Validity of Self-Assessment of Pubertal Maturation

Anna R. Rasmussen, MD, Christine Wohlfahrt-Veje, MD, PhD, Katrine Tefre de Renzy-Martin, MD, Casper P. Hagen, MD, PhD, Jeanette Tinggaard, MD, Annette Mouritsen, MD, PhD, Mikkel G. Mieritz, MD, Katharina M. Main, MD, PhD

abstract

BACKGROUND AND OBJECTIVES: Studies of adolescents often use self-assessment of pubertal maturation, the reliability of which has shown conflicting results. We aimed to examine the reliability of child and parent assessments of healthy boys and girls.

METHODS: A total of 898 children (418 girls, 480 boys, age 7.4–14.9 years) and 1173 parents (550 daughters, 623 sons, age 5.6–14.7 years) assessed onset of puberty or development of breasts, genitals, and pubic hair according to Tanner stages by use of a questionnaire and drawings. Physicians’ assessments were blinded and set as the gold standard. Percentage agreement, κ, and Kendall’s correlation were used to analyze the agreement rates.

RESULTS: Breast stage was assessed correctly by 44.9% of the girls (κ = 0.28, r = 0.74, P < .001) and genital stage by 54.7% of the boys (κ = 0.33, r = 0.61, P < .001). For pubic hair stage 66.8% of girls (κ = 0.55, r = 0.80, P < .001) and 66.1% of boys (κ = 0.46, r = 0.70, P < .001) made correct assessments. Of the parents, 86.2% correctly assessed onset of puberty in girls (κ = 0.70, r = 0.71, P < .001) and 68.4% in boys (κ = 0.30, r = 0.37, P < .001). Children who underestimated were younger and children who overestimated older than their peers who made correct assessments. Girls and their parents tended to underestimate, whereas boys overestimated their pubertal stage.

CONCLUSIONS: Pubertal assessment by the child or the parents is not a reliable measure of exact pubertal staging and should be augmented by a physical examination. However, for large epidemiologic studies self-assessment can be sufficiently accurate for a simple distinction between prepuberty and puberty.

WHAT’S KNOWN ON THIS SUBJECT: Many population-based studies including pubertal children are based on self-assessment of pubertal maturation, the reliability of which is uncertain.

WHAT THIS STUDY ADDS: Self-assessment is not reliable for precise pubertal staging. Simple distinctions between prepuberty and puberty showed moderate agreement with clinical examinations. Parents and girls tended to underestimate and boys to overestimate pubertal development by up to 50% and 30%, respectively.

University Department of Growth and Reproduction, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

Dr Rasmussen performed clinical examinations of children and was the main person responsible for data analysis and writing of the manuscript; Dr Wohlfahrt-Veje conceptualized, designed, and supervised the data collection and analysis, performed clinical examinations of the participants, participated in the drafting of the manuscript, and critically reviewed the final manuscript; Drs Hagen, Mouritsen, Tefre de Renzy-Martin, Tinggaard, and Mieritz participated in the puberty study design and performance of clinical examinations and revised the final manuscript; Professor Main conceptualized, designed, and supervised the data collection and analysis and critically revised the final manuscript; and all authors approved the final manuscript as submitted.


DOI: 10.1542/peds.2014-0793

Accepted for publication Sep 28, 2014

Address correspondence to Anna R. Rasmussen, MD, Department of Growth and Reproduction, Section 5064, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark. E-mail: anna.roe.rasmussen@sund.ku.dk

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2015 by the American Academy of Pediatrics
Pubertal development includes a multitude of physiologic and psychological changes, which strongly affect observations linked to outcome parameters such as biology, behavior, and intellectual performance. Thus, a study of older children and adolescents requires having a valid assessment of pubertal onset and preferably also maturation stages.

Pubertal development is traditionally classified into 5 stages for breast (B1–B5) and genital (G1–G5) development and pubic hair growth (PH1–PH5). In addition, testicular volume is usually measured, and a volume >3 mL by orchidometry is generally accepted as a marker of pubertal onset with testicular secretion of testosterone. Pubic hair usually follows testicular growth and breast development in boys and girls, respectively. The age at pubertal onset varies between genders and individuals and is known to be influenced by many factors such as ethnicity, nutrition, genetics, and environment.

