):246–253.
Schell LM, Gallo MV, Deane GD, Nelder KR, DeCaprio AP, Jacobs A. 2014. Relationships of polychlorinated biphenyls and dichlorodiphenyldichloroethylene (p,p′-DDE) with testosterone levels in adolescent males. Environ Health Perspect. 122(3):304–309.
Sorensen S, Brix N, Ernst A, Lauridsen LLB, Ramlau-Hansen CH. 2018. Maternal age at menarche and pubertal development in sons and daughters: a Nationwide Cohort Study. Hum Reprod. 33(11):2043–2050.
Talma H, Schönbeck Y, van Dommelen P, Bakker B, van Buuren S, Hirasing RA. 2013. Trends in menarcheal age between 1955 and 2009 in the Netherlands. PLOS One. 8(4):e60056.
Tanner JM, Whitehouse RH. 1976. Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. Arch Dis Child. 51(3):170–179.
Waaler PE, Thorsen T, Stea KF, Aarskog D. 1974. Studies in normal male puberty. Acta Paediatr Scand Suppl. 63(249):1–36.