Glandular breast tissue volume by magnetic resonance imaging in 100 healthy peripubertal girls: evaluation of clinical Tanner staging

Louise Fugl[1], Casper P. Hagen[1], Mikkel G. Mieritz[1], Jeanette Tinggaard[1], Eva Fallentin[2], Katharina M. Main[1] and Anders Juul[1]

BACKGROUND: Appearance of glandular breast tissue may be difficult to distinguish from fat tissue by palpation, especially in obese girls. To our knowledge, validation of the clinical assessment of pubertal breast stages by magnetic resonance imaging (MRI) has never been performed. Our objective was to report normative data of glandular breast tissue volume and validate the clinical evaluation of pubertal breast staging by MRI of breast tissue and to evaluate circulating reproductive hormone levels and estrogen-dependent transabdominal ultrasound (TAUS) parameters as markers of glandular breast tissue.

METHODS: Glandular breast tissue volume quantified by MRI and breast stage evaluation was performed in 100 healthy peripubertal girls. Circulating levels of follicle-stimulating hormone (FSH), luteinizing hormone (LH), inhibin B, and estradiol were measured by immunoassays. Ovarian volume, uterine volume, and endometrial thickness were assessed by TAUS.

RESULTS: Glandular breast tissue volume was positively associated with Tanner stages ($r = 0.858$, $P < 0.001$). The sensitivity and specificity of breast palpation to detect presence of glandular breast tissue using MRI as gold standard were 96 and 95%, respectively. The best parameters to distinguish prepubertal girls from girls with breast development were: LH (area under the curve (AUC) by receiver operating characteristic analysis = 0.871), inhibin B (AUC = 0.847) and estradiol (AUC = 0.830).

CONCLUSION: Clinical palpation reliably detects the presence of glandular breast tissue.

Since the late 1960's, pubertal breast development has been based on Tanner's classification defining pubertal onset as the first appearance of glandular breast tissue, Tanner stage B2 (1). The original description of breast stages was a visual description of changes based on photographs. Although additional palpation is usually performed, the first appearance of glandular breast tissue may be difficult to distinguish from the surrounding subcutaneous fat tissue, especially in obese girls. Alternative methods to evaluate pubertal development, such as self-reporting, have been questionable (2,3). To our knowledge, validation of the clinical assessment of pubertal stages by comparison with quantitative imaging techniques like magnetic resonance imaging (MRI) has never been performed. Especially, the distinction between prepubertal stage (B1) and the onset of puberty (B2) is crucial in the clinical setting when evaluating girls with precocious and delayed puberty.

The primary aim of the present study was to report normative data of glandular breast tissue volume and to validate the clinical assessment of pubertal breast staging against MRI of breast tissue. Subsequently, we assessed how well basal circulating reproductive hormone levels and estrogen-dependent parameters (ovarian volume, endometrial thickness, and uterine volume determined by transabdominal ultrasound) distinguished between absence and presence of glandular breast tissue.

RESULTS

Clinical Breast Staging (Tanner) vs. MRI Quantification of Glandular Volume

Glandular breast tissue volume quantified by MRI increased with increasing Tanner breast stage. Tanner stages were highly associated with MRI volume, $r = 0.858$, $P < 0.001$ (**Figure 1**). The median volume (range) was B1: 0.0 ml (0.0–0.3)), B2: 3.8 ml (0.0–111.8), B3: 33.0 ml (4.2–106.8), B4+5: 74.2 ml (19.7–241.5) (**Figure 1**). Marked interindividual variation occurred in late puberty (B4+B5), interindividual coefficients of variation (CV) = 50.6%. Glandular breast tissue (MRI) was present in 52 of 72 breasts in girls classified as B1 or B2. In the clinical assessment, only one false positive and two false negatives occurred compared to MRI. Thus, the sensitivity and specificity for breast inspection and palpation to detect the presence of glandular breast tissue (MRI) were 50/52 = 96% and 19/20 = 95%, respectively. The positive predictive value and negative predictive value was 50/51 = 98% and 19/21 = 91%, respectively.