For clinical or epidemiologic studies in which exact pubertal stages are required, many aspects must be taken into consideration. Some children may feel uncomfortable with the physical examination, and individual assessments are time consuming, logistically challenging, and expensive in large populations. In some cultures, physical examinations of healthy children may not be ethically acceptable. Many studies therefore rely on self-assessment questionnaires with pictures or questions or on the use of a pubertal development scale, where children are asked to rank their development.

Previous investigations of the reliability of self-assessment have shown conflicting results. Some studies found reasonable agreement between self-assessment and examination by a physician, whereas others found discrepancies. Most studies include few children (girls, n = 37–182; boys, n = 23–172) in view of the broad age range of puberty. We therefore aimed to validate self-assessment of sexual maturation compared with clinical examination by trained physicians in a large cohort of boys and girls.

Other studies have compared self-assessment with parental assessment or have used parental assessment as the only evaluation method. To our knowledge only 1 previous study of girls and mothers has compared parental assessment with physician’s examination. This study found reasonable correlation rates. We therefore aimed to examine the reliability of parental assessment also.

**METHODS**

**Study Population**

This study was based on data from an ongoing population-based mother–child cohort conducted in Copenhagen, Denmark. Mothers were recruited between 1997 and 2003 in early pregnancy from 3 university hospitals. Only white mothers of Danish origin were included in the cohort. The cohort has previously been described in detail.

Between 2010 and 2012 all 2647 children were invited for a longitudinal puberty follow-up, and 1284 (48.5%) agreed to participate (Fig 1). Children who did not participate in the follow-up study (n = 1363) did not differ significantly from the included children in gender, socioeconomic status, birth weight (weight for gestational age), or BMI (SD scores for age and gender) at any time during the examinations before the pubertal follow-up (all Ps ≥.2).

At the first examination, at median age 10.9 years for daughters (range 6.2–14.7) and 10.6 years for sons (5.6–14.2), parents (biological mother or father) were requested by questionnaire to assess pubertal development of their child before clinical examination (Table 1). A total of 111 children were excluded; 56 did not feel comfortable having a physical examination, and 23 never returned the questionnaire or returned the questionnaire ≥90 days before (n = 19) or after (n = 13) the examination date.

At the second examination, median age 11.8 years for daughters (range 7.4–14.9) and 11.4 years for sons (7.9–14.9), children were requested to self-assess pubertal development (Table 1). All 1284 children were invited, and 79.8% (n = 1025) agreed to participate. A total of 127 children were excluded; 60 did not feel comfortable having a physical examination, 46 never returned the questionnaire, 7 answered incorrectly, 2 did not want to answer the questionnaire, and the rest answered ≥90 days before (n = 7) or after (n = 5) the examination date.

The study was conducted according to the Helsinki II Declaration and approved by the local ethics committee (KF 01-030/97. KF 01276357, H-1-2009-074) and the Danish Data Protection Agency (1997-1200-074, 2005-41-5545, 2010-41-4757). The parents and children gave their written informed consent before examination.

**Clinical Examination**

Pubertal stages were assessed according to Tanner and Marshall by 6 trained physicians, and their ratings were set as the gold standard. The physicians were blinded to both parental and adolescent questionnaires. Palpation was used to differentiate between fat and breast tissue. If breast stage differed between left and right side, the highest stage was used for analysis.

The examiners in the study participated in repetitive workshops to ensure and maintain standardization. In a pilot study (n = 26), evaluation of breast stages (right side) was done independently by 2 examiners, and agreement for onset of breast development was 100%.
Interobserver agreement was 84.6% \( (n = 22) \), \( \kappa \) was 0.78 \( (P < .001) \), and the correlation (Kendall’s coefficient) was 0.87, \( P < .001 \).

Height was measured by using a wall-mounted stadiometer to the nearest millimeter (Holtain Ltd, Crymych, United Kingdom) and weight was measured to the nearest 0.1 kg using electronic scales (SECA \( \delta \) model 707, Hamburg, Germany; and Bisco model PERS 200, Farum, Denmark).

**Self-Assessment and Questionnaires**

Fourteen days before the first scheduled examination a questionnaire was sent to the parents. They were asked, “Is your child in puberty YES/NO.” They were also asked what the first signs of puberty had been by choosing between breast and/or pubic hair for girls and pubic hair for boys. Parents were not asked about their son’s genital stage. The questionnaire was returned before the examination.

