Published in final edited form as:

J Pediatr. 2017 July ; 186: 172–178.e3. doi:10.1016/j.jpeds.2017.03.050.

Validity of self-assessed sexual maturation against physician assessments and hormone levels

Jorge E. Chavarro, MD, ScD[1], **Deborah J. Watkins, PhD**[2], **Myriam C. Afeiche, PhD**[1,3], **Zhenzhen Zhang, PhD**[4,5], **Brisa N. Sánchez, PhD**[4], **David Cantonwine, PhD**[6], **Adriana Mercado-García, MD, MPH**[7], **Clara Blank-Goldenberg**[8], **John D. Meeker, ScD**[2], **Martha María Téllez-Rojo, PhD**[7], and **Karen E. Peterson, ScD**[1,5,9]

[1]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[2]Department of Environmental Health Sciences, University of Michigan School of Public Health, Ann Arbor MI, USA

[3]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[4]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor MI, USA

[5]Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor MI, USA

[6]Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA

[7]Research Center for Nutrition and Health, Instituto Nacional de Salud Pública, Cuernavaca, MOR, MEXICO

[8]American British Cowdray Medical Center, Mexico City, DF, MEXICO

[9]Center for Human Growth and Development, University of Michigan, Ann Arbor MI, USA

Abstract

Objective—To compare self-report and physician assessments of sexual maturation against serum hormone markers to evaluate the hypothesis that the validity of self-assessed sexual maturation is underestimated in traditional validation studies.

Study design—We adapted a self-assessment instrument that 248 Mexican children and adolescents, age 8–13 years, completed. Participants were examined by a trained pediatrician and provided fasting blood samples for measurement of reproductive (testosterone, estradiol, sex hormone-binding globulin (SHBG), Inhibin B) and other hormones (C-peptide, insulin-like growth factor 1 (IGF-1), leptin, dehydroepiandrosterone sulfate (DHEA-S)) known to change during adolescence. Spearman correlations (*r*) were calculated among the average rank of all hormones, self-, and physician-assessed Tanner stage. The method of triads was used to assess validity of

Correspondence: Jorge E. Chavarro, MD, ScD, Department of Nutrition, Harvard T.H. Chan School of Public Health, 665 Huntington Ave., Boston, MA 02115, jchavarr@hsph.harvard.edu, Phone: 617-432-4584, Fax: 617-432-2435.

Portions of this study were presented as a poster during the Society for Epidemiologic Research Meeting, Seattle, WA, June <<Dates>>, 2014.

self-reports by estimating correlations between self-assessments and true, but unobservable, sexual maturation based on all available data. 95% confidence intervals (CI) were constructed using bootstrap sampling.

Results—Validity of self-reported genitalia staging for boys was modest (*r*[95%CI]=0.50[0.31–0.65]) and inferior to physician assessment (0.75[0.56–0.93]). Breast stage was well reported (0.89[0.79–0.97]) and superior to physician assessment (0.80[0.70–0.89]). Pubic hair stage reported by boys (0.91[0.79–0.99]) and girls (0.99[0.96–1.00]) were superior to physician assessment (0.79[0.57–0.97] and 0.91[0.83–0.97], respectively).

Conclusion—Self-assessment can be validly used in epidemiologic studies for evaluation of sexual maturation in children. Physician assessment may be necessary for accurate assessment of genitalia development in boys.

Keywords

Puberty; validation; reproductive hormones; biomarkers; epidemiology

---

The age at which girls and boys enter puberty has decreased over the past several decades, partially due to changes in nutrition, hygiene, and improved health and socioeconomic status (1, 2). However, there is concern that environmental factors may also be contributing to earlier pubertal onset, with potential adverse effects (1, 3). For example, children who enter puberty early have a higher risk of alcohol and substance use (4, 5) and risky behavior during adolescence (6) compared with their peers. Earlier pubertal onset has also been associated with increased risk of a variety of disorders in adolescence (7–13) and adults (14–30). As a result, self-reported sexual maturation is extensively used in epidemiologic studies as a study outcome or as a critical covariate when evaluating other associations.

Validation studies have found reasonable agreement between self-reported and physician-observed Tanner stages of sexual maturation (31–34). Sources of error in self-assessment include, for example, misidentification by obese adolescents of fat tissue as breast tissue (35–37). Furthermore, although physician assessment has always been used as the gold standard in validation studies, these assessments may also be subject to measurement error and thus result in apparent lower validity of self-reports. Specifically, Tanner staging is dependent on observer training and experience, and basing validity of self-reports solely on correlations between self- and physician-assessed sexual maturation may underestimate children's ability to rate their own pubertal development.

In this study we evaluated the hypothesis that the validity of self-assessed sexual maturation is underestimated in traditional validation studies. To address this question we examined relationships between self-reported sexual maturation, physician-assessed sexual maturation, and a panel of serum hormone concentrations that serve as an objective marker of pubertal development. We then use the method of triads, a technique first proposed for the validation of dietary assessment tools (38, 39), to obtain an estimate of the relation between self-reported and true, but unobservable, pubertal status.

METHODS

The study participants were children enrolled in an ongoing longitudinal birth cohort study in which mothers were recruited from maternity hospitals in Mexico City, as described elsewhere (40–43). Our analysis includes mothers who were recruited in their first trimester of pregnancy, between 1997 and 2004, into the second and third of three sequentially enrolled cohorts. Women were eligible to participate if they were >14 years of age, pregnant, did not have a high-risk pregnancy, and had plans to reside in the area for at least 5 years. The children of enrolled mothers were followed from birth to 5 years of age. In 2010, 250 child participants were selected based on availability of archived maternal biological specimens and ages ranging from 8–13 years, thus likely to be undergoing the pubertal transition, and invited to participate in a follow-up study on growth and sexual maturation. Participants completed a questionnaire on self-reported sexual maturation (described in detail below), had a physical exam, and provided a blood sample for hormone analysis. Of the 250 adolescents who filled out the questionnaire, 131 girls and 117 boys had information on physician and self-reported sexual maturation and serum hormone levels. Research protocols were approved by the Institutional Review Board at the University of Michigan and the Ethics Committee of the Mexico National Institute of Public Health. All child participants provided informed assent and were accompanied by their mother or guardian, who signed a letter of informed consent prior to participation.