Examples of girls with low and high BMI suggest that Tanner staging is based on the combined volume of both glandular and fat tissue (**Figure 2**). Thus, the girls with high BMI seemed to have a smaller amount of actual glandular tissue in the

estimated glandular breast tissue volume (MRI cone volume) than the girls with lower BMI in the corresponding breast stage. Stratifying the girls in tertiles of BMI (**Figure 3a**) and body fat percentage (**Figure 3b**) revealed a tendency toward an inverted U-shaped association with glandular breast tissue volume in girls in mid-puberty (B3) as well as late puberty (B4+5); i.e., B3 total body fat percentage medians were 60.4 (19–192) ml in tertile 1, 88.8 (30–195) ml in tertile 2 and 25.0

(11–92) ml in tertile 3. Kruskal-Wallis *N*-test showed a significant difference between the tertiles ($P = 0.025$).

Circulating Reproductive Hormone Levels and Estrogen-Related Parameters as Predictors of Absence or Presence of Glandular Breast Tissue

All the tested hormones (luteinizing hormone (LH), inhibin B, estradiol and follicle-stimulating hormone (FSH)) showed a good ability to distintiguish B1 from B2 girls, based on the AUC (>0.80). The transabdominal ultrasound (TAUS) parameters (uterine volume, ovarian volume, and endometrial thickness) were less able to predict thelarche (AUC < 0.80). The predictors with the highest AUC were LH and inhibin B. LH concentration above 0.125 IU/l predicted the onset of puberty with 79% sensitivity, 86% specificity and a PPV and NPV of 92 and 67%, respectively (see **Table 1**).