Statistics

All statistical analyses were carried out using SPSS software (version 20.0; IBM SPSS Statistics, IBM Corporation). Outcomes were analyzed either as ordinal data (Tanner stages 1–5) or dichotomized (1 vs 2–5) according to prepubertal or pubertal status.

To examine the agreement between child or parental assessment and clinical examination, we used 3 approaches.

Cohen’s \( \kappa \) coefficient is a statistical measure of interrater agreement. Strength of agreement: <0.00, poor; 0.00 to 0.20, slight; 0.21 to 0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; >0.80, almost perfect.\(^{27}\)

Kendall’s \( \tau \)-b provides an estimate of the similarity of the ordering of data. Perfect agreement corresponds to a coefficient of 1, whereas total disagreement (1 ranking is the opposite of the other) corresponds to a coefficient of –1. Zero indicates the absence of association.

Sensitivity and specificity for parental and participant assessment in comparison with clinical examination were calculated on dichotomized data: prepubertal (Tanner 1) versus pubertal (Tanner 2–5). The sensitivity showed how accurately the children and parents assessed the presence of puberty or secondary sex characteristics (true positive), and the
specificity showed how accurate they were in assessing the absence this development (true negative).

Differences between population characteristics of included, excluded, and nonparticipating children were tested by Mann–Whitney U test (continuous variables) and $\chi^2$ (categorical variables). The first mentioned test was also used to analyze was also used to analyze differences in BMI and age between children who underestimated or overestimated and children who made correct assessments.

**RESULTS**

No significant differences were found for the girls between participants and excluded children with regard to age, BMI, height, and weight ($P > .1$). In boys, median age (11.2 years, range 7.5–13.6; $P = .004$), weight (38.0 kg, range 23.2–62.8; $P = .005$), and height (148.6 cm, range 122.2–172.7; $P = .004$) at first examination were significantly higher in excluded children. The agreements between physician assessment and self-assessment are shown in Table 2 (girls) and Table 3 (boys).
We found that 44.9% of the girls (n = 179) were able to correctly assess breast stage (Table 2). Independent of pubertal maturation, girls tended to underestimate breast development by 1 or 2 stages (52.9%, n = 211). Only 2.3% (n = 9) of girls overestimated their pubertal development by 1 or 2 stages. The highest agreement in breast stages was observed for breast stage B1, at 92.4% (n = 73).

Girls who underestimated breast stages B2 (median age = 11.1 vs 11.8 years, P = .006) and B3 (median age = 12.0 vs 12.4 years, P = .002), but not B4 and B5, were significantly younger than girls with correct assessments. Girls who overestimated breast stage tended to be older than girls who made the correct assessments for B1 (median age = 10.7 vs 10.1 years, P = .27) and B2 (median age = 12.0 vs 11.8 years, P = .69), but this was not significant.

When examining pubic hair stage, 66.8% (n = 265) of the girls made correct assessments (Table 2). Underestimation by 1 or 2 stages was found in 24.7% (n = 98), and overestimation was found in 8.6% (n = 34). The highest concordance was observed in PH1 (90.8%, n = 118) and the lowest in PH5 (25.0%, n = 5).

Girls who underestimated pubic hair stage were younger than those who evaluated correctly (PH2: median age = 11.2 vs 11.8 years, P = .082; PH3: median age = 12.0 vs 12.5 years, P = .003; PH4: median age = 12.4 vs 12.9 years, P = .12; and PH5: median age = 13.0 vs 14.0 years, P = .008). The girls who overestimated their pubic hair stage tended to be older than those who evaluated correctly, but this was not significant for any stages (data not shown). No significant association was found between BMI and accuracy of breast or pubic hair assessment (data not shown).

Boys’ Self-Assessments

Self-assessment of boys showed only fair agreement with clinical
examination both for onset of puberty and pubic hair: Sensitivity was high, and specificity was lower for assessment of both puberty and pubic hair development (Table 4). Among the boys 73.8% (n = 354) correctly assessed whether they were in puberty. Overestimation of pubertal stage was more likely than underestimation (17.1%, n = 82 vs 9.2%, n = 44, respectively, Table 3).

