Testicular ultrasound to stratify hormone references in a cross-sectional Norwegian study of male puberty

Andre Madsen[1,2], Ninnie B. Oehme[2,3], Mathieu Roelants[4], Ingvild S. Bruserud[2,3], Geir Egil Eide[5,6], Kristin Viste[1], Robert Bjerknes[2,3], Bjørg Almås[1], Karen Rosendahl[7,8], Jørn V. Sagen[1,2], Gunnar Mellgren[1,2] and Petur B. Juliusson[2,3,9]

[1] Hormone Laboratory, Haukeland University Hospital, Bergen, Norway

[2] Department of Clinical Science, University of Bergen, Bergen, Norway

[3] Department of Pediatrics, Haukeland University Hospital, Bergen, Norway

[4] Environment and Health, Department of Public Health and Primary Care, KU Leuven, University of Leuven, Leuven, Belgium

[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, Bergen, Norway

Keywords: Puberty, hormones, reference interval, testicular volume

Corresponding author, to whom reprint requests should be addressed:

Andre Madsen, PhD.

Department of Clinical Science, University of Bergen

N-5021 Bergen, Norway

Telephone: +47 97986919

E-mail: andre.madsen@uib.no

Grants: This study was funded by grants from the Western Norway Regional Health Authority, grant no. 912221.

Disclosure summary

I accept the terms of the Endocrine Society Copyright Transfer and Affirmation of Originality agreement on behalf of myself and all co-authors.

I certify that neither I nor my co-authors have a conflict of interest as described above that is relevant to the subject matter or materials included in this Work.

Abbreviations: FSH, follicle-stimulating hormone; LH, luteinizing hormone; SHBG, sex hormone-binding globulin; US, ultrasound; OM, orchidometer; TV, testicular volume; LC-MS/MS, liquid chromatography tandem-mass spectrometry; CLSI, Clinical Laboratory Standards Institute; CV, coefficient of variation; SD, standard deviation; ANOVA, analysis of variance.

ABSTRACT

Context: Testicular growth represents the best clinical variable to evaluate male puberty, but current pediatric hormone references are based on chronological age and subjective assessments of discrete puberty development stages. Determination of testicular volume (TV) by ultrasound provides a novel approach to assess puberty progression and stratify hormone reference intervals.

Objective: To establish references for serum testosterone and key hormones of the male pituitary-gonadal signaling pathway in relation to TV determined by ultrasound.

Design Setting and Participants: Blood samples from 414 healthy Norwegian boys between 6 - 16 years of age were included from the cross-sectional "Bergen Growth Study 2". Participants underwent testicular ultrasound and clinical assessments and serum samples were analyzed by liquid chromatography tandem-mass spectrometry (LC-MS/MS) and immunoassays.

Main Outcome Measures: We present references for circulating levels of total testosterone, luteinizing hormone (LH), follicle-stimulating hormone (FSH) and sex-hormone binding globulin (SHBG) in relation to TV, chronological age and Tanner pubic hair (PH) stages.

Results: In pubertal boys, TV accounted for more variance in serum testosterone levels than chronological age (Spearman's $r = 0.753$, $p<0.001$ vs $r = 0.692$, $p<0.001$, respectively). Continuous centile references demonstrate the association between TV and hormone levels during puberty. Hormone reference intervals were stratified by TV during the pubertal transition.

Conclusions: Objective ultrasound assessments of TV and stratification of hormone references increase the diagnostic value of traditional references based on chronological age or subjective staging of male puberty.

Précis

Ultrasound assessments of testicular volume in Norwegian boys were leveraged to establish improved hormone reference intervals for male puberty.

INTRODUCTION

Biochemical reference intervals are conventionally defined by the limits of the central 95% of the distribution within a reference population (1-3). Reliable reference limits are important to avoid misdiagnosis and accordingly, medical laboratories should continuously strive to verify and update references when new assays are implemented. The main methodological framework for establishing reference intervals is outlined in the proprietary Clinical Laboratory Standards Institute (CLSI) EP28-A3c guidelines and elsewhere (3). Whereas several clinical laboratories are still limited to immunoassay detection of steroid hormones such as testosterone, liquid chromatography tandem-mass spectrometry (LC-MS/MS) permits highly selective quantification of serum testosterone in the sub- to low nanomolar range. Hormone references are essential to identify endocrinopathies in children that exhibit abnormal progression towards a biochemical or developmental end-point and may require medical follow-up (4). Most pediatric hormone references are based on chronological age, but age is not indicative of physiological status during puberty. Furthermore, well-documented secular trends to earlier puberty timing may have implications for male reproductive health (5).

Hormones of the hypothalamic-pituitary-gonadal (HPG) axis orchestrate the development of reproductive organs and sex-specific somatic traits during puberty. Pituitary-derived follicle-stimulating hormone (FSH) and luteinizing hormone (LH) stimulate the development and function of spermatogenic Sertoli cells and testosterone-producing Leydig cells, respectively (6). The bioavailable component of circulating testosterone, unbound by sex hormone-binding globulin (SHBG), contributes to male fertility and secondary sex characteristics and provides negative feedback to the HPG signaling axis. On the somatic level, the traditional Tanner stages are still widely used to define pubertal progression and milestones of sex-specific developmental processes related to reproduction or sexual maturation (7). Prader orchidometry employs a string of beads to measure TV in whole-number milliliter increments.

The onset of male puberty is defined by attainment of testicular volume (TV) greater than 3 ml or equal to 4 ml when judged by traditional Prader orchidometry (8, 9). By convention, testicles are defined as prepubertal in the 1 - 3 ml range, pubertal in the 4 - 12 ml range and adult in the 15 - 25 ml range. However, the clinical practice of determining male puberty onset and progression in terms of Tanner stages and orchidometry may be prone to inaccuracy due to the subjective process of assigning visual and palpative impressions to discrete ordinal stages (10, 11). Notably, palpation systematically overestimates TV for small testicles due to the methodological inability to differentiate the central testicle from surrounding epididymis, scrotal skin and tunica capsule layers (12). Recently, ultrasound assessment of TV has been adopted as a modern alternative to orchidometry. Previous studies have demonstrated that data obtained from ultrasound determination of TV correlates accurately with traditional orchidometry (13-16).

Previous studies have profiled endocrine changes throughout male puberty, but such studies have not considered TV (3, 17, 18) or did not succeeded in obtaining a sufficient population sample to estimate valid reference intervals (19). The evident ongoing decline in timing of puberty in the Danish population (20, 21) prompted us to benchmark male puberty timing in a representative sample of healthy Norwegian children. Apart from our work, the only previous study that reported on contemporary puberty timing in Norwegian boys was conducted in 1974 (22). We have recently conducted a puberty study in which TV was resolved by ultrasound based on our previously published radiological protocol (15). In the current study, we wanted to address endocrine aspects of male puberty in relation to testicular growth. Specifically, we aimed to establish a novel set of hormone references taking into account actual ultrasound-determined TV, in addition to standard age-partitioned references.

