Journal of Pediatric Endocrinology & Metabolism, 22, 291-299 (2009)

# Assessing Inter-Rater Reliability (IRR) of Tanner Staging and Orchidometer Use with Boys: A Study from PROS

Eric J. Slora[1], Alison B. Bocian[1], Marcia E. Herman-Giddens[2], Donna L. Harris[1], Steven E. Pedlow[3], Steven A. Dowshen[4] and Richard C. Wasserman[1,5]

[1]*Pediatric Research in Office Settings, American Academy of Pediatrics, Elk Grove Village, IL,*
[2]*University of North Carolina School of Public Health, Chapel Hill, NC,*
[3]*NORC, University of Chicago, Chicago, IL,* [4]*A.I. Du Pont Hospital for Children, Wilmington, DE and*
[5]*Department of Pediatrics, University of Vermont College of Medicine, Burlington, VT, USA*

## ABSTRACT

***Background:*** **Few studies have systematically assessed the reliability of pubertal markers; most are flawed by limited numbers of markers and ages studied.**
***Aim*: To conduct a comprehensive examination of inter-rater reliability in the assessment of boys' sexual maturity.**
***Subjects:*** **Eight pairs of practitioners independently rated 79 consecutive boys aged 8-14 years.**
***Methods*: Two raters in each of eight practices independently rated boys aged 8-14 years, presenting for physical examinations, on key pubertal markers: pubic hair and genitalia (both on 5-point Tanner scales), testicular size (via palpation and comparison with a four-bead Prader orchidometer), and axillary hair (via a three-point scale).**
***Results:*** **Intraclass correlations assessing degree of inter-rater reliability for pubertal markers ranged from 0.61 to 0.94 (all significant at p <0.001). Rater Kappas for signs of pubertal initiation ranged from 0.49 to 0.79.**
***Conclusions:*** **Practitioners are able to reliably stage key markers of male puberty and identify signs of pubertal initiation.**

## KEY WORDS

secondary sexual characteristics, puberty, primary care, practice-based research network (PBRN)

## INTRODUCTION

The reliability of Tanner staging by clinicians conducting cross-sectional or longitudinal assessments of pubertal status (as opposed to patient self-rating) has not been questioned in most puberty studies or even in discussions of ratings of pubertal status[1-3]. Clearly, however, a consistent and reliable method for Tanner staging must be established as a precursor to data collection, since inter-rater reliability is a concern in both clinical and research settings[2]. Most puberty studies have used an unspecified number of examiners without any validation of sexual maturity staging[4-8]. Puberty data for the US National Health and Nutrition Examination Surveys were collected without validating Tanner stage assignments[9]. Several studies have attempted to address the problem of potential differences among raters; however, these studies have involved only one to several raters, and the issue of inter-observer agreement was not addressed[10-12].

Few studies have directly examined inter-observer agreement in Tanner staging of males. Carlsen *et al.* found "great variation" among raters, although only the most mature Tanner stages (stages 4-5) were considered in this assessment since the study population was comprised of 23 adult men[13]. Espeland *et al.* found examiner over- or under-estimation rates of up to 19% among approximately 250 boys in each of the four pubertal

Reprint address:
Eric J. Slora, Ph.D.
American Academy of Pediatrics
141 Northwest Point Blvd
Elk Grove Village, IL 60007
USA

Tanner stages in their Cooperative Study of Sickle Cell Disease[14]. No "central training" was provided for the examiners, however. In a Swedish study on self-assessment of sexual maturity, an unspecified number of clinicians without specific training in staging examined 100 students and found "excellent inter-rater agreement (88%, 95% CI 84-92%)[15]. For the most part, it has apparently been assumed that unvalidated clinician Tanner staging can be accurately performed by examiners and used as the "gold standard" for rating pubertal development[2].

