Reference data for testicular volume measured with ultrasound and pubic hair in Norwegian boys are comparable with Northern European populations

Short running title: Pubertal development in Norwegian boys

*Ninnie Helen Bakken Oehme[1,2], MD, Mathieu Roelants[3], PhD, Ingvild Særvold Bruserud[1,2], MSN, Andre Madsen[1,4], PhD, Geir Egil Eide[5,6], PhD, Robert Bjerknes[1,2], MD, PhD, Karen Rosendahl[7,8], MD, PhD, Petur B. Juliusson[1,2,9], MD, PhD

1 Department of Clinical Science, University of Bergen, Bergen, Norway

2 Department of Pediatrics, Haukeland University Hospital, Bergen, Norway

3 Environment and Health, Department of Public Health and Primary Care, KU Leuven – University of Leuven, Belgium

4 The Hormone Laboratory, Haukeland University Hospital, Bergen, Norway

5 Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway

6 Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway

7 Department of Radiology, Haukeland University Hospital, Bergen, Norway

8 Department of Clinical Medicine, University of Bergen, Bergen, Norway

9 Department of Health Registries, Norwegian Institute of Public Health

Corresponding author:

Ninnie Oehme

Department of Clinical Science, University of Bergen

N-5020 Bergen, Norway

Telephone (mobile): +47 90540556

Email: ninnie.oehme@uib.no

Abstract

Aim

To estimate references for testicular volume measured with ultrasound and Tanner stages of pubic hair in Norwegian boys, and to compare the timing of puberty with data from similar populations.

Methods

Testicular volume was derived from ultrasound measurements of testicular volume in a cross-sectional study of 514 healthy boys. A continuous testicular volume for age reference curve was estimated with the LMS method. Tanner stages for pubic hair were clinically assessed in 452 boys. Age references for pubertal milestones were estimated with probit regression.

Results

Puberty onset, defined by an ultrasound testicular volume of 2.7 ml, equivalent to an orchidometer volume of 4 ml, occurred at a mean (SD) age of 11.7 (1.1) years. The reference range was 9.7 (3^{rd}) to 13.7 years (97^{th} percentile). Pubic hair (Tanner stage 2) appeared on average at 11.8 (1.2) years with a corresponding reference range of 9.5 to 14.1 years.

Conclusion

The references for testicular volume measured with ultrasound are continuous in age, and allow for the quantification of pubertal development. The age distribution of reaching pubertal milestones was comparable with data from other Northern European countries.

Keywords: Puberty, Secular trend, Testis, Ultrasonography

Key Notes

The testicular volume is usually determined with a Prader orchidometer, however ultrasound examinations is possibly a better alternative.

We estimated up-to-date references for testicular volume, measured with ultrasound, and pubic hair in healthy Norwegian boys.

Implementation of ultrasound for assessing testicular volume implies a transition towards an objective measurement on a continuous scale, which allows to detect smaller changes in the testicular volume and to quantify pubertal development.

Introduction

During the past two decades, several authors have demonstrated renewed trends towards earlier puberty in girls, after a relatively stable period of almost 60 years (1, 2). Although results in boys are more equivocal (3, 4), some studies have suggested similar trends (5, 6). Overweight and obesity have been proposed as possible drivers for this renewed trend (6-8), as well as exposure to endocrine disrupting chemicals (9).

Population based studies of puberty are more challenging in boys than in girls, due to the lack of an easily measured, yet reliable pubertal marker like menarche (4). Testicular examination with a Prader orchidometer is useful and widely used in clinical practice, but it is regarded as impractical for population studies (10). In boys, attainment of a testicular volume (TV) of ≥ 4ml when measured with a Prader orchidometer is considered the best indicator for the onset of male puberty (11). It is therefore desirable that population studies also include assessments of TV. However, measuring TV with a Prader orchidometer may be perceived as intrusive outside of clinical context. In such situations, the use of ultrasound (US) to measure TV could be more acceptable because of the more technical nature of the examination. US is also the preferred method when the accuracy of TV is important (12). In addition, this method has the advantage of detecting testicular pathology, which may explain developmental differences of testicular growth. We have previously shown that the measurement of TV with US is methodologically feasible and appropriate for the generation of pubertal references for TV (13).

The aim of the current study was to estimate references for TV based on US measurements of the testicle and Tanner pubic hair (PH) staging in a representative cohort of healthy Norwegian boys.

Materials and Methods

Childhood population: This study is a part of the Bergen Growth Study 2 on pubertal growth and development in Norway. All boys attending one of six randomly selected schools that

provide primary- and secondary education in the city of Bergen, Norway, were invited to participate in the study from January through June 2016. Parental consent was obtained for 493 (37%) out of 1329 eligible boys, but two boys did not assent and six were absent on the day of examination. In addition, we included US measurements from 58 boys who participated in a reliability study in 2017 (13). The age at examination was calculated from date of birth and date of examination. Eight boys with a disease or condition that could affect growth and 21 boys with a history of scrotal pathology (including cryptorchidism, hydrocele and microlithiasis) were excluded, and the Tanner PH stage was not registered in 62 boys. A parental questionnaire was obtained for 340 of the 514 (66.1%) boys included in the analysis. The questionnaire contained items on origin, chronic disease and previous genital pathology. Origin was grouped as both parents from Norway, one or both parents from the European region, and one or both parents from outside the European region. Height and weight were measured in 457 of the boys included in the analysis. Based on the International Obesity Task Force (IOTF) body mass index (BMI) reference values (14), 11.8% of boys were classified as overweight (IOTF-BMI ≥ 25 kg/m^2) and 2.1% as obese (IOTF-BMI ≥ 30 kg/m^2). This closely matches the 12.8% and 2.1% reported for Norwegian boys in this age range (15).