MATERIALS AND METHODS

Study sample

The "Bergen Growth Study 2" (BGS2) is a cross-sectional cohort study with the purpose of characterizing and benchmarking the timing and progression of puberty in contemporary Norwegian children. Subjects from six schools in the Bergen area in Western Norway were recruited and examined during the year 2016. A total of 491 boys between 6 and 16 years of age (total participation rate 36.6%) were voluntarily recruited in the study. Participants with known chronic diseases that could affect growth were excluded (n=6). A total of 428 boys consented to ultrasound examination of the testicles, of whom 19 participants were excluded due to findings of non-threatening scrotal pathologies including microcalcifications, hydrocele or cryptorchidism (total n=19). Blood samples were acquired from 414 healthy individuals, of which there were 414 accounts of participant chronological age, 406 accounts of ultrasound testicular volume and 403 accounts of Tanner pubic hair (PH) stages. The vast majority of our male sample were children with two Norwegian parents (77.5%) or one or both parents from the European region (10%), but a minority of participants of African, Asian and Hispanic origin were also included to provide references for all children living in Norway.

Clinical inspection and ultrasound assessment

Tanner PH stages were visually determined with respect to the quantity, characteristics and distribution of pubic hair in accordance to an illustrated descriptive reference based on Marshall and Tanner (23). The dimensions of the right testicle were examined by one technician using a SonoSite Edge device (FUJIFILM SonoSite, Inc, USA) as described previously (15). Only in cases where the left testicle appeared visually larger than the right testicle (n=3 cases), the left testicle was measured by ultrasound. In the literature, consensus is that there is no significant difference between left and right testicular volumes in healthy males (24, 25). Ultrasound measures of elliptical length (L), width (W) and depth (D) were converted to

ultrasound testicular volume (TV_{US}) using Lambert's equation (0.71 x L x W x D). Traditional Prader orchidometry defines testicle maturation intervals as pre-pubertal (<3 ml), pubertal (4 - 12 ml) and adult (15 - 25 ml), respectively. TV_{US} was converted to the equivalent orchidometer volume (TV_{OM}) using the non-linear formula $TV_{OM} = 1.96 \times TV_{US}^{0.71}$ as described previously (15). By the reciprocal formula, equivalent intervals for TV_{US} measured by ultrasound were defined: pre-pubertal (<2.7 ml), pubertal (2.7 - 12.8 ml) and adult (>17.6 ml).

Laboratory and blood sample analyses

Blood samples were collected between 0800 and 1400 h. Isolated serum was stored at -80°C until analysis at the Hormone Laboratory at Haukeland University Hospital, Bergen, Norway. The laboratory and its methods are accredited according to NS-EN ISO 15189. Total testosterone was assayed by LC-MS/MS using a previously published method (26). The analytical inter-assay coefficient of variation (CV) was 4 % in the range 1.5 – 37 nmol/L. Peptide hormones LH, FSH and SHBG were analyzed using the IMMULITE 2000 xpi platform (Siemens Healthcare, Erlangen, Germany). The analytical inter-assay CVs were 7 % at 10 IU/L LH, 5 % at 17 IU/L FSH and 6 % at 60 nmol/L SHBG. Free testosterone index was calculated as the percentage of total testosterone divided by SHBG. Standard international units were converted for testosterone (1 mM = 28.8 ng/dL) and SHBG (1 mM = 9.5 µg/dL).

Reference intervals

Data were processed to conform to published guidelines proposed in the CLSI EP28-A3c guidelines and the CALIPER white paper (3). Reference intervals included a minimum of 40 observations and ideally more than 120 observations where possible. Partitioning was based on clinical considerations in terms of attainment of pubertal testicle volume (≥ 4 ml TV_{OM} by orchidometer, corresponding to ≥ 2.7 ml TV_{US} by ultrasound) marking the definition puberty

onset. None of the 414 healthy participants were removed as outliers from the data set. Serum analyte levels in respective partitions based on age, testicle volume or Tanner PH stages were subjected to the pairwise Harris-Boyd standard deviation test (27). For this purpose, a log transformation was used to obtain a Gaussian approximation when sample data were not intrinsically normally distributed by the Shapiro-Wilk criteria. The Harris-Boyd operation considers sample size (n), mean analyte value (μ) and variance (σ) of consecutive partitions. Pairwise partitions were considered justifiably separated only if the Harris-Boyd z score exceeded the corresponding 'critical' z* which is calculated using the formula: $z^* = 3(n_1+n_2/120)^{1/2}$ (28). The central 95% reference intervals and corresponding 90% confidence limits of the lower limits (LL) and upper limits (UL) were calculated with the nonparametric method when the sample size was 120 or higher or with the robust method when the sample size was between 40 and 120. Nonparametric calculations were based on the binomial distribution of observation ranks, while the robust method was based on (n=500) resampled (bootstrapped) datasets.

Statistical analyses

The nonparametric Mann-Whitney U tests, Kruskal-Wallis one-way analysis of variance (ANOVA) tests and Spearman rho correlations were computed with GraphPad Prism v7 (GraphPad Software, San Diego, CA, USA) and SPSS (IBM Corporation, New York, NY, USA). P-values were not adjusted for multiple comparisons, but the number of pairwise tests is limited since only adjacent partitions were compared. Statistical significance was defined as $p<0.05$ (*), $p<0.01$ (**) or $p<0.001$(***). The *referenceIntervals* package in *R* (R Development Core Team, Vienna, Austria) was used to estimate continuous reference charts based on a moving window of n=40-120 observations with increments of 20 observations. Continuous centiles were calculated from no less than n=40 observations in the dataset tail ends. The 90%

confidence intervals associated with the 2.5th and 97.5th percentiles were calculated by resampling using *referenceIntervals* package in *R* and the Analyse-it (Analyse-it, Leeds, UK) integration in Microsoft Excel (Microsoft, Redmond, WA, USA). Figures were generated using *R* and GraphPad Prism.

Ethical considerations

This study was approved by the Norwegian Regional Ethics Committee (2015/128/REK, 2015/235/REK) and conformed to good clinical practice and the ethical standards of the Helsinki Declaration. In accordance with the regulatory specifications, children under the age of 16 were only examined in cases where both written informed parental consent and child assent was obtained. Written and informed consent was obtained according to regulatory protocol prior to inclusion in the cohort study. Participation in the study was rewarded with a cinema voucher. In cases where non-threatening scrotal pathology was detected by ultrasound, parents were informed and boys were referred to a pediatrician.

