# Concordance of testicular measurement in male adolescents with three methods of orchidometry

Gonzalo Agüero, M.D.<sup>a</sup> and Enrique Berner, M.D.<sup>a</sup>

#### ABSTRACT

*Introduction.* The Prader orchidometer is the standard method used to measure testicular volume (TV) in children and adolescents.

*Objective.* To assess the concordance in the estimation of TV and puberty onset with the Prader, Chipkevitch, and Sotos orchidometric techniques.

*Methods.* Cross-sectional descriptive study conducted among male children and adolescents aged 9-20 years. For each adolescent, TV was measured with the methods by Prader (gold standard), Chipkevitch (graphic model), and Sotos (measurement of testicular width with a plastic ruler and use of a formula equivalent to the ellipsoid equation). Male children and adolescents with urogenital conditions and disorders affecting testicular growth were excluded. Kappa statistics was used to determine concordance among methods for puberty onset, and intraclass correlation coefficient (ICC) and Bland-Altman (B&A) plots for TV.

*Results*. In total, 377 healthy males were included. Regarding the concordance for TV (mL), the Prader-Chipkevitch comparison obtained an ICC of 0.994 and a p < 0.001; while the Prader-Soto comparison obtained an ICC of 0.312 and a p < 0.001. With the B&A plots, mean differences were close to 0 mL in the Prader-Chipkevitch comparison and close to 8 mL in the Prader-Sotos comparison. Concordance for puberty onset obtained a kappa value of 0.93 and 0.75 in the Prader-Chipkevitch and Prader-Sotos comparisons, respectively.

*Conclusion.* The Prader and Chipkevitch orchidometers show an excellent concordance in estimating TV and puberty onset; therefore, both methods could be used interchangeably in the daily care of male adolescents. The Sotos method showed a high concordance in estimating pubertal onset, but low in measuring TV. *Key words:* puberty; testis; adolescent growth;

adolescent medicine.

http://dx.doi.org/10.5546/aap.2021.eng.251

To cite: Agüero G, Berner E. Concordance of testicular measurement in male adolescents with three methods of orchidometry. *Arch Argent Pediatr* 2021;119(4):251-258.

## INTRODUCTION

The measurement of testicular volume (TV) in children and adolescents is a clinical indicator that allows to assess pubertal development and the testicular component of urogenital disorders, such as varicocele, and of systemic diseases on the gonadal function.<sup>1,2</sup> Several scientific societies recommend an annual genital exam for male children and adolescents<sup>3,4</sup> to determine Tanner stages and estimate TV with the Prader orchidometer. This tool is the gold standard in clinical orchidometry and in the detection of male puberty onset.5,6

Concordance is the extent to which two or more observers, methods or techniques agree on the same phenomenon being measured. It gains relevance when considering whether it is possible to obtain equivalent results with a method or instrument different from the one used regularly, so that, if necessary, one or the other can be replaced or interchanged for ease of use, cost, and safety, among others.<sup>7</sup>

Our objective was to evaluate the concordance in the estimation of TV and puberty onset with the Prader, Chipkevitch,<sup>2</sup> and Sotos<sup>1</sup> orchidometry, being the Prader method the gold standard.

## POPULATION AND METHODS

Design: Descriptive, cross-sectional study conducted in a public hospital from June 1<sup>st</sup>, 2018 to February 28<sup>th</sup>, 2020.

Population: Male children and adolescents.

**Inclusion criteria**: Argentine male children and adolescents aged 9-20 years, with no relevant medical history, who attended primary health

a. Department of Adolescence, Hospital de Agudos "Dr. Cosme Argerich," Autonomous City of Buenos Aires, Argentina.

*E-mail address:* Gonzalo Agüero, M.D.: aguerogonzalo@gmail. com

*Funding:* SAP research fellowship awarded in 2018.

*Conflict of interest:* None.

Received: 10-31-2020 Accepted: 2-8-2021 FIGURE 1. Measurement of testicular volume with three orchidometric techniques



A: Prader orchidometer; B: Chipkevitch graphic method. Techniques A and B allow to compare the testicular major or long axis; therefore, the orchidometer beads or the card image should be parallel to the testicular long axis and as close as possible to it; C: the Sotos method; with this technique, it is possible to measure the testicular width, and the visual measurement should be done at a distance of 30-40 cm to neutralize parallax.

Source: Gonzalo Agüero, M.D., is the author of the charts.

care centers for a health check-up.