Although Tanner staging is limited to only visual inspection of the subject's secondary sexual characteristics and comparison of the findings with standard photographs and descriptions, assessment of peripubertal testicular volume has now become accepted by pediatric endocrinologists and other experts as the most objective, reliable, and practical method for detecting the physical onset of central puberty in boys. In the distant past, testicular volume was assessed by comparing the testis to common objects, such as beans or grapes[16]. In the early part of the 20th century, techniques included measurements with calipers, graded series of ellipses, and Schonfeld's rubber models of actual testes (the first orchidometer). Early studies recognized problems in over- and underestimation depending on the technique used as well as errors introduced by the thickness of the scrotal skin and difficulties in distinguishing the boundary between the testis and epididymis[16,17].

Orchidometer use became common after Prader introduced his series of graded rotation ellipsoid "beads" of volumes ranging from 1 to 25 ml which were compared to the patient's testes[18]. Prader did not examine inter- or intra-rater reliability, nor has this been done in many pediatric investigations of testicular size[6,12,19]. Zachmann *et al.* in a small study involving comparisons of the measurement results obtained by an unspecified number of examiners with the results obtained by one standard examiner in 21 measurements in six boys, found the correlation coefficient to be +0.81[6]. Mul *et al.* assessed agreement between two observers for measurements of testicular volume in 79 boys as part of their puberty study in The Netherlands[4]. They found the Spearman correlation coefficient to be 0.82 using the complete set of orchidometer beads, 1-25 ml. Rivkees *et al.*, in a report involving the measurements obtained by three clinicians examining 12 boys with central precocious puberty, found that their interobserver coefficient of variation for orchidometer estimates averaged 20.4 ± 10.2%[20]. A prospective Turkish study of 50 boys that investigated the agreement among three examiners using the Prader orchidometer found a high correlation (r = 0.95-0.97) using the Pearson correlation statistic[17].

Despite the greater consistency in results obtained using orchidometers, there have still been problems with the standardization of studies on testicular volume due to the lack of consistency in rating testes when sizes fell between the standard bead volumes. Some studies always assigned the lower size[17,21], while others assigned the size that seemed closest[7,12,13], allowed a half size assignment[22], or did not specify[5,6,18,20]. There is, therefore, some preliminary evidence suggesting that adequate inter-rater reliability of sexual maturity staging can be accomplished, but that evidence is based on the results of studies limited by their use of small samples, statistical tests not suitable for assessment of inter-rater reliability, and different standards and instruments for measurement of testicular volume.

Other testicular volume measurement techniques have been developed and evaluated, including the Seager calipers, the Takihara elliptical punched-ring orchidometer, and ultrasonography[20-22]. Evidence suggests that ultrasonography is the most accurate method, in part due to the enhanced ability to define the boundaries between the testis and surrounding extraneous tissues. However, due to its cost, issues of patient acceptance, and time required, ultrasonography is not used routinely for this purpose clinically or in large-scale pubertal staging studies. The available literature suggests that, while not as sensitive as ultrasonography, the Prader orchidometer is an acceptable and useful alternative for the determination of the size of individual testes[20,22]. The Prader orchidometer is now generally accepted as the most clinically practical and economical tool and it has been used in many studies worldwide involving the measurement of boys' testicular size at various pubertal stages[1,7,17,18,20].

In summary, while a limited number of studies have examined aspects of the measurement of pubertal markers, most are flawed by issues such as the limited number of pubertal markers studied, the limited numbers of raters and subjects, the collection of adult data not applicable to pediatric studies, and the use of data obtained from groups of study subjects that may not generalize to larger populations. To date, no study has systematically examined inter-rater reliability between two examiners of the generally accepted full set of key markers of puberty in populations of boys encountered in primary care settings.

As part of a larger Pediatric Research in Office Settings (PROS) project to determine the stages of secondary sexual development and mean ages of onset of pubertal testicular enlargement among boys seen in pediatric practices in the United States, the aim of this study was to conduct a systematic examination of inter-rater reliability in the assessment of sexual maturity stages in boys using Tanner staging and the Prader orchidometer.

## METHODS

### Practitioner study population and setting

This study was conducted by PROS, the national pediatric primary care practice-based research network of the American Academy of Pediatrics (AAP), Elk Grove Village, Illinois[24]. At the time of the study, PROS included almost 2,000 practitioners in more than 750 practices from 49 states, Puerto Rico, the District of Columbia, and Canada.