RESULTS

Serum hormone levels throughout puberty

Serum concentrations of total testosterone, FSH, LH and SHBG were plotted against the chronological age of each study participant (Figure 1a-d). Table format references from the figure models are provided in Supplemental Table 1 (29). For serum testosterone levels, we calculated a Spearman correlation matrix to resolve associations with regard to both chronological age and TV_{US}. Overall, both chronological age and TV_{US} were strongly associated with each other ($r = 0.864$, $p<0.001$) and with testosterone levels ($r = 0.876$, $p<0.001$ and $r = 0.849$, $p<0.001$, respectively). In prepubertal boys with $TV_{US} < 2.7$ ml, chronological age accounted for more variation in circulating testosterone levels ($r = 0.615$, $p<0.001$) than

TV_{US} (r = 0.420, p<0.001). In contrast, TV_{US} accounted for more variation in testosterone levels (r = 0.753, p<0.001) than age (r = 0.692, p<0.01) in pubertal boys with attained $TV_{US} \geq 2.7$ ml. These findings illustrate the biological relevance of TV during puberty and prompted us to establish an additional set of references for hormone levels in relation to TV.

Serum hormone levels and testicular volume

Serum levels of total testosterone, FSH, LH and SHBG were plotted against a continuous scale of TV determined by ultrasound (Figure 2a-d). The age corresponding to attainment of a mean TV_{US} of 2.7 ml that marks the definition of puberty onset in our cohort was 11.7 years. Reference intervals from Figure 2 are provided in Supplemental Table 2 (29).

Orchidometer references

In order to make our results backwards compatible with traditional Prader orchidometry we also compiled a table to account for hormone levels in terms of TV corresponding to the closest discrete increment on the orchidometer TV_{OM} scale (Table 1). We included the LH/total testosterone ratio as a metric of Leydig cell function to ascertain whether the definition of puberty onset at 4 ml TV_{OM} or 2.7 ml TV_{US} would associate with a trend change of this ratio. Indeed, we observed that this transition was characterized by a unidirectional shift in the LH/Testosterone ratio, implying that serum levels of testosterone increased more rapidly than that of LH from this point of puberty onset. Calculations for free testosterone, unbound by SHBG, were also included.

Serum hormone levels by Tanner PH staging

Lastly, hormone levels were evaluated in relation to the study participants' designated Tanner (PH) stage obtained by visual inspection of pubic hair (Figure 3a-d). Notably, all

analyzed hormones exhibited significant median changes throughout the five Tanner PH stages (Kruskal-Wallis 1-way ANOVA, $p < 0.0001$ for all hormones). With respect to testosterone and SHBG, each Tanner PH stage increment was characterized by a significantly increased median compared to the previous stage (Mann-Whitney test, $p < 0.05$ for paired Tanner stage increments). Mean age (SD) in years ± SD for the indicated Tanner PH stages were: I 9.1 (1.7); II 12.1 (1.2); III 12.7 (0.9); IV 14.2 (1.1) and V 15.0 (0.7).

Cohort partitions and reference intervals

Hormone reference intervals for indicated variables and cohort partitions were organized in Table 2. Indicated partition pairs fulfilled the requirement for the standard deviate score (z^{SD}) to exceed that of the critical sample power test (z^*) based on the actual number of cohort observations in respective partitions, without resampling. Notably, we introduced TV as a covariate for the age-based reference interval 10 - <13 years. This decision was based on the finding that the earliest occurrences of pubertal ultrasound $TV_{US} \geq 2.7$ ml in our dataset were observed from the age of 10 years. Conversely, our data included no observations of boys aged 13 or older that exhibited prepubertal $TV_{US} < 2.7$ ml. We therefore leveraged TV as a binary covariate and partitioned this transition age interval with respect to whether or not the study participants had attained $TV_{US} \geq 2.7$ ml, corresponding to the pediatric definition of pubertal onset. Importantly, these two partitions overlapping in terms of age exhibited significant median differences (Mann-Whitney, $p < 0.001$) for all hormones. As confirmed by the Harris-Boyd standard deviate test, this conditional partitioning of this age interval produced two separate and statistically valid reference intervals. Extrapolated reference intervals for Tanner PH stages were also included. Notably, Tanner PH stages II, III and IV were combined to obtain a statistically valid sample size partition.

DISCUSSION

In this study, we combined ultrasound determination of TV with state-of-the-art LC-MS/MS and immunoassays to respectively quantitate serum levels of testosterone and peptide hormones FSH, LH and SHBG. Our findings document normative endocrine changes associated with puberty progression in terms of both chronological age and gonadal development in a representative sample of contemporary and healthy Norwegian boys. We employed ultrasound to objectively determine testicle dimensions and calculate metric volume. This novel protocol presents an improvement for determining TV compared to traditional palpation and subjective assessments by Prader orchidometry. This is the first study to establish valid reference intervals for pubertal hormones in relation to ultrasound-determined TV. The reference intervals presented in our study are available to supplement existing diagnostic criteria of pediatric endocrinopathies and disorders of sexual development.

Blood samples from healthy children are notoriously hard to obtain in large numbers due to ethical regulations. Accordingly, establishing empirical and valid pediatric reference intervals is a challenge to clinical laboratories. Determination of reference ranges on leftovers of in-house blood samples from children that were previously enrolled in hospital care is not recommended, but this remains common practice. Such population samples are likely to represent individuals that may hamper their suitability to serve as reference material. The data in our cohort is representative for regular school children and the application of exclusion criteria based on a clinical exam and medical history ensured a population reference sample of normal and healthy children. In terms of ethnicity, the current cohort is representative of the Norwegian general population which is comprised of approximately 90% Caucasians. This degree of population homogeneity may limit the generalizability of provided reference intervals in other parts of the world.

Unlike traditional Prader orchidometry, ultrasound analysis allows for direct quantification of the testicle while disregarding the volume contributed by the surrounding epididymis. Orchidometry overestimates TV in smaller testicles where these structures are impossible to distinguish by palpation. Ultrasound provides an assessment of the actual testicle volume on a continuous scale, as opposed to the ordinal scale of the Prader orchidometer. As a result, the two methods produce a readout discrepancy in terms of milliliter (ml) volumes that empirically conform to a non-linear relationship described by the formula $TV_{OM} = 1.96 \times TV_{US}^{0.71}$ (15). This allows for more refined statistical operations and modeling. In addition, ultrasound also has the advantage of being perceived as less personally invasive compared to palpation, while being able to detect testicular lesions (e.g. presence of a hydrocele or varicocele) or scrotal pathologies which may warrant medical attention and also confound the determination of TV (16). In the current study, we calculated ultrasound TV_{US} from testicular height, width and depth multiplied with a constant of 0.71 according to the Lambert equation. It should be noted that some studies used the standard ellipsoid formula with a constant of 0.52, but conversion between both methods is straightforward (15).