Self-reported sexual maturation

We developed a questionnaire for self-report of sexual maturation based on adaptations to the original Tanner stages for secondary sexual characteristics (44). The questionnaire contained line drawings depicting the five Tanner stages and descriptions of each stage. Modifications to the layout of the line drawings were based on those made by Taylor et al (45) (Figure 2; available at www.jpeds.com). The questionnaire was translated into Spanish, reviewed by native speakers and field staff, and piloted among 12 participants aged 7 to 14 years prior to being administered to the study population. At the study visit, a member of the research team explained the objective of the questionnaire to the mother, showed her a sample of the figures, and explained that she would be given the option of discussing the questionnaire with their child prior to completion. The researcher reviewed the questionnaire with the child, explained that he/she had the option of having their mother present while they (the child) filled it out, and left the room. Children were asked to select their self-perceived stage of development by choosing the drawings and descriptions closest to their current stage of sexual development, and girls were asked to report their attainment of menarche (yes/no; if yes, at what age). Both boys and girls were asked to report any practices of pubic hair shaving as this might bias their perceived Tanner staging. After completion, participants folded the questionnaire and returned it to the researcher.

Physical exam and standardization

A pediatrician (either CB-G or AM-G) trained according to standard methods (by JC) assessed Tanner staging for breast and pubic hair development in girls and for genitalia and pubic hair development in boys. Testicular volume was assessed in boys using a Prader orchidometer. Trained nurses also measured height and weight at this visit. Prior to

launching the study, one of the investigators (JEC) conducted a standardization of anthropometry protocol with the research nurses and of Tanner staging with the pediatricians; the latter focused on defining rules to address key assessment issues including differentiation of adipose and breast tissue in overweight girls, asymmetric breast stages and pubic hair removal.

Hormone analysis

During the study visit, a trained phlebotomist collected a fasting blood sample from each child for hormone analysis. Samples were centrifuged, separated into aliquots, and the serum was stored at −80°C until shipment on dry ice to the University of Michigan School of Public Health. We measured total estradiol, total testosterone, inhibin B, and sex hormone-binding globulin (SHBG), DHEA-S, leptin, c-peptide, and insulin-like growth factor 1 (IGF-1) in serum as objective but unspecific biomarkers of sexual and somatic development during puberty. DHEA-S, E2, SHBG, and T were measured using an automated chemiluminescent immunoassay (Bayer Diagnostics ACS:180) and active inhibin B was assayed using Gen II ELISA (Beckman Coulter, Webster, TX), all at the Clinical Ligand Assay Service Satellite Laboratory at the University of Michigan (Ann Arbor, MI). Leptin, c-peptide, and IGF-1 were measured at the Michigan Diabetes Research and Training Center Chemistry Lab using an automated chemiluminescence immunoassay (c-peptide, IGF-1; Immulite 1000), or radioimmunoassay (leptin; Millipore).

Statistical analyses

Because all the sex hormones measured are known to change during adolescence (46–48) but none is a specific marker of the progression through puberty, we constructed a summary score of all the hormones measured (E2, T, inhibin B, SHBG, DHEA-S, leptin, c-peptide and IGF-1). Given that hormones are measured in different units, we ranked the measurements for each hormone, from lowest to highest measured value, and then calculated for each participant the average of the ranks across the eight hormones measured. We used the average of the ranks, instead of actual levels of any one hormone, as our objective biomarker of pubertal status. We estimated pair-wise Spearman correlations between self-assessed sexual maturation status, physician assessed sexual maturation status and the average hormone rank. We used the method of triads (38, 49) to estimate correlations of self-assessed and physician assessed sexual maturation status with true, but unobservable, sexual maturation status (Figure 1; available at www.jpeds.com). Specifically, the correlation between self-assessment and true sexual maturation was estimated as:

$$r_{ST} = \sqrt{\frac{r_{SH} \times r_{SP}}{r_{PH}}}$$

and the correlation between physician-assessment and true sexual maturation was estimated as:

$$r_{PT} = \sqrt{\frac{r_{PH} \times r_{SP}}{r_{SH}}}$$

where r_{ST} is the Spearman correlation between self-assessment and true sexual maturation; r_{PT} is the Spearman correlation between physician-assessment and true sexual maturation; r_{SH} is the Spearman correlation between self-assessment and the average rank of hormone levels; r_{SP} is the Spearman correlation between self-assessment and physician-assessment; and r_{PH} is the Spearman correlation between physician-assessment and the average rank of hormone levels. We used bootstrap sampling to construct confidence intervals around correlations of interest. A total of 1000 bootstrap samples were obtained by random sampling with replacement. For each bootstrap sample, the same estimation procedure using the method of triads were repeated and the 2.5th and 97.5th percentiles of all the bootstrap estimates were used as the nonparametric confidence interval. These analyses were conducted separately for each sex, and separately for each Tanner staging variable.

We also calculated the specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV) of self-assessed pubertal onset, defined as having a Tanner stage >1 for pubic hair, breast, or genital development, compared with physician-assessed pubertal onset using the formulas listed in Figure 3 (available at www.jpeds.com). In addition, we calculated the percent of self-assessments that were: a) in exact agreement with, and b) ± 1 stage agreement with physician-assessments for pubic hair, breast, and genital Tanner stage.

Age-specific BMI and height z-scores were calculated using the SAS macro based on the 2007 World Health Organization (WHO) growth reference for 5–19 year olds (50).

RESULTS

The demographic characteristics of the study participants are shown in Table I. Mean age at assessment was 10 years. Their BMI was, on average, 0.8 standard deviations higher and their height 0.2 standard deviations lower than the WHO reference (51). Eighteen percent of boys and 25 percent of girls were considered overweight (≥85th and <95th percentile BMI for age), while 31 percent of boys and 24 percent of girls were considered obese (≥95th percentile BMI for age). Most participants were pre-pubertal according to their physician-assessed secondary sexual characteristics but there was variability in stage of sexual maturation among post-pubertal children (Table IV; available at www.jpeds.com). Circulating levels of reproductive hormones were reflective of a primarily pre-pubertal study population (Table V; available at www.jpeds.com).

Pairwise correlations between self-assessed sexual maturation, physician-assessed Tanner staging and the average rank of hormone levels are shown in Table II. Girls' self-assessments were closer to physician's assessments than boys' self-assessments. For boys, pubic hair self-assessment was more strongly correlated to physician's assessments than their genitalia development assessment when compared against physician's Tanner staging or testicular volume. Hormone levels were more strongly correlated to physician's assessment of genitalia staging than to boy's self-assessment of genitalia development.