DISCUSSION

In this comprehensive study of 100 healthy girls, we provide normative data on glandular breast tissue volume measured by MRI across breast development stages Tanner B1 through B5. To our knowledge, we are the first to validate clinical staging of breast development by MRI. Our data show that clinical evaluation by inspection and palpation is highly sensitive and specific to detect true glandular breast tissue, which thus validates a long-standing clinical practice. In addition, basal serum levels of LH and inhibin B as well as estradiol were strongly associated with the presence or absence of glandular breast tissue in peripubertal girls.
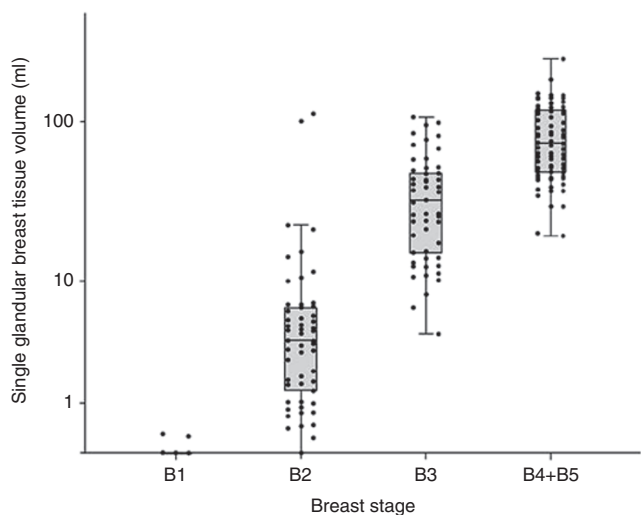
Figure 1. Volume of glandular breast tissue (magnetic resonance imaging) in each breast grouped by Tanner stage of breast development, Spearman's rho 0.858, $P < 0.001$. Boxplots show medians and interquartile ranges (IQR), whiskers show range excluding outliers. The shaded area covers the IQRs above and below the medians.
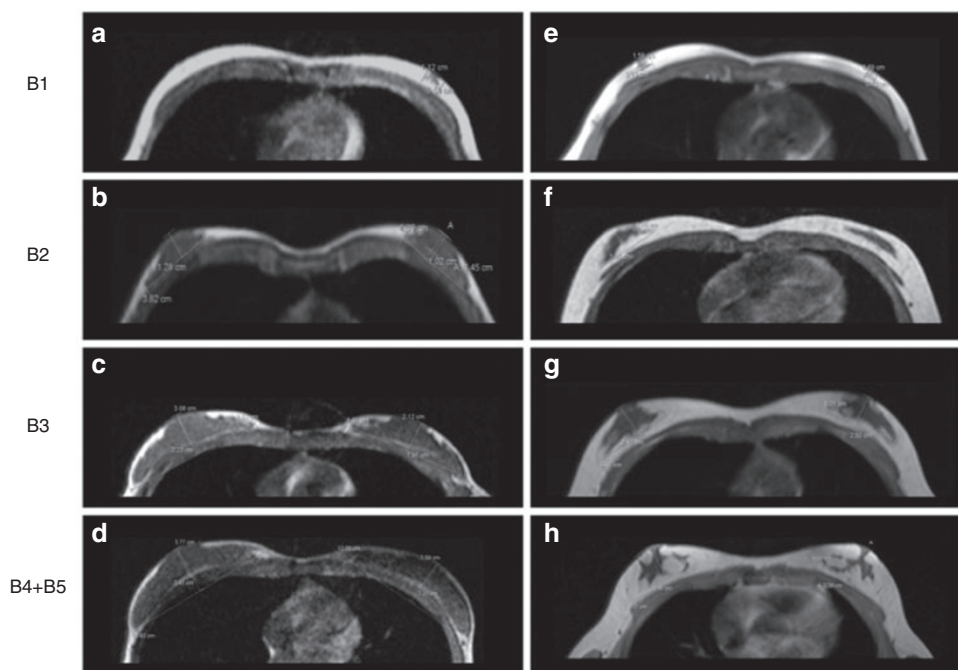
Figure 2. Representative examples of magnetic resonance imaging (transverse chest projection at the level of the nipple) showing the breast volume in girls (B1-B4+5, respectively) with low-normal BMI (left panel: **a–d**) and high BMI (right panel: **e–h**). The low-normal BMI category ranged from 14.7–16.6 (Z-score: −2.28; −0.66) and the high BMI category ranged from 20.1–24.5 (Z-score: 0.92–2.26). Fat tissue is projected as white tissue, glandular tissue as gray.
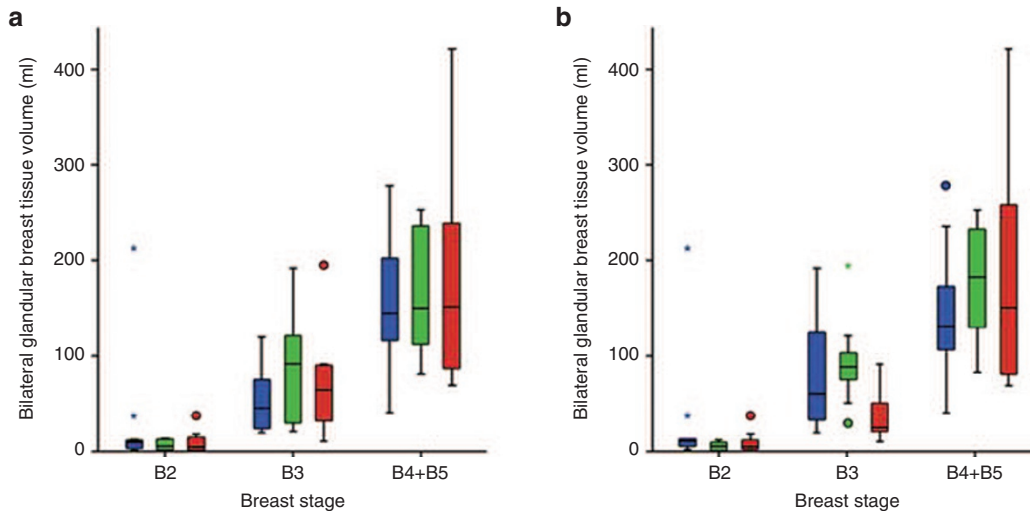
Figure 3. Boxplots showing the association between Tanner stage of breast development (B2, B3, B4+5) and the sum of bilateral glandular breast tissue volume (measured by magnetic resonance imaging) stratified by (**a**) BMI tertiles and (**b**) total body fat percent tertiles (blue = lowest tertiles, green = middle tertiles and red = highest tertiles). The boxplot represents the median and whiskers show the range excluding outliers, dots represent outliers, and stars represent extreme outliers.