A recently published Swedish study investigated the association between Prader orchidometer-resolved TV and serum levels of several androgen hormones including testosterone (19). Unfortunately, the authors did not group orchidometer volumes according to the established practice to distinguish the allocated prepubertal (1 - 3 ml), pubertal (4 - 12 ml) and adult (≥ 15 ml) stages in terms of TV_{OM}. Additionally, the study sample size was insufficient to infer statistically valid hormone reference intervals.

Sertoli cells constitute the majority volume of the adult testicle and our data demonstrates that serum levels of pituitary-derived FSH driving this maturation increase progressively from the age of 6 years. In contrast, testosterone synthesis in Leydig cells is activated by pituitary-derived LH, and we observed only modest increases in LH levels prior to

the age of 10 years. Notably, we observed that the traditional definition of puberty onset at orchidometer TV of 4 ml indeed corresponds to the near maximal LH/testosterone ratio and precedes a unidirectional decline in the ratio, which may represent a critical checkpoint for Leydig cell maturation. Importantly, an elevated LH/testosterone ratio is a hallmark of Klinefelter syndrome (30). Low serum levels of testosterone, LH and FSH are clinically relevant for diagnosing hypogonadotropic conditions including Kallmann syndrome (31). Importantly, circulating testosterone is sequestered by SHBG and, to a lesser extent, albumin. We therefore also present data for calculated free testosterone in relation to TV.

A limitation of the present study is the collection of non-fasting blood samples between 0800 h and 1400 h of the day. It has previously been shown that diurnal variation in testosterone secretion is more pronounced in early and mid-puberty (32). However, the largest temporal changes in serum testosterone levels are observed during the nighttime. Interestingly, we did not observe significant inter-individual differences in terms of serum testosterone levels when comparing prepubertal (TV_{US} < 2.7 ml) boys grouped according in two-hour intervals of blood sample acquisition (Kruskal-Wallis test, $p = 0.11$). In contrast, pubertal boys ($TV_{US} \geq 2.7$ ml) exhibited a significant decrease in testosterone levels throughout the day, unadjusted for age (Kruskal-Wallis, $p < 0.01$). Since our data are primarily based on non-fasting blood samples collected before noon, diurnal hormone variations and feeding effects should be considered when interpreting our data. Additionally, we used the IMMULITE 2000 xpi immunoassay platform for peptide hormone quantification. Although analytical performance of the main proprietary immunoassay platforms exhibit strong consistency, careful evaluation is recommended when implementing references that were established using a different brand platform (33).

With regard to the hormone reference intervals presented in this study, we have aimed to comply with the CLSI EP28-A3c guidelines, the CALIPER white paper and internal

laboratory protocols. Although partitions should ideally be defined by 120 observations, a minimum sample size of 40 is sufficient to estimate a reference interval with robust resampling (34). The Harris-Boyd test was performed on Gaussian distributed transformed data to establish the validity of neighboring reference intervals. When stratifying the boys between 10 and 13 years of age by attainment of a pubertal $TV_{US} \geq 2.7$ ml, all hormones included Table 2 were satisfactorily separated by the Harris-Boyd and z* criteria. This implies that stratifying age with TV as a covariate is statistically warranted and may be of clinical importance.

In conclusion, we have described the association between serum levels of pertinent hormones of the pituitary-gonadal hormone axis and pubertal progression in terms of testicular growth, chronological age and Tanner PH stages. Furthermore, we have established CLSI-compliant reference intervals for serum levels of total testosterone, FSH, LH and SHBG based on a representative population sample of healthy Norwegian children. Pediatric hormone reference intervals may be improved by accounting for TV. By leveraging TV as a covariate, we were able to establish statistically valid partitions to distinguish pubertal states despite overlapping age.

Acknowledgements

The authors wish to thank the children and parents who participated in the BGS2 study. We also thank Hege Skavøy for contributing with arranging the blood sample analyses at the Hormone Laboratory, Haukeland University Hospital. The authors also thank the pediatricians who participated in data collection and particularly radiologist Magnus Sveen for conducting the testicular ultrasound assessments in the BGS2 study. This project was supported by the Western Regional Norwegian Health Authority, grant no. 912221.

References

1. Jung, B, Adeli, K. Clinical laboratory reference intervals in pediatrics: the CALIPER initiative. Clinical biochemistry 2009;42(16-17):1589-1595.
2. Tahmasebi, H, Higgins, V, Fung, AWS, Truong, D, White-Al Habeeb, NMA, Adeli, K. Pediatric Reference Intervals for Biochemical Markers: Gaps and Challenges, Recent National Initiatives and Future Perspectives. EJIFCC 2017;28(1):43-63.
3. Adeli, K, Higgins, V, Trajcevski, K, White-Al Habeeb, N. The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper. Critical reviews in clinical laboratory sciences 2017;54(6):358-413.
4. Konforte, D, Shea, JL, Kyriakopoulou, L, Colantonio, D, Cohen, AH, Shaw, J, Bailey, D, Chan, MK, Armbruster, D, Adeli, K. 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. Clinical chemistry 2013;59(8):1215-1227.
5. Skakkebaek, NE, Rajpert-De Meyts, E, Buck Louis, GM, Toppari, J, Andersson, AM, Eisenberg, ML, Jensen, TK, Jorgensen, N, Swan, SH, Sapra, KJ, Ziebe, S, Priskorn, L, Juul, A. Male Reproductive Disorders and Fertility Trends: Influences of Environment and Genetic Susceptibility. Physiological reviews 2016;96(1):55-97.
6. Clavijo, RI, Hsiao, W. Update on male reproductive endocrinology. Translational andrology and urology 2018;7(Suppl 3):S367-S372.
7. Marshall, WA, Tanner, JM. Variations in the pattern of pubertal changes in boys. Archives of disease in childhood 1970;45(239):13-23.
8. Juul, A, Teilmann, G, Scheike, T, Hertel, NT, Holm, K, Laursen, EM, Main, KM, Skakkebaek, NE. Pubertal development in Danish children: comparison of recent European and US data. International journal of andrology 2006;29(1):247-255; discussion 286-290.
9. Prader, A. Testicular size: assessment and clinical importance. Triangle; the Sandoz journal of medical science 1966;7(6):240-243.
10. Behre, HM, Nashan, D, Nieschlag, E. Objective measurement of testicular volume by ultrasonography: evaluation of the technique and comparison with orchidometer estimates. International journal of andrology 1989;12(6):395-403.
11. Carlsen, E, Andersen, AG, Buchreitz, L, Jorgensen, N, Magnus, O, Matulevicuus, V, Nermoen, I, Petersen, JH, Punab, M, Suominen, J, Zilaitiene, B, Giwercman, A. Inter-observer variation in the results of the clinical andrological examination including estimation of testicular size. Int J Androl 2000;23(4):248-253.
12. Sakamoto, H, Saito, K, Oohta, M, Inoue, K, Ogawa, Y, Yoshida, H. Testicular volume measurement: comparison of ultrasonography, orchidometry, and water displacement. Urology 2007;69(1):152-157.
13. Goede, J, Hack, WW, Sijstermans, K, van der Voort-Doedens, LM, Van der Ploeg, T, Meij-de Vries, A, Delemarre-van de Waal, HA. 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.
14. Joustra, SD, van der Plas, EM, Goede, J, Oostdijk, W, Delemarre-van de Waal, HA, Hack, WW, van Buuren, S, Wit, JM. New reference charts for testicular volume in Dutch children and adolescents allow the calculation of standard deviation scores. Acta paediatrica 2015;104(6):e271-278.
15. Oehme, NHB, Roelants, M, Bruserud, IS, Eide, GE, Bjerknes, R, Rosendahl, K, Juliusson, PB. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys - intra- and interobserver agreement and comparison with Prader orchidometry. Pediatr Radiol 2018;48(12):1771-1778.