However, hormone levels were more strongly related to girls' self-assessment of breast and pubic hair development and to boys' self-assessment of pubic hair development than to physician's assessments of these traits. The concordance between self- and physician assessments was high, particularly for pubic hair (Table VI; available at www.jpeds.com); 84% of girls and 83% of boys agreed exactly with the physician's assessment of pubic hair stage. Exact agreement for breast and genitalia stage was lower. Nevertheless, 78% of self-assessments of genital development and 97% of self-assessments of breast development agreed with physician-assessments within ±1 stage (Table VI). Furthermore, the tool was highly sensitive for classifying children as pre-pubertal (Tanner stage 1) or post-pubertal (Tanner stage >1) on all four traits assessed. For example, 95% of boys and 100% of girls classified as having a pubic hair Tanner stage >1 by a physician also rated themselves in a pubic hair Tanner stage >1 (Table III).

Estimates of the correlations of self-assessed and physician assessed Tanner staging with true sexual maturation status are also shown in Table II. Physician's assessment of true sexual maturation status was highly valid. The correlations between physician assessment and true sexual maturation status ranged from 0.75 for boys genitalia to 0.91 for girls pubic hair. Girls were very accurate reporters of their sexual maturation status and outperformed physicians in their assessment of breast and pubic hair assessment. Boys also outperformed physicians in reporting their pubic hair development. However, physician assessment was considerably better than self-assessment of genitalia development when compared against physician's Tanner staging or testicular volume measurement.

DISCUSSION

We used a novel approach to evaluate the validity of self-assessed sexual maturation status in a cohort of Mexican children and adolescents. The observed correlations between self- and physician-assessments were comparable with those observed in previous validation studies where physician assessment has been considered the gold standard (31, 52, 53). By introducing a third independent measure of sexual maturation status we were able to evaluate the relation of both self- and physician assessment with the true, but unobservable, sexual maturation status. This novel approach showed that self-assessed sexual maturation is very accurately reported by females and that males can accurately report pubic hair development. These results suggest that self-reported sexual maturation aided with line drawings and brief descriptions of each Tanner stage could meet the needs of most epidemiologic studies assessing sexual maturation. Our data suggests, however, that physician assessments may be necessary in studies where differentiating genitalia development from pubic hair development in males is of interest.

Our observed correlations between self- and physician-assessed sexual maturation status were similar to those reported in a recent study in Denmark (52), where assessments of pubic hair development were highly correlated in both males and females (r=0.70 and 0.80, respectively), as were girls' assessments of breast development (r= 0.74). However, the correlations between self- and physician-assessed genital development among boys was weaker in the present study (r=0.38) in comparison with the Danish study (r=0.61). Among girls, correlations between self- and physician-assessed breast and pubic hair development

were also similar to those previously reported in the US (31, 34), Hong Kong (54), and Brazil (55) whereas correlations for genital and pubic hair development among boys were somewhat lower in the current study (34, 54, 55). Boys tended to overestimate their genital development stage, while girls' assessments were generally more highly correlated with physician reports (Table V), a pattern that has been observed in several previous studies (32, 53, 56). However, using physician assessment as the gold standard has the implicit assumption that it is without error when there is evidence that this assumption is weak (57–60). Another method of validating self-assessed sexual maturation status is comparing it with hormone-based measures of sexual maturation, although previous studies have been limited. One study of girls aged 8–18 years reported a correlation of 0.61 between average Tanner stage and estradiol (61), which is slightly lower than our observed correlation of 0.73 between breast development stage and overall hormonal milieu measure. In a study of overweight and obese adolescents, many participants overestimated their maturation status in comparison with hormone-based measures, although the weight status of participants likely influenced the accuracy of their self-assessments (62). By simultaneously comparing self-reports with physician assessments and hormone levels we were able to estimate correlations between self-assessed and true underlying sexual maturation status. In the current study, self-reports were very highly correlated with estimated true maturation status for breast development in girls, and pubic hair development in both boys and girls. In fact, these correlations were higher than correlations between physician assessments and estimated true status (Table II), and higher than previously reported correlations between self- and physician-assessments, suggesting that previous work may have consistently underestimated the validity of self-reported sexual maturation.

As mentioned above, using physician assessment as the gold standard for evaluation of sexual maturation status has the implicit assumption that it is free of error. However, there is evidence that physician assessment of physical traits in general, and of sexual maturation in particular, are subject to considerable within-observer and between-observer variability. For example, substantial variation in assessments of skin fold thickness, waist circumference, and other anthropometry measures (63, 64), as well as blood pressure measurements (65–67) have been demonstrated. The reliability of physician assessed sexual maturation status has generally been poor, especially for assessment of testicular volume (58, 60) and breast development (57), and is highly dependent on appropriate training (59, 68). Our results further suggest that physician assessments are not free of error. Although strongly correlated to hormone levels and true sexual maturation status, correlations between physician assessment and true status tended to be slightly lower than self-assessment vs. true status correlations and self-assessment vs. hormone correlations. The exception to this was physician assessed genital development in boys, which was more highly correlated with true status (r=0.80) compared with self-assessments (r=50). To our knowledge, this is the first report of the validity of physician assessed sexual maturation status, and thus further research is needed.

The method of triads is a technique often used in nutritional epidemiology to validate dietary assessments by estimating correlations with true, but unknown dietary intake. This method involves making triangular comparisons between three distinct measures (e.g. food records, questionnaires, and biomarkers) to estimate relationships of these variables with true dietary

intake (38, 39). However, the method of triads can be used for evaluation of validity of constructs beyond diet as long as specific assumptions are met. Appropriate application of the method of triads requires that associations between the three measurements being compared are solely due to their relationships with the true variable of interest, relationships between variables are linear, and errors are independent (69). The relationships between the three measurements of sexual maturation used in the present analysis – self-reports, physician assessments, and serum hormone levels – meet these required assumptions. Although hormone concentrations are subject to random measurement error, this is likely to be independent of errors in either self- or physician- assessed sexual maturation (69). Errors in self-assessed sexual maturation are likely to be based on age, sex, weight status, and other similar factors (52, 53, 56), whereas errors in physician assessments are likely not due to these issues, but rather differences in training and standardization. Therefore, errors in assessments are not likely to be correlated. One exception to this may be in cases of overweight or obese adolescent girls, where fat tissue may lead to inaccurate characterization of breast Tanner stage status in both self-assessments and physician assessments (35–37). However, proper training of physicians, including both observation and palpation of breast tissue, would limit related errors in physician-assessments, and thus minimize correlations with errors in self-assessments. Characteristics of our study including pre-study training in standard of assessment procedures to address this specific issue suggest that this source of correlated error may be minimal in our study. Our use of the method of triads to evaluate the validity of both self- and physician assessments of sexual maturation demonstrates that this method may be applied to a number of situations outside of dietary assessment. A limitation of this study is that the majority of participants were in the early stages of puberty, with very few children at Tanner stages 4 or 5 for breast, genital, or pubic hair development. This limits our ability to evaluate sexual maturation assessments in the later stages of puberty. As girls generally develop earlier than boys, this could also play a role in the observed differences by sex in self-assessment validity by sex. However, it allowed us to evaluate assessments at the onset of puberty, the developmental transition that is the focus of most epidemiological studies. Another limitation was the measurement of steroid hormones by immunoassay rather than LC-MS/MS, the method recommended by the Endocrine Society particularly for low testosterone levels in women and children. However, we used several hormone concentrations to create a hormone profile rank, minimizing the influence of a small number of values below the detection limit for one hormone. Our use of serum hormones as a third measure of sexual maturation status was also an important strength of our study, as hormone levels are an objective biomarker, independent of both self- and physician assessments. This allowed us to evaluate the validity of both self- and physician- assessment by estimating correlations between these measures and true sexual maturation status.