Table 1. The ability of hormone and ultrasound parameters to distinguish between B1 vs. B2 (ranked by AUC)

Parameter	Unit	N	N(B1)	N(B2)	AUC	Cut-off	Sensitivity	Specificity	PPV	NPV
LH	IU/l	43	14	29	0.871	0.125	0.79	0.86	0.92	0.67
Inhibin B	pg/ml	43	14	29	0.847	28.5	0.86	0.79	0.89	0.73
Estradiol	pmol/l	43	14	29	0.830	43	0.76	0.93	0.96	0.65
FSH	IU/l	43	14	29	0.813	2.24	0.79	0.79	0.88	0.64
Uterine volume	ml	41	12	29	0.769	6.99	0.67	0.75	0.86	0.47
Ovarian volume	ml	42	13	29	0.722	7.13	0.67	0.75	0.86	0.47
Endometrial thickness	mm	40	12	28	0.696	1.05	0.67	0.67	0.82	0.47

AUC, area under the curve based on the receiver operating characteristic (ROC) analysis. Cut-off: indicates the threshold value where the sum of sensitivity and specificity for distinguishing B1 from B2 girls is highest—corresponding to the point of the ROC curve that is furthest from the line of no-discrimination.
FSH, follicle-stimulating hormone; LH, luteinizing hormone; NPV, negative predictive value; PPV, positive predictive value.

Previously, we reported that the median age at onset of breast development in healthy Danish girls declined from 10.9 to 9.9 y from 1991 to 2006 (ref. (4)). Similar findings have been reported in other regions, e.g., North America (5–7) and Thailand (8). These reports relied on physical examination including Tanner staging by palpation of breast development. Parallel to the decline in age of pubertal onset, BMI has increased in many children (5,9,10). Thus, criticism of the reports on declining age of pubertal onset has centred on methodology, speculating that the apparent early breast development may result from difficulties to distinguish between fat tissue and glandular breast tissue (11). The present study supports that palpation of the breast is in fact a sensitive and specific method to detect glandular tissue. However, generally, the girls in our study were comparatively lean and we are therefore not able to evaluate if the risk of misclassification increases within higher BMI categories.

In the present study, MRI of the breast tissue in the girls with highest BMI showed a substantial amount of fat diffusely scattered between and around the glandular tissue. Furthermore, subanalyses stratified by Tanner stages and tertiles of body fat percentage revealed that the volume of glandular tissue was lowest in girls with most body fat. This indicates that the clinical examination cannot distinguish between glandular tissue and fat in mid to late puberty which may result in overestimation of Tanner stages in girls with higher BMI. Data indicated an inverted U-shaped association between body fat and glandular tissue. We speculate that reduced glandular tissue in girls with very little body fat may be caused by less activity of peripheral aromatase.

Additionally, we found that basal measurements of serum LH, inhibin b, and estradiol were the best parameters to distinguish between B1 and B2 girls. The estimated LH cut-off value (0.125 IU/l) in our present study is similar to a previously found cut-off value (0.1 IU/l) for prediction of a positive (pubertal) GnRH-test (12). To our knowledge, no literature exists containing comparable estradiol or inhibin B cut-off values. We propose, that the above-mentioned parameters may be helpful in clinical cases where the absence or presence of breast tissue is difficult to distinguish.

A limitation of our study is the low prevalence of overweight girls (2/36 in Tanner stage B1 and B2), which is a likely contributing factor to the very high accuracy of clinical assessment in this sample.

By assessing healthy peripubertal girls using MRI of the breasts, we have applied a novel approach with which we have validated a common and important clinical examination used worldwide. However, in subanalysis, the study of a larger study population of overweight/obese girls is needed to demonstrate the impact of high BMI on glandular tissue volume.