16. Diamond, DA, Paltiel, HJ, DiCanzio, J, Zurakowski, D, Bauer, SB, Atala, A, Ephraim, PL, Grant, R, Retik, AB. Comparative assessment of pediatric testicular volume: orchidometer versus ultrasound. J Urol 2000;164(3 Pt 2):1111-1114.
17. Andersson, AM, Juul, A, Petersen, JH, Muller, J, Groome, NP, Skakkebaek, NE. Serum inhibin B in healthy pubertal and adolescent boys: relation to age, stage of puberty, and follicle-stimulating hormone, luteinizing hormone, testosterone, and estradiol levels. The Journal of clinical endocrinology and metabolism 1997;82(12):3976-3981.
18. Soeborg, T, Frederiksen, H, Mouritsen, A, Johannsen, TH, Main, KM, Jorgensen, N, Petersen, JH, Andersson, AM, Juul, A. Sex, age, pubertal development and use of oral contraceptives in relation to serum concentrations of DHEA, DHEAS, 17alpha-hydroxyprogesterone, Delta4-androstenedione, testosterone and their ratios in children, adolescents and young adults. Clinica chimica acta; international journal of clinical chemistry 2014;437:6-13.
19. Ankarberg-Lindgren, C, Dahlgren, J, Andersson, MX. High-sensitivity quantification of serum androstenedione, testosterone, dihydrotestosterone, estrone and estradiol by gas chromatography-tandem mass spectrometry with sex- and puberty-specific reference intervals. The Journal of steroid biochemistry and molecular biology 2018;183:116-124.
20. 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):263-270.
21. 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):e932-939.
22. Waaler, PE, Thorsen, T, Stoa, KF, Aarskog, D. Studies in normal male puberty. Acta paediatrica Scandinavica Supplement 1974;(249):1-36.
23. Rasmussen, AR, Wohlfahrt-Veje, C, Tefre de Renzy-Martin, K, Hagen, CP, Tinggaard, J, Mouritsen, A, Mieritz, MG, Main, KM. Validity of self-assessment of pubertal maturation. Pediatrics 2015;135(1):86-93.
24. Bahk, JY, Jung, JH, Jin, LM, Min, SK. Cut-off Value of Testes Volume in Young Adults and Correlation Among Testes Volume, Body Mass Index, Hormonal Level, and Seminal Profiles. Urology 2010;75(6):1318-1323.
25. Tatsunami, S, Matsumiya, KI, Tsujimura, A, Itoh, N, Sasao, T, Koh, E, Maeda, Y, Eguchi, J, Takehara, K, Nishida, T, Miyano, S, Tabata, C, Iwamoto, T. Inter/intra investigator variation in orchidometric measurements of testicular volume by ten investigators from five institutions. Asian Journal of Andrology 2006;8(3):373-378.
26. Methlie, P, Hustad, SS, Kellmann, R, Almas, B, Erichsen, MM, Husebye, E, Lovas, K. Multisteroid LC-MS/MS assay for glucocorticoids and androgens, and its application in Addison's disease. Endocrine connections 2013;2(3):125-136.
27. Harris, EK, Boyd, JC. On dividing reference data into subgroups to produce separate reference ranges. Clinical chemistry 1990;36(2):265-270.
28. Horn, PS, Pesce, AJ. Reference intervals: an update. Clinica chimica acta; international journal of clinical chemistry 2003;334(1-2):5-23.
29. Madsen A, Oehme NB, Roelants M, Bruserud IS, Eide GE, Viste K, Bjerknes R, Almås B, Rosendahl K, Sagen JV, Mellgren G, Juliusson PB. Figshare Digital Repository 2019. Deposited 11 September 2019. https://doi.org/10.6084/m9.figshare.9798248.
30. Aksglaede, L, Skakkebaek, NE, Almstrup, K, Juul, A. Clinical and biological parameters in 166 boys, adolescents and adults with nonmosaic Klinefelter syndrome: a Copenhagen experience. Acta paediatrica 2011;100(6):793-806.

31. John, H, Schmid, C. Kallmann's syndrome: clues to clinical diagnosis. International journal of impotence research 2000;12(5):269-271.
32. Ankarberg-Lindgren, C, Norjavaara, E. Changes of diurnal rhythm and levels of total and free testosterone secretion from pre to late puberty in boys: testis size of 3 ml is a transition stage to puberty. European journal of endocrinology 2004;151(6):747-757.
33. Radicioni, A, Lenzi, A, Spaziani, M, Anzuini, A, Ruga, G, Papi, G, Raimondo, M, Foresta, C. A multicenter evaluation of immunoassays for follicle-stimulating hormone, luteinizing hormone and testosterone: concordance, imprecision and reference values. Journal of endocrinological investigation 2013;36(9):739-744.
34. Horn, PS, Pesce, AJ, Copeland, BE. A robust approach to reference interval estimation and evaluation. Clinical chemistry 1998;44(3):622-631.

TABLE LEGENDS

Table 1. Cohort hormone levels in relation to traditional orchidometry scale.

Ultrasound testicular volume (TV_{US}) was converted to the orchidometer TV_{OM} scale using the non-linear formula outlined in the methods section. For each discrete increment on the orchidometer TV_{OM} scale, the equivalent TV_{US} is provided. Cohort participants were assigned to the closest TV_{OM} ordinal. Empirical hormone levels corresponding to the median (p50), along

with the 2.5th (p2.5) and 97.5th (p97.5) percentiles are provided. Calculations for LH/testosterone (T) ratio and free testosterone (T) index were also included. Observations where TV_{OM} was over 15 ml were discarded due to small sample size. Abbreviations: TV, testicular volume; SD, standard deviation; FSH, follicle-stimulating hormone; SHBG, sex hormone-binding globulin; T, testosterone.

Table 2. Hormone reference intervals extrapolated from cohort data.