Acknowledgments

Supported by the U.S. National Institute of Environmental Health Sciences (NIEHS) (R01 ES007821, R01 ES021446, R01 ES014930, R01 ES013744, P01 ES012874 and P30 ES017885), Consejo Nacional de Ciencia y Tecnología (CONACyT) (37210-M, 29192-M and 41912-M), U.S. Environmental Protection Agency (EPA) STAR Research Assistance Agreement (RD-83172501), and NIEHS/EPA Formative Children's Environmental Health and Disease Prevention Center (P20 ES018171/RD834800) and NIEHS/EPA Children's Environmental Health and

Disease Prevention Center (P01 ES02284401/RD 83543601). M.A. is currently an employee of the Nestlé Research Center, Lausanne, Switzerland. The other authors declare no conflicts of interest.

We gratefully acknowledge the contribution of the American British Cowdray Hospital in Mexico City for use of its research facilities.

Abbreviations

DHEA-S	dehydroepiandrosterone sulfate
GnRH	gonadotropin releasing hormone
HPA	hypothalamic-pituitary-adrenal
HPG	hypothalamic-pituitary-gonadal
IGF-1	insulin-like growth factor 1
NPV	negative predictive value
PPV	positive predictive value
SHBG	sex hormone-binding globulin

References

1. Lee Y, Styne D. Influences on the onset and tempo of puberty in human beings and implications for adolescent psychological development. Hormones and Behav. 2013; 64:250–61.
2. 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 Rev. 2003; 24:668–93. [PubMed: 14570750]
3. Buck Louis GM, Gray LE Jr, Marcus M, Ojeda SR, Pescovitz OH, Witchel SF, et al. Environmental factors and puberty timing: expert panel research needs. Pediatrics. 2008; 121:S192–207. [PubMed: 18245512]
4. Kaltiala-Heino R, Koivisto AM, Marttunen M, Frojd S. Pubertal timing and substance use in Middle Adolescence: A 2-year follow-up study. JYouth Adolesc. 2011; 40:1288–301. [PubMed: 21533658]
5. Patton GC, McMorris BJ, Toumbourou JW, Hemphill SA, Donath S, Catalano RF. Puberty and the onset of substance use and abuse. Pediatrics. 2004; 114:E300–E6. [PubMed: 15342890]
6. Collado-Rodriguez A, MacPherson L, Kurdziel G, Rosenberg LA, Lejuez CW. The relationship between puberty and risk taking in the real world and in thelaboratory. Personality and Individual Differences. 2014; 68:143–8. [PubMed: 24999291]
7. Mendle J, Turkheimer E, Emery RE. Detrimental Psychological Outcomes Associated with Early Pubertal Timing in Adolescent Girls. DevRev. 2007; 27:151–71.
8. Graber JA. Pubertal timing and the development of psychopathology in adolescence and beyond. Hormones and Behav. 2013; 64:262–9.
9. Hamilton JL, Hamlat EJ, Stange JP, Abramson LY, Alloy LB. Pubertal timing and vulnerabilities to depression in early adolescence: differential pathways to depressive symptoms by sex. JAdolesc. 2014; 37:165–74. [PubMed: 24439622]
10. Kaltiala-Heino R, Marttunen M, Rantanen P, Rimpela M. Early puberty is associated with mental health problems in middle adolescence. Soc Sci Med. 2003; 57:1055–64. [PubMed: 12878105]
11. Tremblay L, Lariviere M. The influence of puberty onset, Body Mass Index, and pressure to be thin on disordered eating behaviors in children and adolescents. Eating Behaviors. 2009; 10:75–83. [PubMed: 19447348]
12. Klump KL. Puberty as a critical risk period for eating disorders: a review of human and animal studies. Horm Behav. 2013; 64:399–410. [PubMed: 23998681]