The study was publicized at a semiannual meeting of the PROS network. Recruitment materials, including an overview of the study protocol, were given to 38 eligible meeting attendees and volunteers were requested by the study team. Of those meeting attendees, nine PROS practitioners agreed to participate in the study. Thereafter, the volunteers received a participation fax-back sheet, reiterating the responsibilities of the study and confirming their willingness to participate. The initial nine volunteers were asked to recruit another PROS practitioner in their practice who was willing to be one of two raters needed for the study. One of the original nine volunteers subsequently declined to participate and ultimately eight practitioner pairs participated in data collection. Participants were from a geographically dispersed sample of eight non-solo, English-speaking PROS practices in seven states.

### Procedures

The Institutional Review Board of the AAP approved the study protocol. The study protocol required written informed consent from parents or guardians and verbal assent from all boys for study participation.

Practitioner members of the research team (in consultation with state practitioner representatives of PROS during semiannual national network meetings) developed the protocol and materials for this study as well as the larger PROS study, including a training manual to instruct clinicians in the assessment of secondary sexual characteristics and pubertal maturation in boys by visual inspection and palpation[25].

Data collection occurred from October 2004 through November 2005. The study occurred in two phases. In phase 1 (training), a two-part qualifying examination was used to 1) assess competence in rating the stages of male secondary sexual development (Tanner staging), 2) evaluate knowledge regarding the correct use of a four-bead modification of the Prader orchidometer - a set of beads used to measure testicular volumes (see Fig. 1 for photograph of modified orchidometer), and 3) to assure the accuracy of puberty ratings. This was accomplished by requiring the practitioners to study the training manual on the assessment of sexual maturity stages in boys developed for the study[25] and pass two examinations to qualify before enrolling boys in the study. Part I of the examination consisted of multiple choice questions drawn from material in the training manual such as Tanner staging and measuring testicular volumes. Part II involved a set of 12 photographs of male genitalia and pubic hair (four of which were taken from the classic van Wierengen *et al.* collection[26], three from the original Reynolds and Wines article[27], and five unpublished photographs from the collections of three pediatric endocrinologists) to be rated for Tanner stage. The sexual maturity stages for the

Tanner and van Wierengen photographs were as stated by those authors. The ratings for the unpublished photographs were validated through assessment and agreement by the study's consultant panel of five pediatric endocrinologists. To pass the examinations, practitioners were required to score at least 90% on Part I and 87.5% on Part II. Practitioners were permitted to refer to the training manual during the examination and retake either or both parts of the examination as many times as necessary to pass. Once both practitioner partners received passing scores, patient enrollment materials were sent to their practice. Practitioners received Continuing Medical Education (CME) credits for passing the qualifying examination.

In phase 2 (data collection for inter-rater reliability assessment), two raters in each of eight PROS practices independently rated consecutive boys aged 8-14 years presenting for complete physical examinations on key pubertal markers: pubic hair and genitalia (both on 5-point ordinal Tanner scale), left and right testicular size (rated via palpation of the testicle and comparison with a four-bead orchidometer [≤1, 2, 3, or ≥4 ml] held in the rater's other hand), and axillary hair (three-point ordinal scale: none, sparse, or mature). Only the first four beads, 1 to 4 ml, of the Prader orchidometer were used to simplify the examiners' task, because measurement of testicular volume beyond the size indicating the onset of central puberty (>3-4 ml) was not relevant to the study. (Post-pubertal testicular volume has not been correlated with specific Tanner stages.)

Each practitioner pair was asked to conduct 10 independent assessments. The practitioners decided the order of their assessments and how they would introduce the second examiner into the examination room. Each practitioner performed an independent assessment without sharing findings with the other practitioner. Data were obtained for a total of 79 boys. Study practices kept recruitment logs to track patient demographic information and participation status (consent or decline). At the end of each week, the practice staff sent completed study forms in sealed envelopes to the PROS central office along with a copy of any completed recruitment logs. Overall, 67.5% of eligible boys were enrolled during data collection. Of the 38 boys who declined participation, the mean age was 11.4 years (significantly higher than the mean of 10.4 years for participants). Also noteworthy were the differences in minority/non-minority distribution among decliners versus participants. While the difference did not achieve statistical significance, the decliners were disproportionately white (74% versus 55% for participants; $p = 0.054$).