Ultrasound: All US examinations of the testis were performed by a single technician using a Sonosite Edge US machine with a 15-6 MHz linear probe. The length (L), width (W) and depth (D) of the right testicle were measured with the boy in a supine position according to a standardized protocol (13). If the left testicle appeared larger by visual inspection, this was also measured, and the volume of the largest testicle was registered. The TV was calculated from the length, width and depth using the Lambert equation as $TV = L \times W \times D \times 0.71$. The observer variability of this method is 9.2% with a technical error of measurement (TEM) of 6.5% (13). An empirical equation to predict Prader orchidometer volume from US volume was previously derived as $Vol_{OM} = 1.96 \times Vol_{US}^{0.71}$. The Prader orchidometer volume of ≥ 4ml that defines puberty onset is thus equivalent to an US measured volume ≥ 2.7 ml (13). A

preliminary logistic regression analysis of pubertal onset (yes/no) according to age and origin showed no statistically significant differences between boys of Norwegian origin and European (p = 0.17) or non-European boys (p = 0.11). All boys were therefore included in the analysis of TV during puberty.

Tanner staging: Tanner PH stages were visually assessed in the supine position by the same observer performing the US examinations. Illustrated descriptions based on the work of Marshall and Tanner served as a reference (16). A preliminary analysis showed no statistical significant difference in the timing of pubarche (Tanner PH stage 2) between boys of Norwegian origin and European (p = 0.82) or non-European boys (p = 0.59). All boys were therefore included in the analysis of pubic hair stages.

Statistical analyses**:** A reference curve of the continuous US testicular volume for age was estimated with the LMS method (17). The LMS method normalizes the distribution of a variable by applying a Box-Cox power transformation to remove skewness from the data. The reference is summarized by three curves, representing the Box-Cox power to remove skewness (L), the mean (M), and the approximate coefficient of variation (S) along the independent covariate age. The amount of smoothing is expressed in terms of smoothing parameters or equivalent degrees of freedom (edf). For the TV, the optimal Box-Cox power L was determined to be constant at 0.5 (i.e. a square root transformation), the M-curve was fitted with 8 edf, and the S-curve with 4 edf. The tabulated values of L, M and S by age contain all the information that is needed to calculate any percentile, or to convert measurements into z-scores. Because L is a constant of 0.5, centiles can be derived using the simplified formula $C_\alpha = M\,(1 + Sz_\alpha/2)^2$, and z-scores as $Z = 2\times(\sqrt{X/M} - 1)/S$, where z_α is the normal equivalent deviate that corresponds to the desired percentile.

In addition to the LMS reference curves for the continuous TV, we used probit regression within a generalized linear model (GLM) to estimate cumulative incidence curves for reaching TVs that correspond to selected discrete Prader orchidometer volumes, and for each of the Tanner PH stages. Non-parametric generalized additive models (GAM) provided identical

results, which confirmed our assumption of a normal age distribution at the different pubertal milestones (data not shown).

All statistical analyses were performed using R version 3.4 (R foundation for Statistical Computing, Vienna, Austria) or IBM SPSS statistics version 24 (IBM Corp.,)

Ethical considerations: 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 cinema voucher was given as an incentive. The study was approved by the Regional Committee for Medical and Health Research Ethics West (REC-WEST 2015/128).

Results

Childhood population: The number of boys varied from 28 to 66 per age year between six and 16 years, and 12 boys were 16 years of age. Based on information from the questionnaire, 77.4% had two Norwegian parents, 10% had one or two parents from another European country, and 12.5% had one or two parents from outside the European region. US examination of the scrotum, revealed microlithiasis in one boy and testis located in the inguinal canal in two boys. In addition, we observed twelve cases of unilateral and six cases of bilateral cryptorchidism.

Testicular volume: A total of 514 boys with a mean age of 11.0 years (range: 6.1 to 16.4 years) were included for the references. Figure 1 shows the US testicular volumes by age and the fitted median and ± 2 SD lines. The corresponding L, M and S values are listed in Table 1 and selected percentiles are provided in supplementary Table S1. Figure 2 shows the cumulative incidence curves of selected discrete Prader orchidometer volumes by age, which were derived from the US volumes using the formula given in the methods section. The corresponding age quantiles are listed in Table 2 as SD scores and in the supplementary Table S2 as age percentiles. The mean age (SD) for attainment of a US

measured TV of 2.7 ml (equivalent to a Prader orchidometer volume of 4 ml) was 11.7 (1.1) years and the 3rd and 97th percentiles were respectively 9.7 and 13.7 years.