Conclusion

Clinical inspection and palpation reliably detects glandular breast tissue in healthy girls. In addition, basal LH and inhibin B as well as estradiol were valuable parameters to distinguish prepubertal girls from girls with pubertal breast development.

METHODS

Participants

A total of 129 healthy girls (9.8–14.7) from The Copenhagen Mother-Child cohort were invited to participate in the study. From 1997–2001, 2,688 Danish women were consecutively included at their first routine obstetric visit at three hospitals in Copenhagen (participation rate 22%). The children from this cohort were examined at several time points during infancy and childhood, and 1,293 peripubertal children (584 girls) agreed to participate in an ongoing longitudinal follow-up study with annual examinations (participation rate 43%). Detailed information on the original cohort has previously been published (13,14). A nested cohort in the present cross sectional study was selected to undergo an MRI study with several outcomes (breast tissue, abdominal fat mass, adrenal glands, internal genitalia) of which some previous results have been published (15,16). For the nested cohort, the selection criterion was high attendance rate at previous examinations. Of the 129 girls invited to participate and 121 consented and underwent chest MRI ($n = 106$) and pelvic ultrasound examination ($n = 112$). In the present study, we focused on chest MRI and used TAUS to measure the size of the internal genitalia, since the latter is available in most departments and clinics. For the main analyses MRI data and Tanner staging were available in 100 girls. In the subsequent analysis, we compared pubertal staging by either MRI or Tanner with serum hormone levels ($n = 112$) and ovarian volume ($n = 103$), endometrial thickness ($n = 110$), and uterine volume ($n = 112$), respectively.

The number of overweight girls compared to number of girls in each puberty stage was 1 out of 10 girls in B1 (10%), 1 out of 26 girls in B2 (4%), 1 out of 27 girls in B3 (4%), and 6 out of 37 girls in B4-B5 (16%).

Clinical Examination

On the day of the ultrasound and MRI scan, a thorough clinical examination was performed. Height was measured using a wall-mounted stadiometer to the nearest mm (Holtain, Crymych, UK) and weight was measured to the nearest 0.1 kg using electronic scales (SECA delta model 707, Seca, Hamburg, Germany, and Bisco model PERS 200, Bisco, Farum, Denmark). BMI was calculated as body weight (kg)/ height (m²). Skinfolds (SF) were measured at four anatomical sites (triceps, subscapular, suprailiac, and biceps) with a Harpenden caliper (John Bull, British Indicators LTD, Weybridge, UK) with a precision of 0.1 mm after allowing the jaws to close for approximately 2 s, as previously described (17). SF % body fat was calculated using Slaughter's equation (18) for girls: (1.33 (triceps + subscapular) – 0.013(triceps + subscapular)² – 2.5) when sum of triceps and subscapular SF was <35 mm and (0.546 (triceps + subscapular) + 9.7) when the sum was >35 mm.

The stage of pubertal development was evaluated according to Tanner and Marshall (inspection and palpation of breast tissue). Three girls declined pubertal examination. Clinical hyperandrogenism was

not systematically evaluated, however, no girls presented with severe hirsutism or acne. In a questionnaire, the girls reported if they had experienced menarche ($n = 26$). Median (range) gynecological age (time from menarche to examination) was 0.69 (0.03–2.30) years. Examinations were not scheduled according to menstrual cycle. No girls had a history of endocrine, gynecological, or cerebral illness.

MRI

MRI is considered useful as gold standard for measurements of breast glandular tissue because it is the only nonradiating technique, which provides reproducible high-resolution slice imaging of the breasts. In T1-weighted images, there is an excellent soft tissue contrast, because fat has a high signal-intensity (is white) whereas glandular tissue has a considerably lower signal (dark gray), which makes it easy to delineate the border between the two tissues.