### Statistical analysis

Descriptive analyses were done to characterize practitioners' and patients' demographics and to establish a comparative baseline. Intraclass correlation coefficients (ICCs) based on multiple patient observations were generated to assess the reliability of pubertal stage assignment by practitioners within a given location. Further, Kappas were calculated to assess rater agreement (versus chance) regarding stages representing initiation of pubertal maturation. Data analysis was done using SPSS version 14.0 (SPSS Inc., Chicago, IL).

## RESULTS

### Practice characteristics

The eight practices that completed data collection for the study represented a variety of practice settings. As Table 1 indicates, the majority of practices were from suburban locations and urban, non-inner-city settings. The sample was geographically dispersed as well, representing all four US Census regions, with a slight plurality from the West (37.5%).

### Practitioner characteristics

Participating practitioners included 14 pediatricians (88%), one pediatric nurse practitioner and one physician assistant, of whom 75% were male. The majority of the practitioners were white (13/16), with two Asians and one African-American. The mean age of the clinicians was 46.6 years.

### Subject characteristics

A total of 79 boys were assessed in this study. Their mean age was 10.4 years, and, as Table 2


Fig. 1: Modified four-bead Prader orchidometer.

illustrates, the group was racially/ethnically diverse, including 18.2% African-Americans and 16.9% Hispanics. The sample also included a relatively high percentage of Medicaid patients. Slightly more than a quarter of the patients had a chronic disease, with almost half of those patients reporting asthma. None of the 79 boys examined in this study presented for growth or pubertal concerns.

**Reliability of pubertal stage assignment**

Intraclass correlations were used to assess the degree to which the rater pairs were able to independently and reliably stage various markers of

TABLE 1

Practice characteristics (n = 8)

| | |
|---|---|
| **Practice location** | |
| Urban, inner city | 0.0% |
| Suburban | 37.5% |
| Urban, non-inner city | 37.5% |
| Rural | 25.0% |
| **Geographic region** | |
| West | 37.5% |
| South | 25.0% |
| New England | 25.0% |
| Midwest | 12.5% |

TABLE 2

Patient characteristics (n = 79)

| | |
|---|---|
| **Race** | |
| White | 54.5% |
| African-American | 18.2% |
| Asian | 11.7% |
| Hawaiian/Pacific Islander | 2.6% |
| Multiracial | 13.0% |
| **Ethnicity** | |
| Hispanic | 16.9% |
| Non-Hispanic | 83.1% |
| **Payment source** | |
| Medicaid | 38.9% |
| Non-Medicaid | 61.1% |
| **Chronic disease present** | |
| Yes | 27.8% |
| No | 72.2% |
| **Type of chronic disease** | |
| Asthma | 45.4% |
| ADHD | 13.6% |
| Other | 41.0% |

boys' puberty. Variables assessed included left and right testicular size (as measured via a four-bead orchidometer, with beads ranging from 1 to 4 ml in size and ratings that included ≤1, 2, 3, and ≥4 ml), axillary hair (as measured on a three-point ordinal scale previously employed in another study[28] with verbal anchors corresponding to scale values of 1 [none], 2 [sparse], and 3 [mature], as well as genitalia and pubic hair [both measured using the marker-specific traditional five-point ordinal Tanner scales]). As Figure 2 indicates, all of the intraclass correlations were highly significant, with values ranging from 0.61 (left testicular size) to 0.94 (pubic hair), and p-values all exceeding $p < 0.001$.