13. Mendle J, Harden KP, Brooks-Gunn J, Graber JA. Peer relationships and depressive symptomatology in boys at puberty. Dev Psychol. 2012; 48:429–35. [PubMed: 22103302]
14. Feng Y, Hong XM, Wilker E, Li ZP, Zhang WB, Jin DL, et al. Effects of age at menarche, reproductive years, and menopause on metabolic risk factors for cardiovascular diseases. Atherosclerosis. 2008; 196:590–7. [PubMed: 17675039]
15. Jacobsen BK, Oda K, Knutsen SF, Fraser GE. Age at menarche, total mortality and mortality from ischaemic heart disease and stroke: the Adventist Health Study, 1976–88. Int J Epidemiol. 2009; 38:245–52. [PubMed: 19188208]
16. Mueller NT, Duncan BB, Barreto SM, Chor D, Bessel M, Aquino EML, et al. Earlier age at menarche is associated with higher diabetes risk and cardiometabolic disease risk factors in Brazilian adults: Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). Cardiovasc Diabetol. 2014; 13:22. [PubMed: 24438044]
17. Prentice P, Viner RM. Pubertal timing and adult obesity and cardiometabolic risk in women and men: a systematic review and meta-analysis. Int J Obes Relat Metab Disord. 2013; 37:1036–43.
18. Chen L, Zhang C, Yeung E, Ye A, Mumford SL, Wactawski-Wende J, et al. Age at menarche and metabolic markers for type 2 diabetes in premenopausal women: the BioCycle Study. JClinEndocrinolMetab. 2011; 96:E1007–12.
19. Elks CE, Ong KK, Scott RA, van der Schouw YT, Brand JS, Wark PA, et al. Age at Menarche and Type 2 Diabetes Risk The EPIC-InterAct study. Diabetes Care. 2013; 36:3526–34. [PubMed: 24159179]
20. Frontini MG, Srinivasan SR, Berenson GS. Longitudinal changes in risk variables underlying metabolic Syndrome X from childhood to young adulthood in female subjects with a history of early menarche: the Bogalusa Heart Study. Int J Obes Relat Metab Disord. 2003; 27:1398–404. [PubMed: 14574352]
21. He CY, Zhang CL, Hunter DJ, Hankinson SE, Louis GMB, Hediger ML, et al. Age at Menarche and Risk of Type 2 Diabetes: Results From 2 Large Prospective Cohort Studies. Am J Epidemiol. 2010; 171:334–44. [PubMed: 20026580]
22. Janghorbani M, Mansourian M, Hosseini E. Systematic review and meta-analysis of age at menarche and risk of type 2 diabetes. Acta Diabetol. 2014; 51:519–28. [PubMed: 24671509]
23. Lakshman R, Forouhi NG, Sharp SJ, Luben R, Bingham SA, Khaw KT, et al. Early age at menarche associated with cardiovascular disease and mortality. J Clin Endocrin Metab. 2009; 94:4953–60.
24. Stockl D, Meisinger C, Peters A, Thorand B, Huth C, Heier M, et al. Age at menarche and its association with the metabolic syndrome and its components: results from the KORA F4 Study. PloS One. 2011; 6:e26076. [PubMed: 22028807]
25. Widen E, Silventoinen K, Sovio U, Ripatti S, Cousminer DL, Hartikainen AL, et al. Pubertal Timing and Growth Influences Cardiometabolic Risk Factors in Adult Males and Females. Diabetes Care. 2012; 35:850–6. [PubMed: 22338106]
26. Apter D, Reinila M, Vihko R. Some endocrine characteristics of early menarche, a risk factor for breast-cancer, are preserved into adulthood. Int J Cancer. 1989; 44:783–7. [PubMed: 2511157]
27. Beral V, Bull D, Pirie K, Reeves G, Peto R, Skegg D, et al. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol. 2012; 13:1141–51. [PubMed: 23084519]
28. Lacey JV Jr, Kreimer AR, Buys SS, Marcus PM, Chang SC, Leitzmann MF, et al. Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort. BMC Cancer. 2009; 9:84. [PubMed: 19292893]
29. Ali AT. Reproductive factors and the risk of endometrial cancer. Int J GynecolCancer. 2014; 24:384–93.
30. Jordan SJ, Webb PM, Green AC. Height, age at menarche, and risk of epithelial ovarian cancer. Cancer epidemiol, Biomarkers Prev. 2005; 14:2045–8. [PubMed: 16103459]
31. Brooks-Gunn J, Warren MP, Rosso J, Gargiulo J. Validity of self-report measures of girls' pubertal status. Child Dev. 1987; 58:829–41. [PubMed: 3608653]

32. Norris SA, Richter LM. Usefulness and reliability of tanner pubertal self-rating to urban Black adolescents in South Africa. J Res Adolescence. 2005; 15:609–24.
33. Duke PM, Litt IF, Gross RT. Adolescents' self-assessment of sexual maturation. Pediatrics. 1980; 66:918–20. [PubMed: 7454482]
34. Morris NM, Udry JR. Validation of a self-administered instrument to assess stage of adolescent development. J YouthAdolesc. 1980; 9:271–80.
35. Bonat S, Pathomvanich A, Keil M, Field A, Yanovski J. Self-assessment of pubertal stage in overweight children. Pediatrics. 2002; 110:743–7. [PubMed: 12359788]
36. Lee K, Valeria B, Kochman C, Lenders CM. Self-assessment of height, weight, and sexual maturation: validity in overweight children and adolescents. J AdolescHealth. 2006; 39:346–52.
37. Sun Y, Tao FB, Su PY. Self-assessment of pubertal Tanner stage by realistic colour images in representative Chinese obese and non-obese children and adolescents. ActaPaediatr. 2012; 101:e163–6.
38. Kaaks R, Riboli E, Esteve J, Van Kappel AL, Van Staveren WA. Estimating the accuracy of dietary questionnare assessments: validation in terms of structural equation models. Stat Med. 1994; 13:127–42. [PubMed: 8122049]
39. Ocke MC, Kaaks RJ. Biochemical markers as additional measurements in dietary validity studies: application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr. 1997; 65:1240S–5S. [PubMed: 9094928]
40. Gonzalez-Cossío T, Peterson KE, Sanín H, Fishbein E, Palazuelos E, Aro A, et al. Decrease in birth weight in relation to maternal bone-lead burden. Pediatrics. 1997; 100:856–62. [PubMed: 9346987]
41. Hernández-Avila M, Peterson KE, Gonzalez-Cossío T, Sanín LH, Aro A, Schnaas L, et al. Effect of maternal bone lead on length and head circumference of newborns and 1-month-old infants. Arch Environ Health. 2002; 57:482–8. [PubMed: 12641193]
42. Téllez-Rojo MM, Hernández-Avila M, Lamadrid-Figueroa H, Smith D, Hernández-Cadena L, Mercado A, et al. Impact of bone lead and bone resorption on plasma and whole blood lead levels during pregnancy. Am J Epidemiol. 2004; 160:668–78. [PubMed: 15383411]
43. Ettinger AS, Lamadrid-Figueroa H, Tellez-Rojo MM, Mercado-Garcia A, Peterson KE, Schwartz J, et al. Effect of calcium supplementation on blood lead levels in pregnancy: a randomized placebo-controlled trial. Environ Health Perspect. 2009; 117:26–31. [PubMed: 19165383]
44. Tanner, JM. Growth at Adolescence: With a General Consideration of the Effects of Hereditary and Environmental Factors uponGrowth and Maturation from Birth to Maturity. Oxford: Blackwell Scientific Publications; 1962.
45. Taylor SJ, Whincup PH, Hindmarsh PC, Lampe F, Odoki K, Cook DG. Performance of a new pubertal self-assessment questionnaire: a preliminary study. Paediatr Perinat Epidemiol. 2001; 15:88–94. [PubMed: 11237120]
46. Gardner, DG., Shoback, D. Appendix: Normal hormone reference ranges. 9th. Gardner, DG., Shoback, D., editors. 2011.
47. 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. J Clin Endocrinol Metab. 1997; 82:3976–81. [PubMed: 9398699]
48. Sehested A, Juul AA, Andersson AM, Petersen JH, Jensen TK, Muller J, et al. Serum inhibin A and inhibin B in healthy prepubertal, pubertal, and adolescent girls and adult women: relation to age, stage of puberty, menstrual cycle, follicle-stimulating hormone, luteinizing hormone, and estradiol levels. J Clin Endocrinol Metab. 2000; 85:1634–40. [PubMed: 10770209]
49. Ocke MC, Kaaks RJ. Biochemical markers as an additional measurement in dietary validity studies: application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr. 1997; 65:1240S–5S. [PubMed: 9094928]
50. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bull WHO. 2007; 85:660–7. [PubMed: 18026621]