Pubic hair: Tanner PH stage was determined in 452 (88.0%) boys with a mean age of 10.9 years (range 6.1 to 16.3 years). Figure 3 shows the cumulative incidence curves when boys reach Tanner PH stages 2 – 5. The mean age (SD) of pubarche (Tanner PH stage 2) was 11.8 (1.2) years, and the corresponding reference range defined by the 3rd and 97th percentiles was 9.5 to 14.1 years. The mean ages and reference quantiles of other PH stages are listed in Table 3 as SD scores and in supplementary Table S3 as age percentiles.

The distribution of the continuous US TVs in boys who were classified as Tanner PH stages 1 – 5 is shown in Figure 4. There is both a substantial spread within, and overlap between, groups in terms of TV. When boys were classified as pre-pubertal (1-3 ml), pubertal (4-14 ml) or adult (≥ 15ml) based on the equivalent Prader orchidometer volumes, 14.0% of boys with a pubertal testicular volume were characterized as Tanner PH1 (i.e. no pubarche), while among boys with a pre-pubertal testicular volume, 8.1% were characterized as ≥Tanner PH2 (pubarche). All boys with Tanner PH stages 4 or 5 had attained either a pubertal or an adult volume of the testicles.

Discussion

In the current study, we present contemporary references for TV, obtained with US, and Tanner PH stages in 6 to 16 year-old Norwegian boys. By using LMS centile curves to summarize TV by age, the assessment of pubertal growth can now be quantified on a continuous scale since measurements can easily be converted to age adjusted SD scores. In addition, we present age percentiles from probit analyses that document the cumulative incidence of reaching milestones of the development of TV and pubic hair.

In this paper, we report the very first references for pubertal development in Norwegian boys. Although it has been shown that, the timing of puberty in Northern European populations is

very similar (18), national reference data collected at regular intervals can help to detect secular trends earlier (3). Up to date references are also important because early or late puberty may have consequences for the health of individual boys. Studies have found a protective effect of later puberty on testicular cancer (19), but delayed puberty has also been linked to bullying, poor self-esteem and psychosocial distress (20).

A testicular volume of 4 ml measured with a Prader orchidometer is commonly considered as a robust marker of the start of puberty (11). However, we (13) and others (21) have previously shown that the orchidometer overestimates the true volume near this range, and that the actual volume at the start of puberty is about 2.7 ml when measured with ultrasound (13). This also corresponded nicely with increased sex hormone levels (22). While our references of TV are primarily based on US measurements, we previously devised a conversion formula that allow a seamless conversion from one method to the other (13). Apart from being closer to the true TV, the US method has the additional advantage that the volume is measured on a continuous scale, contrary to the Prader orchidometer method which is limited to reaching a discrete set of volumes making it difficult to estimate volumes in between two consecutive beads or beyond all available beads.

The onset of puberty, defined by an US TV of 2.7 ml (4 ml. with the Prader orchidometer), was reached at a mean age of 11.7 years. This is highly comparable with the 11.6 years observed in Dutch boys by Goede et al. (23), and later remodeled by Joustra et al. (24), which is the only US reference for TV for adolescent boys published to date. In order to compare our data with these references we multiplied the Dutch estimates with a factor 0.71/0.52 because these studies used the ellipsoid formula (L×W×D×0.52) to calculate volumes from testicular dimensions, while we used the Lambert formula (LxWxDx0.71) because it was found to give a better approximation (25).

Our data are also in agreement with the age of attainment of an equivalent Prader orchidometer volume of 4 ml. in other European countries, e.g. 11.4 years in Belgium (26), 11.5 years in the Netherlands (27) and 11.7 years in Denmark (6). While the Copenhagen

Puberty Study reported a decline in age at onset of puberty of three months between 1991 and 2006, this trend was no longer significant after adjustment for BMI (6). In the United States, the PROS study from 2005-2010 reported a mean age of 11.5 years in the non-Hispanic white population and 11.8 years in the African American population (28). A comparison of our data with these references does not suggest a secular trend towards earlier puberty in boys over the last decade. However, compared to data from 109 Norwegian boys aged 1.9-16.9 years collected by Waaler in the 1970's (29), contemporary boys reach a pubertal testicular volume approximately 2-3 months earlier, i.e. a rate of less than one month per decade.

Studies in European populations like Denmark (30, 31), Belgium (26), Italy (32) and The Netherlands (27), reported relatively narrow range of mean age at pubarche (Tanner PH2) of 11.5 to 11.9 years. Our finding of 11.8 years corresponds with this range. One study from Denmark reported a mean age of 12.4 years, but this was a surprising finding because it implied a slow down with approximately five months between 1991 and 2006 (6), while other studies during the same period in Denmark reported stable average ages at pubarche of 11.6 (31) and 11.9 years (30). The PROS and NICHD studies from the United States both reported a mean age at Tanner PH2 in the non-Hispanic white population of 11.5 years (28, 33) , which is about three months earlier compared to Norwegian boys.

Longitudinal studies have shown that around 46-90% of boys enter puberty by the "testicular" pathway, i.e. gonadal enlargement before the appearance of pubic hair (pubarche pathway) (11, 31, 33). The mean ages of attainment of a pubertal TV (11.7 years) and Tanner PH2 (11.8 years) in our study are consistent with this. Because our data are cross-sectional, we cannot estimate the duration of each stage, nor the pace of progression throughout the various pubertal stages. However, a direct comparison of different pubertal markers showed that 14.0% of the boys reached a pubertal TV before pubic hair appeared, whereas only 8.1% showed pubic hair (Tanner PH2) prior to reaching a pubertal TV. Importantly, careful assessment of TV is the most reliable method to detect the earliest signs of puberty, whereas

PH staging alone may lead to misclassification of some boys in the earliest segment of pubertal maturation (11).