**Reliability of ratings for initiation of pubertal maturation**

Kappa statistics were employed to examine another facet of inter-rater-reliability – the extent to which raters were able to independently and reliably identify when subjects had begun to demonstrate signs of pubertal maturation for each marker. For this analysis, the scale values previously described were each dichotomized – with Stage 1 (prepubertal) of each scale compared with a derived, collapsed Stage 2, representing all other values (pubertal) on the multivalue scale. As Figure 3 shows, Kappas ranged from 0.49 to 0.79, indicating generally very good agreement for

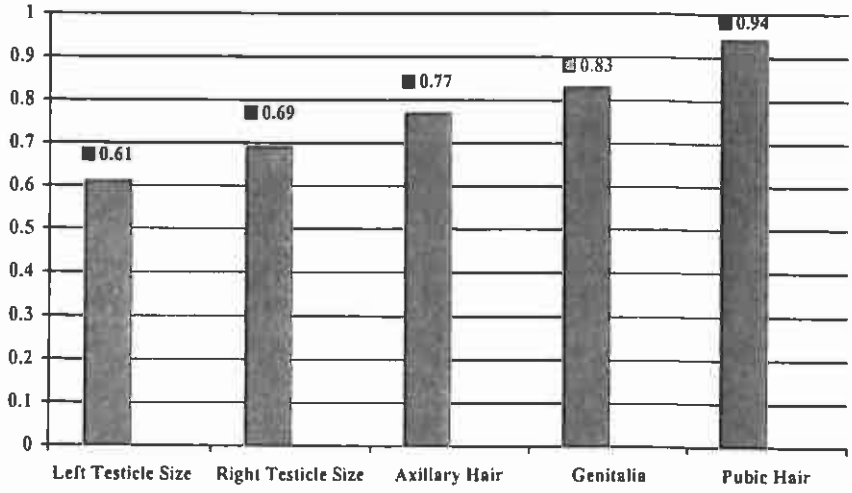


Fig. 2: Intraclass correlations for key pubertal markers.

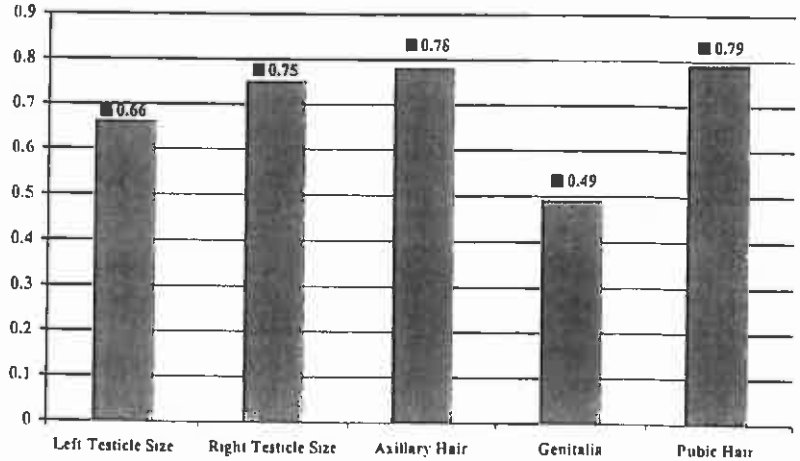


Fig. 3: Kappas for key pubertal markers.

identification of the initiation of pubertal maturation. The Kappa for the staging of genitalia was somewhat lower than those for other markers, but still in a range generally considered adequate[29].

## DISCUSSION

This study provides the first systematic look at the reliability of staging key markers of boys' puberty. Moreover, it provides a first look at staging of male genital maturity through use of a modified Prader orchidometer, appropriate for use in practices that see pediatric populations. The intraclass correlations were all highly significant, and the staging of pubertal initiation, as measured by Kappas, was good as well.

Of note, however, was that the Kappa obtained for examination of genitalia was lower than those obtained for the other markers. This finding is consistent with the subjective nature of visual inspection, as compared with more objective measures such as testicular volume measurement. While still in part visually based, the latter method includes palpation as a critical aspect of the method, along with a referent instrument, the Prader orchidometer, for measurement. The establishment of adequate inter-rater reliability of these methods not only supports their clinical utility, but also provides a sound foundation for the use of these methods in studies of boys' puberty.