Magnetic resonance imaging was performed with a 3-Tesla MRI (Magnetom Verio; Siemens AG, Erlangen, Germany). In all girls ($n = 106$), a 2D T1-weighted flash breath hold sequence was performed in the transverse plane using a 32-channel receiver array coil. The scan parameters were as follows: TR/TE 103/2.5 milliseconds, Flip angle 70°, FOV 380 mm, matrix 320 × 240 and slice thickness 5 mm.

Mammary dimensions were measured by a single observer (C.P.H.) blinded for Tanner stages. In cases of minimal breast tissue (<0.6 ml), final measurements were performed by an experienced radiologist (E.F.). The height and width of the glandular tissue were measured in the transverse plane at the level of the papilla (**Supplementary Figure S1** online). Mammary volume was calculated assuming a conical shape (conical MRI: $\pi/3 \times radius^2 \times height$). By this method, we could avoid to include subcutaneous fat; however in girls with highest BMI, the volume contains both glandular tissue and intramammarian adipose tissue. Mammary volume was determined in 103 girls (93 had glandular tissue). In the remaining girls ($n = 3$), assessment was not possible because the region of interest was not included in the scanned area. Ultimately, 100 girls had both a MRI determined mammary volume and a Tanner evaluation.

TAUS

All girls were scanned with a full bladder. A single experienced operator performed the examinations using a Voluson E8 Ultrasound System (GE Healthcare Medical Systems, Zipf, Austria) with a multifrequency transabdominal probe (RM6C, 3–8 MHz). Ovarian volume was calculated in 104/111/103 girls (L/R/sum) (Ellipsoid TAUS). In the remaining girls, assessment was not possible due to the presence of large cysts (>25 cm³) ($n = 2$), inadequate picture quality ($n = 3$), or insufficient bladder volume ($n = 3$). We report the sum of volumes from both ovaries. Uterine volume was calculated in 112 girls (Ellipsoid TAUS) and the endometrial thickness measured in 110 girls. Details regarding TAUS have previously been described (16,19)

Reproductive Hormone Assays

All blood samples ($n = 112$) were drawn between 8:00 AM and 2:00 PM from an antecubital vein, clotted, and centrifuged; serum was stored at −20 °C until hormone analyses. Blood samples were analyzed after maximum 2 y of storage. All samples were analyzed in the same laboratory blinded for age and pubertal stage.

Serum FSH and LH were measured by time-resolved immunofluorometric assay (Delfia; PerkinElmer, Boston, MA) with detection limits of 0.06 and 0.05 IU/l, respectively. Undetectable LH values ($n = 9$) were assigned 0.025 IU/l (0.5 × detection limit). Intra- and interassay CVs were less than ≤ 3%. Serum concentrations of inhibin B were measured using the Beckman Coulter Gen II assay with a detection limit of 3 pg/ml, and an interassay CV < 11%. Serum estradiol was measured by radioimmunoassay (Pantex, Santa Monica, CA), with a detection limit of 18 pmol/l and an intra- and interassay CVs were < 8 and < 15%, respectively. According to standard practice, all undetectable values ($n = 2$) were assigned 9 pmol/l (0.5 × detectiol limit).

Statistics

In the first analyses, we included each breast as single observation. Inter-individual variation was assessed by coefficience of variation (CV% = SD/mean × 100). The correlation between the glandular breast tissue volume (MRI) and pubertal breast stages (Tanner) was calculated using Spearman's rho. Crosstabs were made to calculate

sensitivity and specificity, using MRI as gold standard to detect glandular tissue. When evaluating Tanner staging B2-B5 as a marker of glandular breast tissue (MRI), we used the volume of each breast ($n = 200$). To evaluate if glandular breast volume was associated with BMI and body fat percentage, breast volume tertiles were compared using the Kruskal-Wallis *N*-test.

In the subsequent analyses, we used receiver operating characteristic analyses to evaluate sensitivity and specificity of different markers (FSH, LH, estradiol, inhibin B, uterine volume, endometrial thickness, ovarian volume) to distinguish between prepubertal (no glandular breast tissue on either side) and pubertal girls (at least one breast with measurable glandular tissue volume, i.e., B2-B5).