Previous population studies have defined the normal physiological range for pubertal development in boys as 2.5 or 3 time the standard deviation below and above the mean (18). In our study, the reference age range (mean ± 2.5 SD) of reaching a pubertal TV (US measured TV of 2.7 ml) is bounded by the ages of 9.0 and 14.3 years. We therefore recommend adhering to the current definition of normal pubertal onset in boys between 9 to 14 years.

A major strength of our study is the use of US for the measurement of TV in a population-based study. US provides the opportunity to obtain more accurate estimates of TV in comparison with a Prader orchidometer, without interference of surrounding tissues, such as the scrotal skin, epididymis or the tunica vaginalis. Furthermore, US provides a continuous measure of volume, in contrast to the discrete ordinal Prader orchidometer beads, which allows for semiparametric data modeling and calculation of z-scores. US has the additional benefit of detecting testicular pathology, which may explain alterations in the timing of testicular growth, as in our cohort, we found one patient with testicular microlithiasis and two boys with testis located in the inguinal canal. Further, the majority (58%) of the participating boys from the test/retest-study reported to prefer the examination of TV with US to direct palpation of the testicle. This may be explained by the less intrusive positioning of the examiner, facing the US machine rather than the scrotum directly, and that there is no direct contact between examiner's hand and the scrotum. Since US equipment and protocols are becoming more user-friendly and accessible, and because they are known to be safe and without a risk of ionizing radiation, they might more readily be adopted for routine use by pediatricians and other clinicians.

Some limitations to the study needs to be addressed. Only 37% of the invited boys agreed to participate. This makes a selection bias possible, for instance, if boys maturing early or very late were less inclined to participate. In addition, only boys up to 16.4 years were included,

potentially omitting the stabilization of testicular growth at the adult range in our reference curve. Due to difficulties of recruitment and a potential reluctance regarding palpation of the testicles, examination with a Prader orchidometer was only performed in the reliability study. Based on these examinations we were able to estimate a conversion equation to calculate Prader orchidometer volume from US volume. Another limitation was that only the right testicle was measured, except when the left testicle appeared larger by visual inspection. However, no statistical significant differences between left and right TV have been found in previous studies (34).

Conclusion

We have presented references for testicular growth based on US measurements of testicular dimensions, and for the clinical assessment of Tanner PH stages. Prader orchidometer has long been considered a subjective clinical tool that is limited to an ordinal scale. Our implementation of an US protocol implies a transition towards an objective measurement on a continuous scale that allows to detect smaller changes in the TV, and allows to quantify pubertal development. Further, US was the preferred examination method amongst the majority of the boys. The high degree of similarity of our data with previously published estimates of puberty onset in boys does not suggest an ongoing secular trend during the past decade.

Acknowledgments

We thank Magnus R. Sveen for substantial help during data collection and to all the boys and schools for participating

Abbreviations

US: ultrasound

TV: testicular volume

PH: pubic hair

PROS: Pediatric Research in Office Settings

NICHD: National Institute of Child Health and Human Development

Statement of financial support: The Western Norway Regional Health Authority, Grant number 911975, supported this study.

Financial Disclosure: The authors have no financial relationships relevant to this article to disclose.

Potential Conflicts of Interest: The authors have no conflicts of interest relevant to this article to disclose.

References:

1. Aksglaede L, Sorensen K, Petersen JH, Skakkebaek NE, Juul A. Recent decline in age at breast development: the Copenhagen Puberty Study. *Pediatrics* 2009; 123 5:e932-9.
2. Herman-Giddens ME, Slora EJ, Wasserman RC, Bourdony CJ, Bhapkar MV, Koch GG, et al. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. *Pediatrics* 1997; 99 4:505-12.
3. Juul A, Teilmann G, Scheike T, Hertel NT, Holm K, Laursen EM, et al. Pubertal development in Danish children: comparison of recent European and US data. *International journal of andrology* 2006; 29 1:247-55; discussion 86-90.
4. Euling SY, Herman-Giddens ME, Lee PA, Selevan SG, Juul A, Sorensen TI, et al. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. *Pediatrics* 2008; 121 Suppl 3:S172-91.
5. Herman-Giddens ME, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988-1994. *Arch Pediatr Adolesc Med* 2001; 155 9:1022-8.
6. Sorensen K, Aksglaede L, Petersen JH, Juul A. Recent changes in pubertal timing in healthy Danish boys: associations with body mass index. *The Journal of clinical endocrinology and metabolism* 2010; 95 1:263-70.
7. Reinehr T, Roth CL. Is there a causal relationship between obesity and puberty? *The Lancet Child & adolescent health* 2019; 3 1:44-54.
8. Lee JM, Wasserman R, Kaciroti N, Gebremariam A, Steffes J, Dowshen S, et al. Timing of Puberty in Overweight Versus Obese Boys. *Pediatrics* 2016; 137 2:e20150164.
9. Parent AS, Franssen D, Fudvoye J, Pinson A, Bourguignon JP. Current Changes in Pubertal Timing: Revised Vision in Relation with Environmental Factors Including Endocrine Disruptors. *Endocrine development* 2016; 29:174-84.
10. Abreu AP, Kaiser UB. Pubertal development and regulation. *The lancet Diabetes & endocrinology* 2016; 4 3:254-64.
11. Biro FM, Lucky AW, Huster GA, Morrison JA. Pubertal staging in boys. *J Pediatr* 1995; 127 1:100-2.
12. Kuijper EA, van Kooten J, Verbeke JI, van Rooijen M, Lambalk CB. Ultrasonographically measured testicular volumes in 0- to 6-year-old boys. *Human reproduction (Oxford, England)* 2008; 23.
13. Oehme NHB, Roelants M, Bruserud IS, Eide GE, Bjerknes R, Rosendahl K, et al. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys - intra- and interobserver agreement and comparison with Prader orchidometry. *Pediatric radiology* 2018; 48 12:1771-8.
14. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ (Clinical research ed)* 2000; 320 7244:1240-3.
15. Juliusson PB, Roelants M, Eide GE, Hauspie R, Waaler PE, Bjerknes R. Overweight and obesity in Norwegian children: secular trends in weight-for-height and skinfolds. *Acta paediatrica (Oslo, Norway : 1992)* 2007; 96 9:1333-7.
16. Rasmussen AR, Wohlfahrt-Veje C, Tefre de Renzy-Martin K, Hagen CP, Tinggaard J, Mouritsen A, et al. Validity of self-assessment of pubertal maturation. *Pediatrics* 2015; 135 1:86-93.
17. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in medicine* 1992; 11 10:1305-19.
18. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocrine reviews* 2003; 24 5:668-93.

19. Weir HK, Kreiger N, Marrett LD. Age at puberty and risk of testicular germ cell cancer (Ontario, Canada). *Cancer causes & control : CCC* 1998; 9 3:253-8.
20. Golub MS, Collman GW, Foster PM, Kimmel CA, Rajpert-De Meyts E, Reiter EO, et al. Public health implications of altered puberty timing. *Pediatrics* 2008; 121 Suppl 3:S218-30.
21. Diamond DA, Paltiel HJ, DiCanzio J, Zurakowski D, Bauer SB, Atala A, et al. Comparative assessment of pediatric testicular volume: orchidometer versus ultrasound. *The Journal of urology* 2000; 164 3 Pt 2:1111-4.
22. Madsen A, Oehme NB, Roelants M, Bruserud IS, Eide GE, Viste K, et al. Testicular ultrasound to stratify hormone references in a cross-sectional Norwegian study of male puberty. *The Journal of clinical endocrinology and metabolism* 2019.
23. Goede J, Hack WW, Sijstermans K, van der Voort-Doedens LM, Van der Ploeg T, Meij-de Vries A, et al. Normative values for testicular volume measured by ultrasonography in a normal population from infancy to adolescence. *Hormone research in paediatrics* 2011; 76 1:56-64.
24. Joustra SD, van der Plas EM, Goede J, Oostdijk W, Delemarre-van de Waal HA, Hack WW, et al. New reference charts for testicular volume in Dutch children and adolescents allow the calculation of standard deviation scores. *Acta paediatrica (Oslo, Norway : 1992)* 2015; 104 6:e271-8.
25. Hsieh ML, Huang ST, Huang HC, Chen Y, Hsu YC. The reliability of ultrasonographic measurements for testicular volume assessment: comparison of three common formulas with true testicular volume. *Asian journal of andrology* 2009; 11 2:261-5.
26. Roelants M, Hauspie R, Hoppenbrouwers K. References for growth and pubertal development from birth to 21 years in Flanders, Belgium. *Annals of human biology* 2009; 36 6:680-94.
27. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, et al. Continuing positive secular growth change in The Netherlands 1955-1997. *Pediatric research* 2000; 47 3:316-23.
28. Herman-Giddens ME, Steffes J, Harris D, Slora E, Hussey M, Dowshen SA, et al. Secondary sexual characteristics in boys: data from the Pediatric Research in Office Settings Network. *Pediatrics* 2012; 130 5:e1058-68.
29. Waaler PE, Thorsen T, Stoa KF, Aarskog D. Studies in normal male puberty. *Acta paediatrica Scandinavica Supplement* 1974; 249:1-36.
30. Wohlfahrt-Veje C, Mouritsen A, Hagen CP, Tinggaard J, Mieritz MG, Boas M, et al. Pubertal Onset in Boys and Girls Is Influenced by Pubertal Timing of Both Parents. *The Journal of clinical endocrinology and metabolism* 2016; 101 7:2667-74.
31. Mouritsen A, Aksglaede L, Soerensen K, Hagen CP, Petersen JH, Main KM, et al. The pubertal transition in 179 healthy Danish children: associations between pubarche, adrenarche, gonadarche, and body composition. *European journal of endocrinology* 2013; 168 2:129-36.
32. De Simone M, Danubio ME, Amicone E, Verrotti A, Gruppioni G, Vecchi F. Age of onset of pubertal characteristics in boys aged 6-14 years of the Province of L'Aquila (Abruzzo, Italy). *Annals of human biology* 2004; 31 4:488-93.
33. Susman EJ, Houts RM, Steinberg L, Belsky J, Cauffman E, Dehart G, et al. Longitudinal development of secondary sexual characteristics in girls and boys between ages 91/2 and 151/2 years. *Arch Pediatr Adolesc Med* 2010; 164 2:166-73.
34. Tatsunami S, Matsumiya K, Tsujimura A, Itoh N, Sasao T, Koh E, et al. Inter/intra investigator variation in orchidometric measurements of testicular volume by ten investigators from five institutions. *Asian journal of andrology* 2006; 8 3:373-8.