Reference intervals for indicated hormones were established with respect to chronological age, ultrasound testicular volume (TV_{US}) and Tanner PH stages. Sample size (n) is specified for each partition and the standard international (SI) unit is denoted for each analyte. TV_{US} was used as a partitioning variable with age or independently to separate prepubertal (TV_{US} < 2.7 ml) and pubertal (TV_{US} 2.7 - 12.8 ml) individuals. Empirical cohort median (p50) hormone levels are presented for each partition. Significant changes in median analyte levels compared to the previous partition above are indicated **p<0.01 or ***p<0.001 (pairwise Mann-Whitney U test). Hormone reference intervals were inferred from resampled datasets based on the population sample. Intervals were nonparametrically defined by the 2.5th percentile lower limit (LL) and 97.5th percentile upper limit (UL). Lower and upper partition limits are presented with respective 90% confidence intervals in parentheses. Using the Harris-Boyd standard deviate test, each partition was tested pairwise with the pertinent and previous partition above it. Where the current and the previous partitions were justifiably separated, the z variable is indicated †. Abbreviations: TV, testicular volume by ultrasound; SI unit, standard international unit; LL, lower limit; UL, upper limit; CI, confidence interval; z, Harris-Boyd standard deviate test; Tanner PH, pubic hair development stage.

FIGURE LEGENDS

Figure 1. Circulating hormone levels in relation to chronological age.

Serum levels of (a) total testosterone, (b) FSH, (c) LH and (d) SHBG from n=414 boys plotted against participant age. Continuous reference interval centiles indicating the moving average median (p50; dashed red line), lower limit (p2.5; lower black line) and upper limit (p97.5; upper black line) were estimated by nonparametric method in areas of n=120 observations and by the robust resampling method when considering n=40-120 observations in the tail ends. The 90% confidence intervals associated with the respective centiles are indicated as shaded area.

Figure 2. Circulating hormone levels in relation to ultrasound testicular volume.

Serum levels of (a) total testosterone, (b) FSH, (c) LH and (d) SHBG from n=409 boys plotted against ultrasound-determined TV_{US}. Continuous reference interval centiles indicate the moving average median (p50; dashed red line), lower limit (p2.5; lower black line) and upper limit (p97.5; upper black line) with shaded 90% confidence intervals. The vertical demarcation line indicates the cut-off for attainment of pubertal TV.

Figure 3. Circulating hormone levels in relation to Tanner PH stages.

Serum levels of (a) total testosterone, (b) FSH, (c) LH and (d) SHBG in n=412 boys in relation to visual determination of participant Tanner pubic hair stages I - V. Median hormone levels associated with each stage is indicated by the red line. The number of observations included for the different Tanner PH stages were I (n=249), II (n=33), III (n=29), IV (n=28) and V (n=64).

Data availability

Restrictions apply to the availability of data generated or analyzed during this study to preserve patient confidentiality or because they were used under license. The corresponding author will on request detail the restrictions and any conditions under which access to some data may be provided.