Several methodological issues are noteworthy. First, our practitioner and patient study samples may not have been representative of US ambulatory practice since they were comprised of volunteer practitioners and patients within a practice-based research network. Recent literature[30] suggests that PROS is reasonably representative of US ambulatory pediatric practice as a whole, but this subsample of practices was self-selected from within the network and contained demographic biases – both with respect to practitioners (e.g., race/ethnicity) and practices (e.g., practice location) that could limit the generalizability of the findings. Moreover, those practitioners willing to participate in such a study may have been more interested in or confident regarding their knowledge of the topic than others and may therefore have been more motivated and skilled at pubertal stage assessment. Second, as noted earlier, demographically our sample of patient participants was younger than those who declined participation, and also more likely to be members of minority groups. Third, our sample was relatively small. While this study is a clear improvement in sample size over its predecessors, it is unclear whether the findings would be replicated in a larger sample of patients and providers. Fourth, establishment of inter-rater reliability with respect to testicular volume is confined to 1 to 4 ml. Last, it is worthy of note that the practitioners who participated in the study had to complete specific training in sexual maturity staging in order to qualify for participation. Any assumption that such inter-rater reliability findings might be obtained in a general sample of pediatricians may be unwarranted.

Keeping these methodological cautions in mind, the present study provides the first methodologically rigorous and comprehensive evidence that with appropriate training, office-based practitioners are able to reliably stage key markers of puberty in males and accurately identify the signs of initiation of puberty as reflected in that staging. The findings are important in that they provide assurance for trained practitioners performing such assessments in clinical settings, and from a research perspective, strong validation of the use of this methodology for the staging of puberty in future studies.

## ACKNOWLEDGEMENTS

This study was supported by grants from the Genentech Center for Clinical Research and Education, Grant #MCJ-177022 from the Health Resources and Services Administration Maternal and Child Health Bureau, Georgia Health Foundation, and Pfizer, Inc.

We especially appreciate the efforts of the PROS practices and practitioners, as well as those families and patients who agreed to participate in the study. The pediatric practices that participated in this study are listed by American Academy of Pediatrics Chapter. The listing of participants' names does not imply their endorsement of the data and conclusions. Alaska: Anchorage Pediatric Group LLC (Anchorage); Hawaii: Children's Medical Association Inc. (Aiea); Iowa: University

of Iowa (Iowa City); Massachusetts: Burlington Pediatrics (Burlington), Mary Lane Pediatric Associates (Ware); South Carolina: MUSC Pediatric Primary Care (Charleston); Virginia: Alexandria Lake Ridge Pediatrics (Alexandria); Washington: Central Washington Family Medicine (Yakima).

## CONFLICT OF INTEREST STATEMENT

The authors of this manuscript hereby affirm that we do not have any affiliation, financial agreement, or other involvement with any company or organization with a financial interest in the subject matter in the submitted manuscript. We have received unrestricted research grant funding for this project from Genentech Center for Clinical Research and Education, and Pfizer, Inc.