Tables

Table 1 Age-specific references for ultrasound testicular volume (ml) estimated from 514 healthy Norwegian boys aged 6-16 years in 2016-2017 using the LMS method[a)]

Age	L	M	S	-2 SD	-1 SD	Mean	+1 SD	+2 SD
6	0.5	0.78	0.24622	0.4	0.6	0.8	1.0	1.2
7	0.5	0.83	0.30388	0.4	0.6	0.8	1.1	1.4
8	0.5	0.91	0.36146	0.4	0.6	0.9	1.3	1.7
9	0.5	1.00	0.41844	0.3	0.6	1.0	1.5	2.0
10	0.5	1.23	0.47844	0.3	0.7	1.2	1.9	2.7
11	0.5	2.02	0.52162	0.5	1.1	2.0	3.2	4.7
12	0.5	3.77	0.54007	0.8	2.0	3.8	6.1	8.9
13	0.5	6.30	0.51981	1.5	3.5	6.3	10.0	14.6
14	0.5	9.39	0.46762	2.7	5.5	9.4	14.3	20.2
15	0.5	12.58	0.39673	4.6	8.1	12.6	18.1	24.5
16	0.5	15.64	0.32065	7.2	11.0	15.6	21.1	27.0

Abbreviations: L: Skewness parameter; M: mean, S: coefficient of variation; SD: standard deviation

a) The LMS model was fitted on the original age scale with squared root transformed volumes (L=0.5) and 8 and 4 equivalent degrees of freedom for the M and S curves.

Table 2 Age distribution for testicular volumes corresponding to equivalent Prader orchidometer volumes (ml) in 514 healthy Norwegian boys aged 6-16 years in 2016-2017

Prader	USV	M	SE	SD	-2 SD	-1 SD	+1 SD	+2 SD
2	1.0	8.95	0.16	2.01	4.9	6.9	11.0	13.0
3	1.8	11.05	0.11	1.25	8.6	9.8	12.3	13.6
4	2.7	11.67	0.11	1.07	9.5	10.6	12.7	13.8
5	3.7	12.25	0.11	1.01	10.2	11.2	13.3	14.3
6	4.8	12.66	0.11	0.97	10.7	11.7	13.6	14.6
8	7.2	13.49	0.12	1.10	11.3	12.4	14.6	15.7
10	9.9	14.08	0.13	1.20	11.7	12.9	15.3	16.5
12	12.8	15.10	0.15	1.32	12.5	13.8	16.4	17.7
15	17.6	16.27	0.32	1.54	13.2	14.7	17.8	19.4

Abbreviations and symbols: Prader = equivalent Prader orchidometer volumes; USV = Ultrasound volume; M = mean; SE = standard error; SD = standard deviation

Table 3 Age distribution (years) by Tanner pubic hair stage (PH) in 452 healthy Norwegian boys aged 6-16 years in 2016-2017

PH	Mean	SE	SD	-2 SD	-1 SD	+1 SD	+2 SD
2	11.78	0.12	1.22	9.3	10.6	13.0	14.2
3	12.68	0.12	1.12	10.4	11.6	13.8	14.9
4	13.46	0.11	0.86	11.7	12.6	14.3	15.2
5	14.42	0.12	0.90	12.6	13.5	15.3	16.2

Abbreviations and symbols: PH = Tanner pubic hair stage; SE = standard error; SD = standard deviation