Figure 1

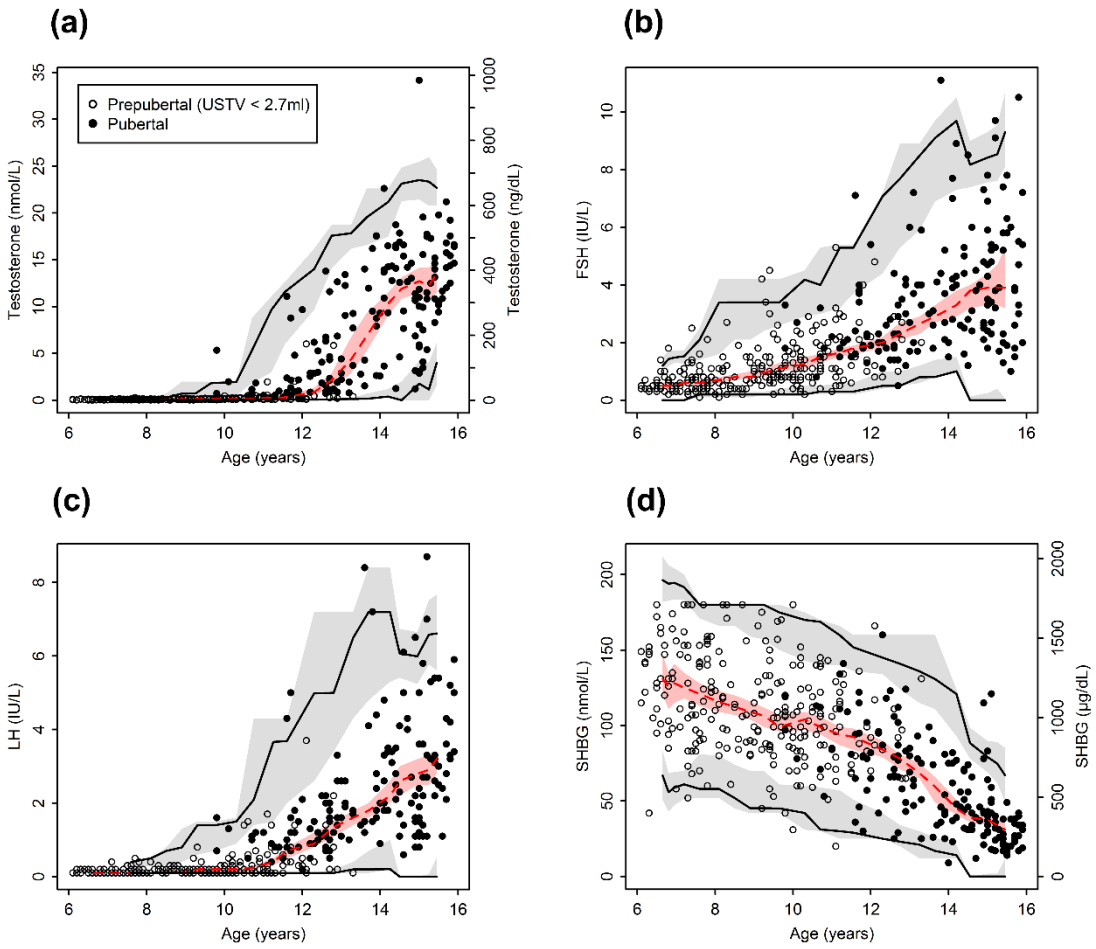

Figure 2

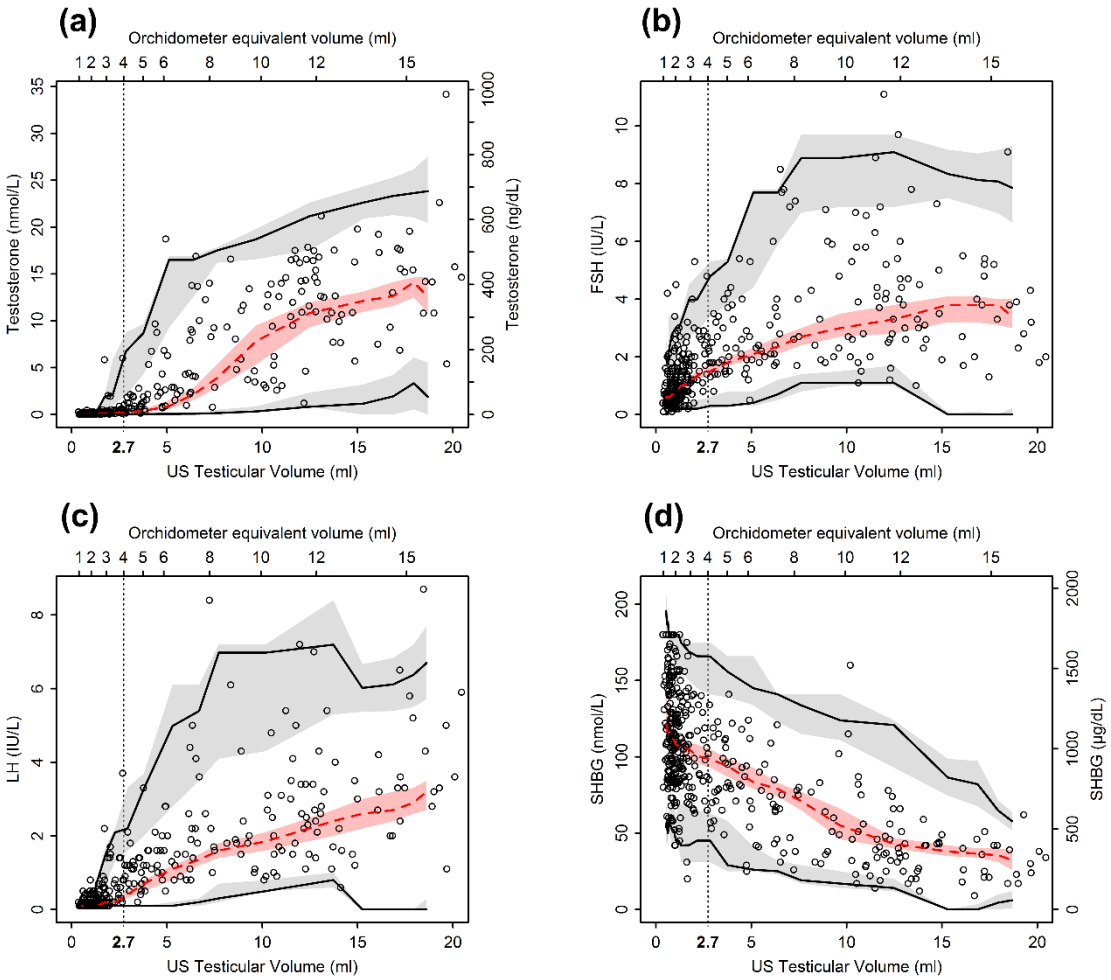

Figure 3

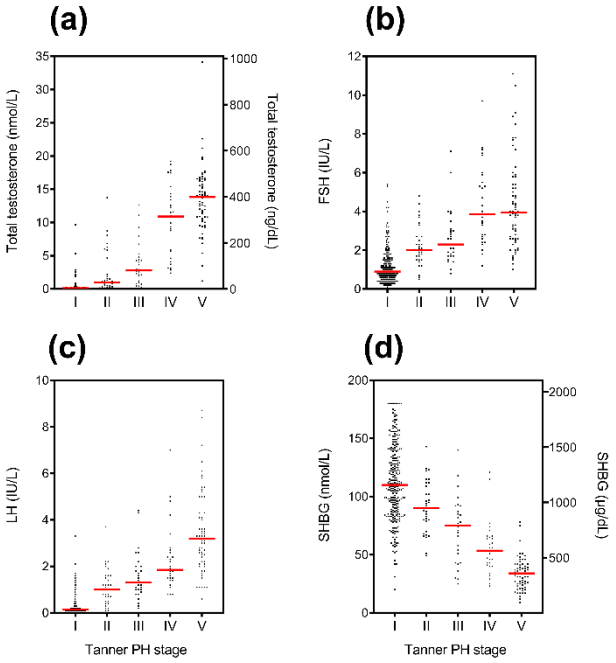

Table 1. Cohort hormone levels in relation to traditional orchidometry scale.

				Hormone levels: p50 (p2.5 - p97.5)					
TV_{OM} (ml)	TV_{US} (ml)	n	Age (SD)	Testosterone (nmol/L)	FSH (IU/L)	LH (IU/L)	SHBG (nmol/L)	LH/T ratio	Free T index
1	0.4	37	8.1 (1.4)	0.09 (0.02 - 0.25)	0.6 (0.2 - 2.3)	0.1 (0.1 - 0.3)	120 (61 - 180)	1.75 (0.49 - 8.26)	0.07 (0.01 - 0.25)
2	1.0	150	8.7 (1.6)	0.10 (0.02 - 0.26)	0.8 (0.2 - 2.8)	0.1 (0.1 - 0.4)	111 (52 - 180)	1.51 (0.46 - 7.78)	0.09 (0.02 - 0.29)
3	1.8	46	10.3 (1.5)	0.17 (0.04 - 2.01)	1.5 (0.3 - 3.9)	0.3 (0.1 - 1.7)	104 (35 - 165)	1.53 (0.39 - 8.49)	0.17 (0.03 - 3.01)
4	2.7	22	11.3 (1.2)	0.46 (0.10 - 3.87)	2.1 (0.8 - 4.1)	0.8 (0.2 - 2.9)	88 (55- 137)	1.54 (0.64 - 5.03)	0.60 (0.10 - 5,02)
5	3.7	20	12.0 (0.9)	1.13 (0.21 - 6.07)	2.3 (1.4 - 4.3)	1.1 (0.3 - 2.5)	96 (57 - 132)	0.80 (0.23 - 2.31)	1.39 (0.25 - 7.59)
6	4.8	19	13.1 (1.1)	2.87 (0.76 - 14.6)	2.2 (0.8 - 5.4)	1.5 (0.6 - 2.8)	81 (27 - 121)	0.45 (0.16 - 1.17)	4.87 (0.84 - 21.2)
8	7.2	21	13.4 (1.1)	8.79 (0.89 - 16.7)	3.4 (1.7 - 8.2)	2.2 (0.8 - 7.3)	66 (29 - 115)	0.36 (0.15 - 1.04)	14.1 (0.96 - 54.2)
10	9.9	23	14.1 (1.4)	6.18 (2.74 - 14.3)	2.8 (1.3 - 7.0)	1.8 (0.9 - 5.1)	51 (20 - 135)	0.25 (0.14 - 0.53)	10.6 (2.31 - 63.2)
12	12.8	37	14.8 (0.9)	12.9 (4.36 - 19.9)	4.0 (1.2 - 9.8)	2.6 (1.1 - 7.0)	40 (14 - 74)	0.21 (0.09 - 0.66)	36.1 (9.39 - 88.8)
15	17.6	22	15.3 (0.6)	14.9 (6.17 - 28.1)	3.6 (1.5 - 7.2)	3.5 (1.6 - 7.5)	32 (13 - 56)	0.22 (0.14 - 0.87)	44.8 (25.4 - 108)

Table 2. Hormone reference intervals extrapolated from cohort data.