All statistical analyses were carried out using the SPSS software (Version 22; SPSS, Chicago, IL).

Ethical Considerations
The Copenhagen Mother-Child cohort was conducted according to the protocols approved by the local ethics committee (KF 01–030/97/KF 01 276 357/H-1–2009–074) and the Danish Data Protection Agency (1997–1200–074/2005–41–5545/2010–41–4757). All parents and girls gave informed written consent to participate in this study.

REFERENCES
1. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Arch Dis Child 1969;44:291–303.
2. Rasmussen AR, Wohlfahrt-Veje C, Tefre de Renzy-Martin K, et al. Validity of self-assessment of pubertal maturation. Pediatrics 2015;135:86–93.
3. Schlossberger NM, Turner RA, Irwin CE Jr. Validity of self-report of pubertal maturation in early adolescents. J Adolesc Health 1992;13:109–13.
4. Aksglaede L, Sørensen K, Petersen JH, Skakkebaek NE, Juul A. Recent decline in age at breast development: the Copenhagen Puberty Study. Pediatrics 2009;123:e932–9.
5. Biro FM, Galvez MP, Greenspan LC, et al. Pubertal assessment method and baseline characteristics in a mixed longitudinal study of girls. Pediatrics 2010;126:e583–90.
6. Herman-Giddens ME. Recent data on pubertal milestones in United States children: the secular trend toward earlier development. Int J Androl 2006;29:241–6; discussion 286–90.
7. Sun SS, Schubert CM, Liang R, et al. Is sexual maturity occurring earlier among U.S. children? J Adolesc Health 2005;37:345–55.
8. Jaruratanasirikul S, Chanpong A, Tassanakijpanich N, Sriplung H. Declining age of puberty of school girls in southern Thailand. World J Pediatr 2014;10:256–61.
9. Herman-Giddens ME, Slora EJ, Wasserman RC, 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;99:505–12.
10. Sun SS, Schubert CM, Chumlea WC, et al. National estimates of the timing of sexual maturation and racial differences among US children. Pediatrics 2002;110:911–9.
11. Euling SY, Herman-Giddens ME, Lee PA et al. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. Pediatrics 2008;121:172–91.
12. Neely EK, Wilson DM, Lee PA, Stene M, Hintz RL. Spontaneous serum gonadotropin concentrations in the evaluation of precocious puberty. J Pediatr 1995;127:47–52.
13. Boisen KA, Kaleva M, Main KM, et al. Difference in prevalence of congenital cryptorchidism in infants between two Nordic countries. Lancet 2004;363:1264–9.
14. Chellakooty M, Schmidt IM, Haavisto AM, et al. Inhibin A, inhibin B, follicle-stimulating hormone, luteinizing hormone, estradiol, and sex hormone-binding globulin levels in 473 healthy infant girls. J Clin Endocrinol Metab 2003;88:3515–20.
15. Mouritsen A, Johansen ML, Wohlfahrt-Veje C, et al. Determination of adrenal volume by MRI in healthy children: associations with age, body size, pubertal stage and serum levels of adrenal androgens. Clin Endocrinol (Oxf) 2014;81:183–9.
16. Hagen CP, Mouritsen A, Mieritz MG, et al. Uterine volume and endometrial thickness in healthy girls evaluated by ultrasound (3-dimensional) and magnetic resonance imaging. Fertil Steril 2015;104:452–9.e2.
17. Wohlfahrt-Veje C, Tinggaard J, Winther K, et al. Body fat throughout childhood in 2647 healthy Danish children: agreement of BMI, waist circumference, skinfolds with dual X-ray absorptiometry. Eur J Clin Nutr 2014;68:664–70.
18. Slaughter MH, Lohman TG, Boileau RA, et al. Skinfold equations for estimation of body fatness in children and youth. Hum Biol 1988;60:709–23.
19. Hagen CP, Mouritsen A, Mieritz MG, et al. Circulating AMH reflects ovarian morphology by magnetic resonance imaging and 3D ultrasound in 121 healthy girls. J Clin Endocrinol Metab 2015;100:880–90.