When examining genital stage, we found that 54.7% (n = 239) made correct evaluations. More boys (n = 127, 29.1%) overestimated themselves by 1 or 2 stages, and only 71 (16.2%) underestimated themselves. The highest agreement was found in G1 (58.4%, n = 129).

Younger ages at G2 (median age = 10.9 vs 11.7 years, P = .001) and G3 (median age = 12.4 vs 12.8 years, P = .073) were associated with underestimation of genital stage. The boys who overestimated genital staging were older than those who assessed correctly (G1: median age = 11.0 vs 10.3 years, P < .001; G2: median age = 12.6 vs 11.7 years, P = .017). There was no correlation between BMI and boys’ assessment of genital stage (data not shown).

Boys made correct assessments of pubic hair stage in 66.1% (n = 310), with a tendency to overestimate (n = 122, 26.0%) rather than underestimate (n = 37, 7.9%) by 1 to 2 stages (Table 3). The highest agreement was observed in PH5 (88.8%, n = 8) and the lowest in PH4 (44.4%, n = 12).

Younger age was associated with underestimation of pubic hair stage (PH2: median age = 11.8 vs 12.3 years, P = .035). Boys who overestimated pubic hair development were older than those who assessed correctly (PH1: median age = 11.3 vs 10.5 years, P < .001; PH2: median age = 12.8 vs 12.3 years, P = .29; and PH4: median age = 13.6 vs 13.1 years, P = .041). There was no systematic influence of BMI on pubic hair assessment in boys (data not shown).

**Parental Assessment of Puberty in Girls and Boys**

The agreement between clinical examination and parental assessment was stronger for parents evaluating daughters than sons. The agreement between evaluation of specific secondary sex characteristics (onset of breast development and pubic hair) was slight to fair for both genders (Table 4). The specificity for assessment of puberty was high in both daughters and sons (.94 and .95, respectively), but sensitivity was lower, in particular for puberty in sons (.33) and pubic hair (.61–.71) in both genders.

Of the parents, 86.2% (n = 473) assessed correctly whether their daughter was in puberty or not, 84.3% (n = 280) correctly assessed onset of breast development, and 69.6% (n = 229) pubic hair. Parents underestimated the onset of puberty in 11.8% (n = 65), breast development in 13.0% (n = 43), and pubic hair in 27.4% (n = 90). They overestimated the onset of puberty in 2% (n = 11), breast development in 2.7% (n = 9), and pubic hair development in 3% (n = 10).

In boys, 68.4% (n = 422) of parents correctly assessed whether they were in puberty, and 67.4% (n = 87) correctly assessed onset of pubic hair growth. They underestimated the onset of puberty in 28.7% (n = 177) and pubic hair development in 20.2% (n = 26), and they overestimated the onset of puberty in 2.9% (n = 18) and pubic hair development in 12.4% (n = 16).

**DISCUSSION**

This large study of 898 Danish children shows that self-assessment and parental assessment of pubertal development are inaccurate in a substantial number of participants when compared with clinical examination by trained physicians. Overall, children were slightly more accurate than their parents in assessment of whether they had entered puberty. However, half of the girls tended to underestimate their exact breast development stage, and one-quarter also underestimated pubic hair. In boys, the opposite was observed, with approximately one-third overestimating genital or pubic hair stage. Parents underestimated physical development in up to one-third of their children.

Our findings are in line with some previous publications, which found an agreement between physician and self-assessment ranging from 48.6% to 52.0% for breast stage, 53.3% to 64.0% for pubic hair in girls, and 27.0% to 49.0% for genital stage. Some studies have reported similar or slightly higher agreement for pubic hair in boys, ranging from 58.0% to 78.0%.

A few studies found high agreement rates for breast stage (86.0%) and pubic hair for girls and boys (80.0% to 93.0%). To our knowledge this study is the largest published to date, and evaluations by the child or parent and the examiner were blinded to pubertal assessment of the other party. One study had very few participants (girls 43, boys 23), and half of the children completed the questionnaires in front of the unblinded physician before the clinical examination. That may have introduced a systemic bias toward better agreement between ratings. Two previous studies used quadratic weighted $\kappa$, which takes into account the relative seriousness of disagreement, whereas the $\kappa$ measure used for our study differentiates only between agreement and disagreement.