Hormone	Reference	Partition	Covariate	n	SI unit	p50	p	Reference intervals		z
								LL p2.5 (90% CI)	UL p97.5 (90% CI)	
Testosterone	Age	06 - < 10 yrs		169	nmol/L	0.09	-	0.02 (0.01 - 0.02)	0.28 (0.19 - 0.72)	-
		10 - < 13 yrs		137	nmol/L	0.28	***	0.07 (0.06 - 0.08)	10.7 (6.85 - 13.2)	†
		13 - < 16 yrs		108	nmol/L	11.6	***	1.02 (0.39 - 1.95)	22.3 (18.9 - 34.2)	†
	Age w/ TV_{US}	10 - < 13 yrs	TV < 2.7 ml	84	nmol/L	0.17	-	0.06 (0.06 - 0.08)	3.42 (0.84 - 6.01)	-
		10 - < 13 yrs	TV ≥ 2.7 ml	52	nmol/L	1.87	***	0.16 (0.14 - 0.27)	12.8 (11.1 - 13.8)	†
	TV_{US}	< 2.7 ml	*Prepubertal*	247	nmol/L	0.11	-	0.02 (0.02 - 0.03)	0.77 (0.45 - 2.01)	-
		2.7 - 12.8 ml	*Pubertal*	115	nmol/L	4.32	***	0.24 (0.14 - 0.48)	17.5 (16.6 - 18.7)	†
	Tanner PH	I		249	nmol/L	0.11	-	0.02 (0.02 - 0.03)	2.17 (1.54 - 2.89)	-
		II / III / IV		90	nmol/L	3.11	***	0.11 (0.07 - 0.17)	18.1 (16.5 - 19.2)	†
		V		64	nmol/L	13.8	***	2.81 (1.21 - 6.86)	25.9 (20.2 - 34.2)	†
FSH	Age	06 - < 10 yrs		169	IU/L	0.7	-	0.2 (0.1 - 0.2)	3.1 (2.4 - 4.0)	-
		10 - < 13 yrs		137	IU/L	1.7	***	0.3 (0.2 - 0.4)	4.9 (4.0 - 6.3)	†
		13 - < 16 yrs		108	IU/L	3.8	***	1.3 (1.0 - 1.6)	9.7 (8.5 - 11.1)	†
	Age w/ TV_{US}	10 - < 13 yrs	TV < 2.7 ml	84	IU/L	1.4	-	0.2 (0.2 - 0.4)	4.3 (3.0 - 5.3)	-
		10 - < 13 yrs	TV ≥ 2.7 ml	52	IU/L	2.1	***	0.7 (0.5 - 1.2)	6.0 (4.3 - 7.1)	*
	TV_{US}	< 2.7 ml	*Prepubertal*	247	IU/L	0.8	-	0.2 (0.2 - 0.2)	3.4 (2.9 - 4.4)	-
		2.7 - 12.8 ml	*Pubertal*	115	IU/L	2.7	***	1.1 (0.5 - 1.3)	8.9 (7.4 - 11.1)	†
	Tanner PH	I		249	IU/L	0.9	-	0.2 (0.2 - 0.3)	3.7 (3.1 - 4.4)	-
		II / III / IV		90	IU/L	2.6	***	0.7 (0.5 - 1.1)	7.6 (6.7 - 9.7)	†
		V		64	IU/L	4.0	***	1.3 (1.0 - 1.6)	10.2 (8.6 - 11.1)	†
LH	Age	06 - < 10 yrs		168	IU/L	0.1	-	≤ 0.1 (n/a)	0.6 (0.4 - 0.8)	-
		10 - < 13 yrs		135	IU/L	0.5	***	≤ 0.1 (n/a)	3.3 (2.2 - 4.7)	†
		13 - < 16 yrs		107	IU/L	2.5	***	0.6 (0.1 - 0.97)	7.5 (6.1 - 8.7)	†
	Age w/ TV_{US}	10 - < 13 yrs	TV < 2.7 ml	83	IU/L	0.2	-	≤ 0.1 (n/a)	2.1 (1.4 - 3.7)	-
		10 - < 13 yrs	TV ≥ 2.7 ml	51	IU/L	1.2	***	0.3 (0.2 - 0.5)	4.2 (2.6 - 5.0)	†
	TV_{US}	< 2.7 ml	*Prepubertal*	245	IU/L	0.1	-	≤ 0.1 (n/a)	1.3 (0.8 - 1.7)	-
		2.7 - 12.8 ml	*Pubertal*	114	IU/L	1.6	***	0.4 (0.2 - 0.7)	6.6 (5.0 - 8.4)	†
	Tanner PH	I		246	IU/L	0.2	-	≤ 0.1 (n/a)	1.5 (1.3 - 1.7)	-
		II / III / IV		90	IU/L	1.4	***	0.1 (0.1 - 0.2)	5.1 (4.3 - 7.0)	†
		V		63	IU/L	3.2	***	0.9 (0.6 - 1.1)	8.1 (6.8 - 8.7)	†
SHBG	Age	06 - < 10 yrs		169	nmol/L	114	-	49 (42 - 58)	≥ 180 (n/a)	-
		10 - < 13 yrs		137	nmol/L	92	***	29 (22 - 39)	155 (142 - 174)	†
		13 - < 16 yrs		108	nmol/L	40	***	13 (9 - 17)	110 (87 - 131)	†
	Age w/ TV_{US}	10 - < 13 yrs	TV < 2.7 ml	84	nmol/L	96	-	35 (20 - 60)	162 (145 - 180)	-
		10 - < 13 yrs	TV ≥ 2.7 ml	52	nmol/L	81	**	27 (25 - 33)	148 (130 - 160)	†
	TV_{US}	< 2.7 ml	*Prepubertal*	247	nmol/L	109	-	50 (42 - 59)	179 (174 - 180)	-
		2.7 - 12.8 ml	*Pubertal*	115	nmol/L	67	***	19 (14 -25)	133 (121 - 160)	†
	Tanner PH	I		249	nmol/L	110	-	50 (42 - 58)	179 (174 - 180)	-
		II / III / IV		90	nmol/L	73	***	25 (23 - 29)	132 (121 - 143)	†
		V		64	nmol/L	34	***	11 (9 - 16)	70 (55 - 78)	†