## REFERENCES

1. Dorn L, Dahl R, Woodward H, Biro F. Defining the boundaries of early adolescence: a user's guide to assessing pubertal status and pubertal timing in research with adolescents. Appl Dev Sci 2006; 10: 30-56.
2. Coleman L, Coleman J. The measurement of puberty: a review. J Adolesc 2002; 25: 535-550.
3. Rockett JC, Lynch CD, Buck GM. Biomarkers for assessing reproductive development and health: Part 1–Pubertal development. Environ Health Perspect 2004; 112: 105-112.
4. Mul D, Fredriks AM, van Buuren S, Oostdijk W, Verloove-Vanhorick SP, Wit JM. Pubertal development in the Netherlands 1965-1997. Pediatr Res 2001; 50: 479-486.
5. Biro F, Lucky A, Huster G, Morrison JA. Pubertal staging in boys. J Pediatr 1995; 127: 100-102.
6. Zachmann M, Prader A, Kind HP, Hafliger H, Budliger H. Testicular volume during adolescence: cross-sectional and longitudinal studies. Helv Paediatr Acta 1974; 29: 61-72.
7. Largo RH, Prader A. Pubertal development in Swiss boys. Helv Paediatr Acta 1983; 38: 211-228.
8. Foster T, Voors A, Webber L, Frerichs R, Berenson G. Anthropometric and maturation measurements of children, ages 5 to 14 years, in a biracial community-the Bogalusa Heart Study. Am J Clin Nutr 1977; 30: 582-591.
9. Herman-Giddens M, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the National Health and Nutrition Examination Survey III, 1988-1994. Arch Pediatr Adolesc Med 2001; 155: 1022-1028.
10. Lee P. Normal ages of pubertal events among American males and females. J Adolesc Health Care 1980; 1: 26-29.
11. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child 1970; 45: 13-23.
12. Taranger J, Engstrom I, Lichtenstein H, Svennberg-Redegren I. VI. Somatic pubertal development. Acta Paediatr Scand Suppl 1976; 258: 121-135.
13. Carlsen E, Andersen AG, Buchreitz L, et al. Inter-observer variation in the results of the clinical andrological examination including estimation of testicular size. Int J Androl 2000; 23: 248-253.
14. Espeland M, Gallagher D, Tell G, Davidson L, Platt O. Reliability of Tanner stage assessments in a multicenter study. Am J Hum Biol 1990; 2: 503-510.
15. Berg-Kelly K, Erdes L. Self-assessment of sexual maturity by mid-adolescents based on a global question. Acta Paediatr 1997; 86: 10-17.
16. Schonfeld W, Beebe G. Normal growth and variation in the male genitalia from birth to maturity. J Urol 1942; 48: 759-777.
17. Karaman MI, Kaya C, Caskurlu T, Guney S, Ergenekon E. Measurement of pediatric testicular volume with Prader orchidometer: comparison of different hands. Pediatr Surg Int 2005; 21: 517-520.
18. Prader A. Testicular size: assessment and clinical importance. Triangle 1966; 7: 240-243.
19. Matsuo N, Anzo M, Sato S, Ogata T, Kamimaki T. Testicular volume in Japanese boys up to the age of 15 years. Eur J Pediatr 2000; 159: 843-845.
20. Rivkees SA, Hall DA, Boepple PA, Crawford JD. Accuracy and reproducibility of clinical measures of testicular volume. J Pediatr 1987; 110: 914-917.
21. Diamond DA, Paltiel HJ, DiCanzio J, Zurakowski D, Bauer SB, Atala A, Ephraim PL, Grant R, Retik AB. Comparative assessment of pediatric testicular volume: orchidometer versus ultrasound. J Urol 2000; 164: 1111-1114.
22. al Salim A, Murchison PJ, Rana A, Elton RA, Hargreave TB. Evaluation of testicular volume by three orchidometers compared with ultrasonographic measurements. Br J Urol 1995; 76: 632-635.
23. Chipkevitch E, Nishimura RT, Tu DG, Galea-Rojas M. Clinical measurement of testicular volume in adolescents: comparison of the reliability of 5 methods. J Urol 1996; 156: 2050-2053.
24. Wasserman R, Slora E, Bocian A. Pediatric Research in Office Settings (PROS): a national practice-based research network to improve children's health care. Pediatrics 1998; 102: 1350-1357.
25. Herman-Giddens M, Bourdony C, Dowshen S. Assessment of sexual maturity stages in boys. Elk Grove Village, IL: Unpublished manual. American Academy of Pediatrics, 2005.

26. van Wieringen JC, Wafelbakker F, Verbrugge HP, De Haas JH. Growth diagrams 1965. In: Yen SSC, Jaffe RB, eds. Reproductive Endocrinology: Physiology, Pathophysiology and Clinical Management, 2nd Ed. Philadelphia, PA: WB Saunders Co, 1978.
27. Reynolds EL, Wines JV. Physical changes associated with adolescence in boys. Am J Dis Child 1951; 82: 529-547.
28. Herman-Giddens M, Slora E, Wasserman R, 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: 505-512.
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-174.
30. Slora E, Thoma K, Wasserman R, Pedlow S, Bocian A. Patient visits to a national practice-based research network: comparing Pediatric Research in Office Settings with the National Ambulatory Medical Care Survey. Pediatrics 2006; 118: e228-234.