Figures

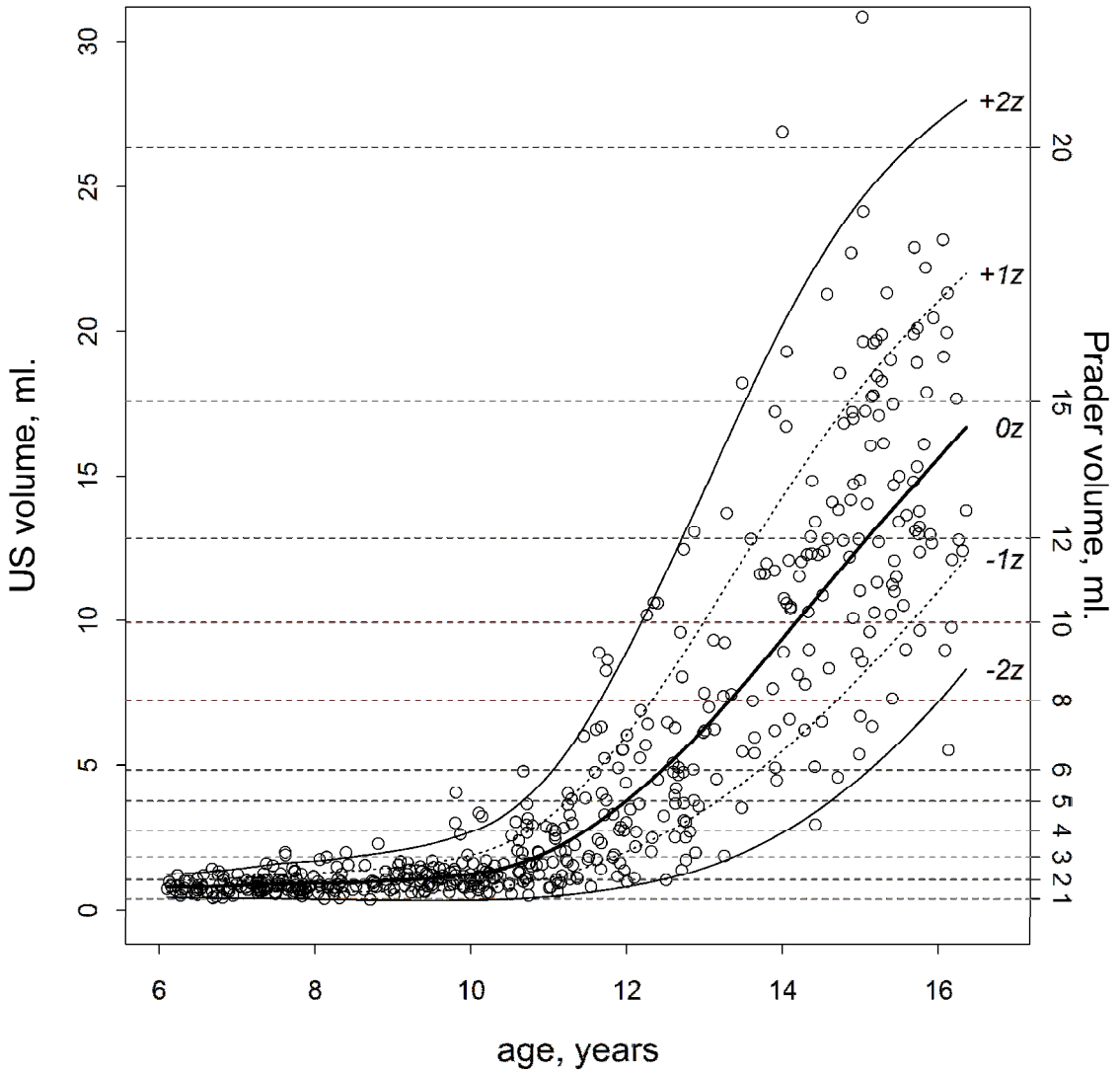

Figure 1 LMS-smoothed reference chart of ultrasound (US) measured testicular volume in 514 healthy Norwegian boys, aged 6-16 years. Corresponding equivalent Prader orchidometer volumes are shown on the right axis.

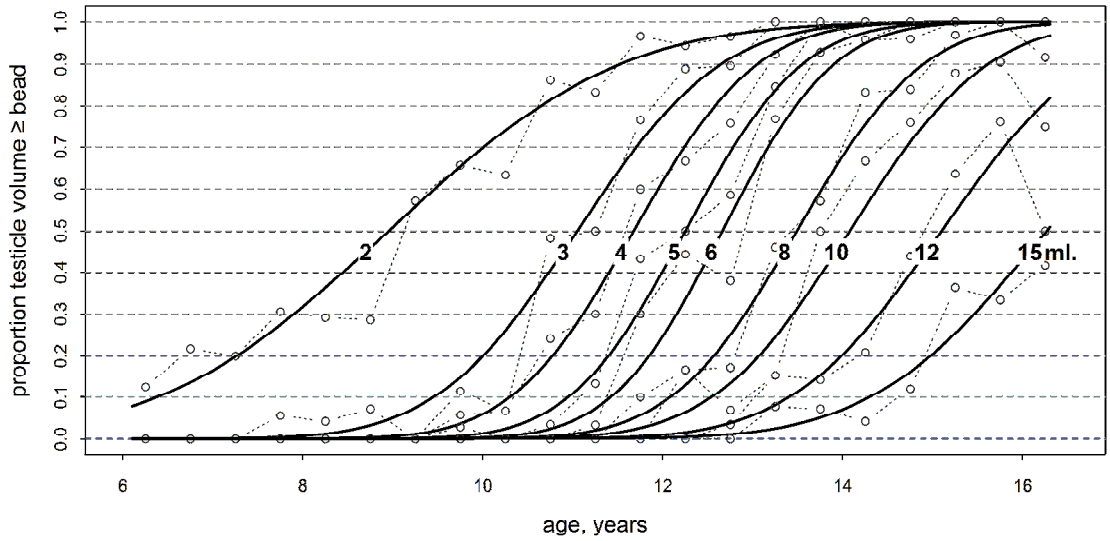

Figure 2 Cumulative incidence of reaching selected equivalent Prader orchidometer volumes estimated with probit regression in 514 healthy Norwegian boys aged 6-16 years. Connected markers show the empirical data and bold lines the corresponding probit models.