51. WHO. Growth reference data for 5–19 years. World Health Organization; 2007. [February 24, 2014] Available from: http://www.who.int/growthref/en/
52. 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; 135:86–93. [PubMed: 25535262]
53. Schlossberger NM, Turner RA, Irwin CE Jr. Validity of self-report of pubertal maturation in early adolescents. J Adol Health. 1992; 13:109–13.
54. Chan NP, Sung RY, Kong AP, Goggins WB, So HK, Nelson EA. Reliability of pubertal self-assessment in Hong Kong Chinese children. J Paediatr Child Health. 2008; 44:353–8. [PubMed: 18476928]
55. Matsudo SMM, Matsudo VKR. Self-assessment and physician assessment of sexual maturation in Brazilian boys and girls–concordance and reproducibility. Am J Hum Biol. 1994; 6:451–5. [PubMed: 28548259]
56. Jaruratanasirikul S, Kreetapirom P, Tassanakijpanich N, Sriplung H. Reliability of pubertal maturation self-assessment in a school-based survey. J PediatrEndocrinol Metab. 2015; 28:367–74.
57. Hergenroeder AC, Hill RB, Wong WW, Sangi-Haghpeykar H, Taylor W. Validity of self-assessment of pubertal maturation in African American and European American adolescents. J Adolesc Health. 1999; 24:201–5. [PubMed: 10195803]
58. Carlsen E, Andersen AG, Buchreitz L, Jorgensen N, Magnus O, Matulevicuus V, et al. Inter-observer variation in the results of the clinical andrological examination including estimation of testicular size. Int J Andrology. 2000; 23:248–53.
59. Slora EJ, Bocian AB, Herman-Giddens ME, Harris DL, Pedlow SE, Dowshen SA, et al. Assessing inter-rater reliability (IRR) of Tanner staging and orchidometer use with boys: a study from PROS. J Pediatr Endocrinol Metab. 2009; 22:291–9. [PubMed: 19554802]
60. 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 J Andrology. 2006; 8:373–8.
61. Rapkin AJ, Tsao JCI, Turk N, Anderson M, Zeltzer LK. Relationships among Self-Rated Tanner Staging, Hormones, and Psychosocial Factors in Healthy Female Adolescents. J Pediatr Adolesc Gynecol. 2006; 19:181–7. [PubMed: 16731411]
62. Raman A, Lustig RH, Fitch M, Fleming SE. Accuracy of self-assessed Tanner staging against hormonal assessment of sexual maturation in overweight African-American children. J Pediatr Endocrinol Metab. 2009; 22:609–22. [PubMed: 19774842]
63. Stomfai S, Ahrens W, Bammann K, Kovacs E, Marild S, Michels N, et al. Intra- and inter-observer reliability in anthropometric measurements in children. Int J Obes Relat Metab Disord. 2011; 35:S45–S51.
64. Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. British J Nutr. 1999; 82:165–77.
65. Canner PL, Borhani NO, Oberman A, Cutler J, Prineas RJ, Langford H, et al. The Hypertension Prevention Trial–assessment of quality of blood-pressure measurements. Am J Epidemiol. 1991; 134:379–92. [PubMed: 1877599]
66. de Graaff J, Ubbink DT, Legemate DA, de Haan RJ, Jacobs M. Interobserver and intraobserver reproducibility of peripheral blood and oxygen pressure measurements in the assessment of lower extremity arterial disease. J Vasc Surg. 2001; 33:1033–40. [PubMed: 11331846]
67. Johenning AR, Karrison TG, Barron WM. Interobserver variability in the measurement of diastolic blood-pressure in pregnancy. Hypertension in Pregnancy. 1995; 14:301–11.
68. Slough JM, Hennrikus W, Chang Y. Reliability of Tanner staging performed by orthopedic sports medicine surgeons. Med Sci Sports Exerc. 2013; 45:1229–34. [PubMed: 23439412]
69. Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. Am J Clin Nutr. 1997; 65:1232S–9S. [PubMed: 9094927]