Our participants have been followed since birth, and many families were therefore familiar with pubertal examinations because physicians had previously explained the procedures before they were asked to make pubertal self-assessments. Thus, we expected a higher agreement between clinical examination and self-assessment. However, the children may have encountered difficulties in
differentiating between the illustrations of pubertal stages (eg, the illustration of B1 could be misinterpreted as B2 if the explanatory text was ignored; Fig 2). The high percentage of misclassification by self-assessment indicates that studies in which pubertal development has a significant influence on the outcome must include standardized clinical examinations. Overall, it was easier for children to assess whether there were any signs of puberty than to evaluate the exact pubertal staging according to Tanner. Parents were better at assessing whether their children were in puberty than at describing the first physical signs of development. The physicians in our study had an intrarater variation for assessment of breast stage of \( \kappa = 0.78, P < .001 \), whereas the agreement for onset of breast development was 100%, probably because of the method of palpation of breast tissue. This finding is in agreement with earlier publications. 17,30

The girls tended to underestimate breast and pubic hair stage, which has also been shown in previous studies. 19,28,31 In contrast, the boys tended to overestimate genital and pubic hair stage, which has also been reported previously.18–20,31 Age had a significant influence on pubertal self-assessment, in that the children who underestimated physical development tended to be younger, whereas older children tended to overestimate. We therefore hypothesize that older prepubertal boys overestimate and younger pubertal girls underestimate because they had a preconceived expectation to be at the same development stage as their peers. Girls may also be misled by the assessment of breast development as bra size rather than breast shape. Only half of the boys in our study were able to assess their genital stage, which has also been shown previously. 15,28 One study using schematic drawings of Tanner stages similar to ours showed that 27% (\( n = 1150 \)) of the boys reported a lower genital stage than they reported 1–1.5 years earlier when asked to repeat their self-assessments. 9 This finding indicates that self-assessment is unreliable. Contrary to our expectations, BMI did not significantly and systematically influence self-assessment. High BMI and adiposity can make it difficult to differentiate lipomastia from breast tissue, which can lead to overestimation of breast development. 20 Most of our study participants were normal-weight children from social class 1 and 2. Social class was determined from educational level and self-reported occupational status in hierarchical order (1 being highest, 5 being lowest). 32 Our data may therefore not be applicable to children from other social classes or have enough statistical power to detect influences of high BMI on pubertal self-assessment.

The majority of our participants were in the early stages of puberty, so our study cannot determine whether self-assessment of puberty is more accurate once menarche or voice break has occurred. One previous study with girls in predominantly late pubertal stages 17 reported similar agreement rates with other studies examining children in early stages of puberty, 15,19 suggesting that pubertal stage may not influence agreement between self-assessment and clinical examination.

In conclusion, our data suggest that pubertal staging by children or their parents instead of physical examination leads to a substantial proportion of misclassification. Girls and parents tend to underestimate and boys to overestimate their development. Thus, studies of outcomes that are strongly dependent on precise staging of pubertal development must ensure correct assessment by standardized clinical examination. However, self-assessment may be sufficiently accurate for large epidemiologic studies in which only the distinction of prepuberty versus puberty is important.

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: The study was supported by a grant from the Danish Agency for Science, Technology and Innovation (grant 09-067180).

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

5. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon...


Validity of Self-Assessment of Pubertal Maturation
Anna R. Rasmussen, Christine Wohlfahrt-Veje, Katrine Tefre de Renzy-Martin, Casper P. Hagen, Jeanette Tinggaard, Annette Mouritsen, Mikkel G. Mieritz and Katharina M. Main

Pediatrics 2015;135;86; originally published online December 22, 2014; DOI: 10.1542/peds.2014-0793
Validity of Self-Assessment of Pubertal Maturation
Anna R. Rasmussen, Christine Wohlfahrt-Veje, Katrine Tefre de Renzy-Martin, Casper P. Hagen, Jeanette Tinggaard, Annette Mouritsen, Mikkel G. Mieritz and Katharina M. Main
Pediatrics 2015;135;86; originally published online December 22, 2014;
DOI: 10.1542/peds.2014-0793

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/135/1/86.full.html