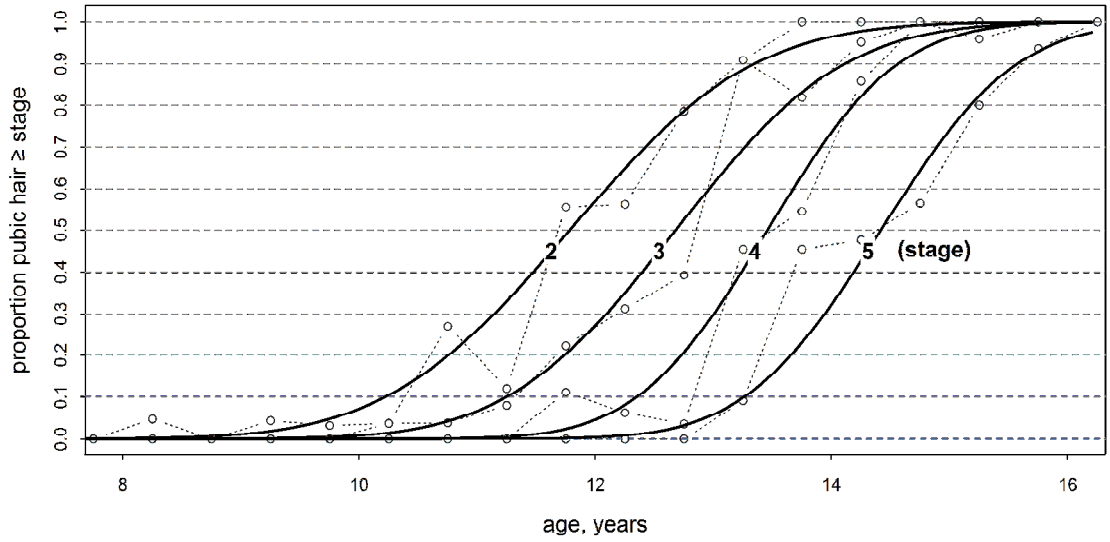

Figure 3 Cumulative incidence of Tanner stages for pubic hair in 452 healthy Norwegian boys aged 6-16 years. Connected markers show the empirical observations and bold lines the corresponding probit models.

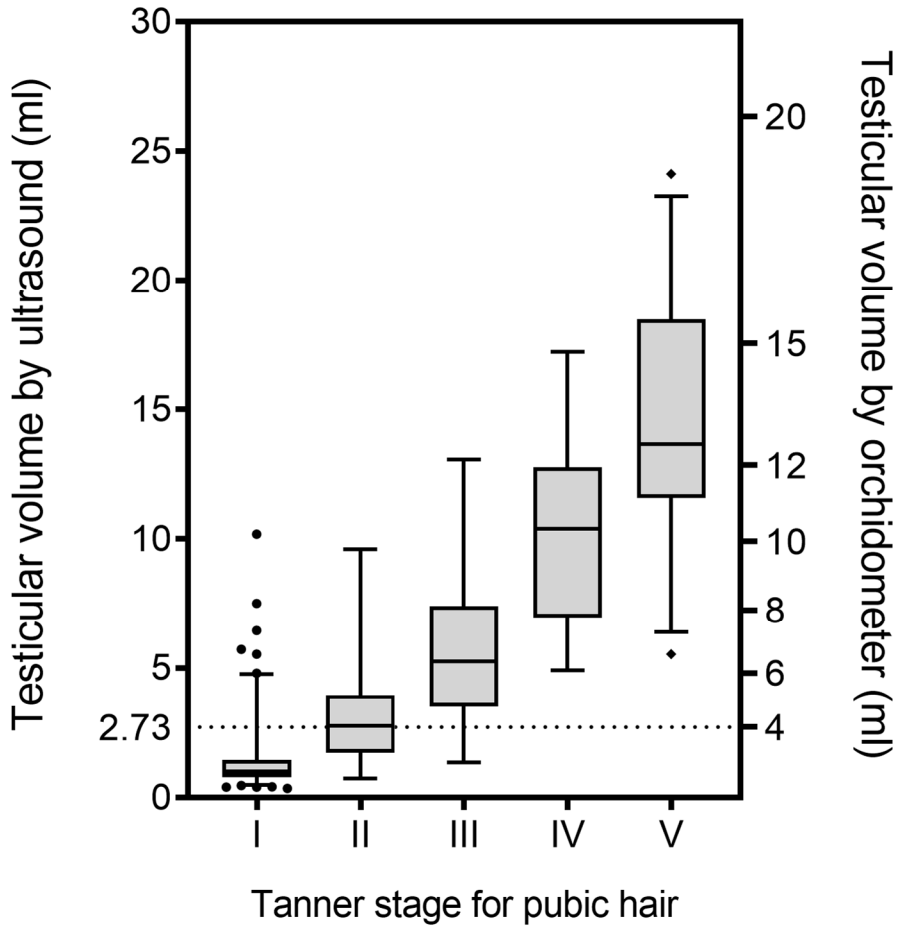

Figure 4 Box and whiskers plots of testicular volumes at different Tanner pubic hair stages in 452 Norwegian boys aged 6.1-16.4 years in 2016-2017. The boundaries of the box are the 1st and 3rd quartile. The median is identified by a line inside the box. The length of the box is the interquartile range (IQR).