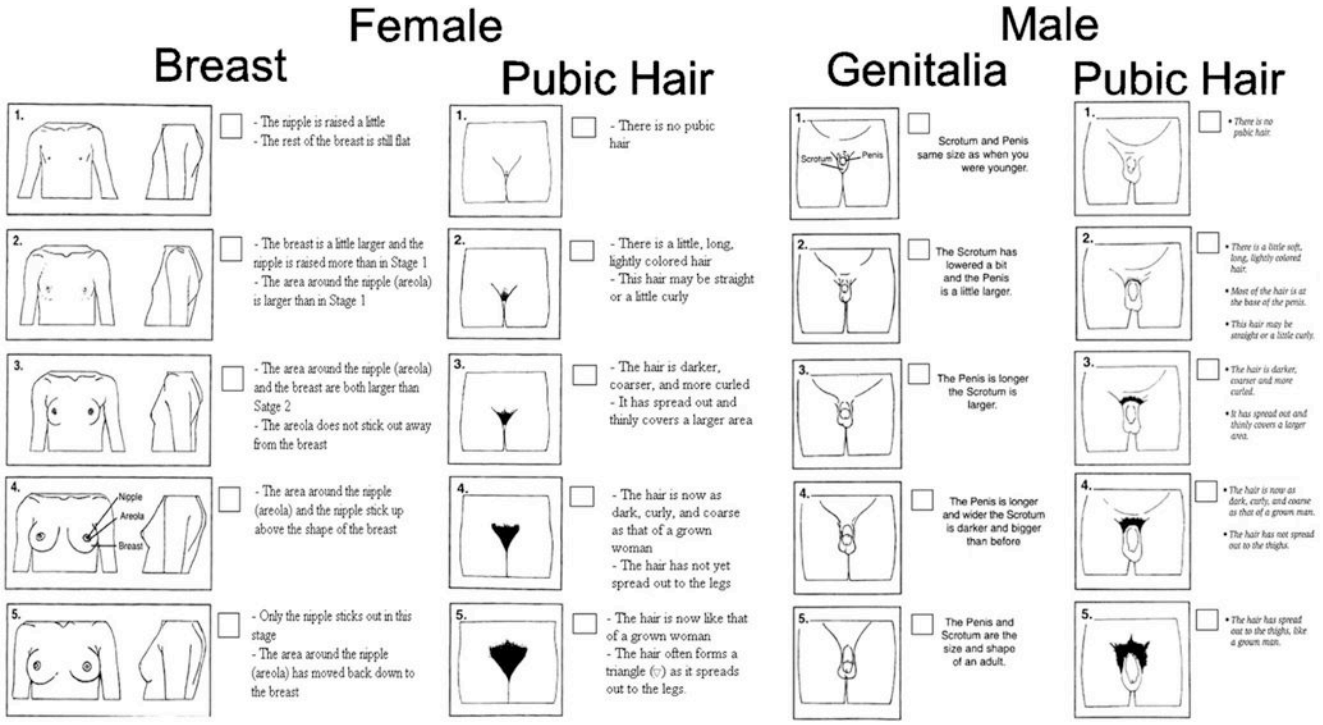


Figure 1.
online. Method of Triads Overview

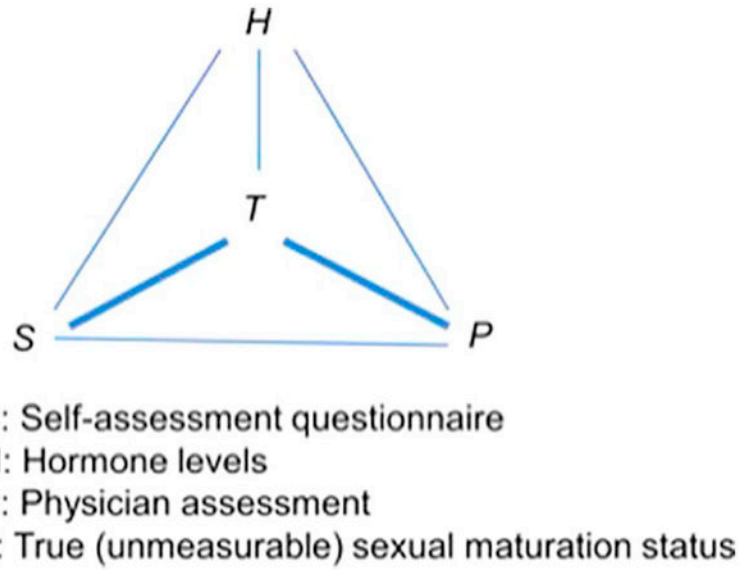
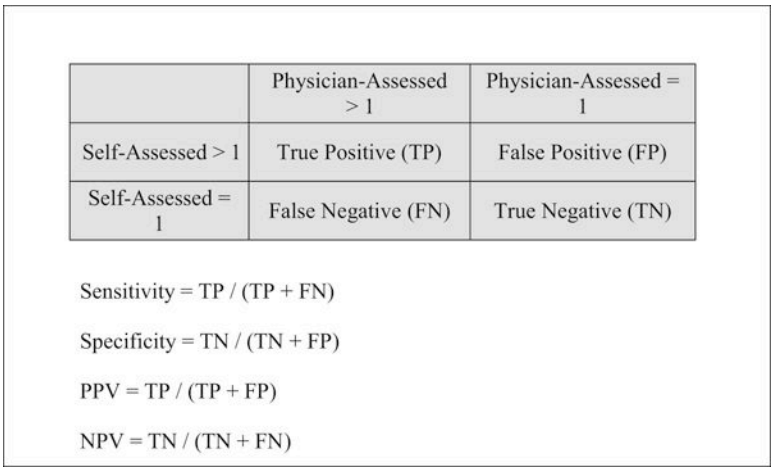


Figure 2.
online. Sexual Maturation Self-Assessment Tool



	Physician-Assessed > 1	Physician-Assessed = 1
Self-Assessed > 1	True Positive (TP)	False Positive (FP)
Self-Assessed = 1	False Negative (FN)	True Negative (TN)

Sensitivity = TP / (TP + FN)

Specificity = TN / (TN + FP)

PPV = TP / (TP + FP)

NPV = TN / (TN + FN)

Figure 3.

online. Formulas for Calculating Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) for Self-Assessed Tanner Staging.

Table 1

Participant Characteristics and Hormone Concentrations.

Characteristic	Mean (SD) or N (%)	
	Girls	Boys
N (%)	131 (52.8)	117 (47.2)
Age, years	10.3 (1.73)	10.4 (1.60)
BMI z-score	0.82 (1.29)	0.87 (1.19)
Height z-score	−0.19 (0.92)	−0.14 (0.81)
Maternal education, years	10.9 (2.83)	11.2 (2.81)
Testosterone, ng/dl	22.0 (14.0)	78.3 (145)
Estradiol, pg/ml	38.1 (50.6)	18.4 (9.34)
SHBG, nmol/L	70.8 (36.8)	82.9 (42.4)
DHEA-S, µg/dl	54.3 (39.5)	67.7 (52.0)
Inhibin B, pg/ml	36.6 (36.9)	124 (71.9)
Leptin, ng/ml	13.9 (10.1)	8.25 (6.37)
C-peptide, ng/ml	1.86 (1.26)	1.60 (1.15)
IGF-1, ng/ml	279 (105)	233 (98.7)
Physician-assessed breast development, N (%) B1[a]	86 (65.6)	
Physician-assessed genitalia development, N (%) G1[b]		56 (49.1)
Physician-assessed pubic hair development, N (%) PH1[c]	97 (74.0)	93 (81.6)

Abbreviations: SD, standard deviation; BMI, body mass index; SHBG, sex hormone-binding globulin; DHEA-S, dehydroepiandrosterone sulfate; IGF-1, insulin-like growth factor 1.

[a]Tanner stage = 1 for breast development.

[b]Tanner stage =1 for genital development, missing for 3 boys.

[c]Tanner stage = 1 for pubic hair development, missing for 3 boys.

Table 2

Spearman Correlations Between Self-Assessment, Physician-Assessment, Hormone Status, and True Sexual Maturation Status.

Self-Assessment	Physician-Assessment	SP[a]	SH[b]	PH[c]	ST[d] (95% CI)	PT[e] (95% CI)
Girls						
Breast	Breast	0.71	0.73	0.66	0.89 (0.79, 0.97)	0.80 (0.70, 0.89)
Pubic Hair	Pubic Hair	0.91	0.62	0.57	1.00 (0.96, 1.00)	0.91 (0.84, 0.97)
Boys						
Genitalia	Genitalia	0.38	0.40	0.61	0.50 (0.31, 0.65)	0.75 (0.56, 0.93)
Genitalia	Testicular volume (largest)	0.40	0.40	0.65	0.50 (0.31, 0.66)	0.80 (0.62, 0.98)
Genitalia	Testicular volume (average)	0.39	0.40	0.66	0.49 (0.30, 0.65)	0.80 (0.62, 0.97)
Pubic Hair	Pubic Hair	0.73	0.61	0.54	0.91 (0.79, 1.00)	0.80 (0.57, 0.97)

[a] SP: Spearman correlation between self-assessment and physician assessment

[b] SH: Spearman correlation between self-assessment and average rank of hormone levels

[c] PH: Spearman correlation between physician assessment and average rank of hormone levels

[d] ST: Spearman correlation between self-assessment and true sexual maturation status

[e] PT: Spearman correlation between physician assessment and true sexual maturation status

J Pediatr. Author manuscript; available in PMC 2018 July 01.

Table 3

Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) of Self-Assessed Pubertal Onset (Girls n=131, Boys n=114).

Girls: Tanner Stage >1 for Breast Development							
	Physician						
Self	**yes**	**no**	**total**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
yes	42	28	70	0.93	0.67	0.60	0.95
no	3	58	61				
total	45	86	131				
Girls: Tanner Stage >1 for Pubic Hair Development							
	Physician						
Self	yes	no	total	Sensitivity	Specificity	PPV	NPV
yes	34	7	41	1	0.93	0.83	1
no	0	90	90				
total	34	97	131				
Boys: Tanner Stage >1 for Genital Development							
	Physician						
Self	yes	no	total	Sensitivity	Specificity	PPV	NPV
yes	54	44	98	0.93	0.21	0.55	0.75
no	4	12	16				
total	58	56	114				
Boys: Tanner Stage >1 for Pubic Hair Development							
	Physician						
Self	yes	no	total	Sensitivity	Specificity	PPV	NPV
yes	20	7	27	0.95	0.92	0.74	0.99
no	1	86	87				
total	21	93	114				

Table 4

online. Percent of Participants at each Tanner Stage According to Self and Physician Assessment by Sex.

	Male (n=114)				Female (n=131)			
	Self-Assessed		Physician-Assessed		Self-Assessed		Physician-Assessed	
Tanner Staging	Genitalia	Pubic Hair	Genitalia	Pubic Hair	Breast	Pubic Hair	Breast	Pubic Hair
1	13.68	76.07	49.12	81.58	46.56	68.70	65.65	74.05
2	44.44	14.53	37.72	14.91	32.82	14.50	15.27	16.79
3	23.93	5.98	8.77	2.63	14.50	13.74	13.74	6.87
4	11.97	2.56	4.39	0.88	5.34	3.05	5.34	1.53
5	5.98	0.85	0	0	0.76	0	0	0.76

J Pediatr. Author manuscript; available in PMC 2018 July 01.

Table 5

online. Distribution of Reproductive and Somatic Hormone Concentrations by Sex

Hormone	Min	5th percentile	25th percentile	50th percentile	75th percentile	95th percentile	Max
Girls (n=131)							
Estradiol (pg/ml)	4.2	9.8	16.5	23.0	43.1	95.7	483
Testosterone (ng/dl)	<2	<2	13.4	19.8	29.6	47.7	73.6
Inhibin B (pg/ml)	<10	<10	11.8	22.4	47.9	105	283
SHBG (nmol/L)	12.6	22.7	39.2	64.0	92.7	151	172
DHEA-S (μg/dl)	<15	<15	23.2	42.6	74.0	150	214
IGF-1 (ng/ml)	102	141	198	250	359	468	606
C-peptide (ng/ml)	0.39	0.66	1.00	1.60	2.20	4.10	10.1
Leptin (ng/ml)	2.4	3.6	6.3	10.7	18.5	34.3	62.2
Boys (n=117)							
Estradiol (pg/ml)	<3.8	9.1	13.3	16.7	20.1	32.8	83.5
Testosterone (ng/dl)	<2	<2	11.1	21.3	59.8	442	720
Inhibin B (pg/ml)	20.9	40.4	64.9	103	177	251	353
SHBG (nmol/L)	18.0	24.9	48.3	76.3	105	163	224
DHEA-S (μg/dl)	<15	<15	32.6	51.0	90.6	183	326
IGF-1 (ng/ml)	92.9	124	166	206	268	453	568
C-peptide (ng/ml)	0.46	0.67	0.96	1.30	1.80	4.30	7.70
Leptin (ng/ml)	1.4	2.2	3.6	6.5	10.7	21.8	34.2

Abbreviations: SHBG, sex hormone-binding globulin; DHEA-S, dehydroepiandrosterone sulfate; IGF-1, insulin-like growth factor 1.

Table 6

online. Cross Tabulation and Percent Agreement of Self-Assessed vs. Physician-Assessed Tanner Staging (Girls n=131, Boys n=114).

Girls: Tanner Staging for Breast Development

	Physician						% Agreement	
Self	**1**	**2**	**3**	**4**	**5**	**Total**	**Exact**	**±1 Stage**
1	58	2	1	0	0	61	67.4	96.5
2	25	15	3	0	0	43	75.0	100
3	3	3	10	3	0	19	55.6	94.4
4	0	0	4	3	0	7	42.9	100
5	0	0	0	1	0	1	n/a	n/a
Total	86	20	18	7	0	131	65.6	96.9

Girls: Tanner Staging for Pubic Hair Development

	Physician						% Agreement	
Self	1	2	3	4	5	Total	Exact	±1 Stage
1	90	0	0	0	0	90	92.8	100
2	7	12	0	0	0	19	54.5	100
3	0	10	7	1	0	18	77.8	100
4	0	0	2	1	1	4	50.0	100
5	0	0	0	0	0	0	0	100
Total	97	22	9	2	1	131	84.0	100

Boys: Tanner Staging for Genital Development

	Physician						% Agreement	
Self	1	2	3	4	5	Total	Exact	±1 Stage
1	12	4	0	0	0	16	21.4	76.8
2	31	16	4	0	0	51	37.2	74.4
3	10	12	3	2	0	27	30.0	90
4	3	6	2	2	0	13	40.0	100
5	0	5	1	1	0	7	n/a	n/a
Total	56	43	10	5	0	114	28.9	78.1

Boys: Tanner Staging for Pubic Hair Development

Girls: Tanner Staging for Breast Development

	Physician							% Agreement	
Self	**1**	**2**	**3**	**4**	**5**	**Total**	**Exact**	**±1 Stage**
	Physician						% Agreement	
Self	1	2	3	4	5	Total	Exact	±1 Stage
1	86	1	0	0	0	87	92.5	100
2	7	9	1	0	0	17	52.9	88.2
3	0	5	0	1	0	6	0.0	100
4	0	1	2	0	0	3	0.0	100
5	0	1	0	0	0	1	n/a	n/a
Total	93	17	3	1	0	114	83.3	98.2