**Exclusion criteria**: Pubertal alterations, hypogonadism, congenital urogenital anomalies, scrotal abnormalities, chronic conditions that affect testicular growth, intellectual disability, ongoing acute illness, and refusal to participate in the study. If adolescents met the inclusion criteria, but in the genital exam any of the aforementioned conditions were detected, they were withdrawn from the study.

## Outcome measures

**Tanner staging.**<sup>8-10</sup> It classifies pubertal development into 5 stages, according to morphological modifications of the external genitalia (G) and the development of pubic hair (PH). Stage 1 is considered prepubertal; stages 2-4, puberty in progress; and stage 5, full development.

Testicular volume (mL). The volume of both testes was measured by 3 orchidometric techniques. The Prader orchidometer is used to estimate TV by comparative palpation with an ellipsoid model of known volume.<sup>11,12</sup> It is the most widely used orchidometer to measure the volume of testes, it complements Tanner staging, and is particularly useful for discriminating prepubertal from peripubertal development in early adolescents.<sup>5,6</sup> The Chipkevitch visual method<sup>2,13</sup> proposes the comparison between the testis and a card with 6 oval beads of known proportions. And finally, with the Sotos method,<sup>1,14</sup> testicular width measurement is performed using a variant of the ellipsoid formula that matches the ultrasound measurement (*Figure 1* and *Table 1*). Ultrasound is considered a more accurate way to estimate TV, although it is not a practical approach for the daily care of children and adolescents; therefore, its use is reserved for patients with urogenital problems.<sup>1,14</sup>

**Onset of puberty.** Male puberty begins upon reaching a TV  $\geq$  4 mL as per the Prader orchidometry; this value corresponds to an ultrasound volume  $\geq$  1.4 mL.<sup>15</sup> Since no cut-off value was described by the authors for puberty onset, a TV  $\geq$  1.4 mL is assumed in the case of Sotos (his method matches ultrasonography) and  $\geq$  3.5 mL in the case of Chipkevitch (value closer to that indicated by the Prader orchidometry, please refer to *Table 1*). The onset of puberty is measured in yes/no terms.

**Sampling.** The confidence interval (CI) for a mean was estimated on the assumption that the greatest variability in TV obtained by the Prader orchidometry is reached at 14 years old (mean: 15 mL, standard deviation [SD]: 7 mL),<sup>16</sup> with a 95 % CI and an accuracy of 2 mL; 50 adolescents are required for every Tanner stage (250 male adolescents and 500 testes). Participants were included sequentially by simple random sampling, limited by the number of daily consultations performed by male adolescents.

**Procedures.** The principal investigator assessed all participants as follows: Tanner staging; palpation of the scrotal contents to rule out testicular, spermatic cord or epididymal conditions; and measurement starting with

the right testis and using the techniques in the following order: Chipkevitch, Sotos, and Prader.

**Bias control.** To avoid the influence of urogenital factors or chronic conditions, only healthy adolescents, without a relevant clinical history, were assessed.

To determine a reliable TV assessment, all measurements were performed as follows: the same instruments were used in each measurement, the environment had to be warm ( $\geq 20$  °C) to avoid the cremasteric reflex, and the surrounding scrotal skin had to be thinned as much as possible without compressing or

	,	I		
Characteristic	Prader orchidometer	Chipkevitch visual method	Sotos method	
Instrument	A string of oval beads of known volume.	Graphic model or card with 6 oval beads of known dimensions.	Clear plastic ruler and a variant of the formula of ellipsoid equation.	
Method rationale	It is assumed that the testis is a rotational ellipsoid. The TV was obtained from the testicular width and length (measured with caliper or ruler), and the ellipsoid equation was applied: $0.52 \times W2 \times L$ where W = testicular width and L = testicular length. The scrotal skin and other adjacent tissues are included in the calculation, so both methods overestimate the TV.		It obtains a TV similar to the ultrasound value by removing adjacent tissues of the testes from its calculation. It uses the ellipsoid equation: $(W-ss)^3 \times 0.88$ where W = testicular width and ss = double scrotal skin, which varies according to the Tanner stage (G1, G2, and G3: ~1.5 mm; G4 and G5: ~2 mm).	
Technique (orchidometry)	The testis is held in one hand and the orchidometer in the other hand, and it is compared with the bead whose volume or long axis is closest.	The testis is held in one hand and the card in the other hand, and the long axis of the testis s compared with the long axis i of the most similar drawing.	<ul> <li>The Tanner stage G is determined.</li> <li>The testis is held between the index finger and the thumb and its width is measured.</li> <li>The double scrotal skin is subtracted from the testicular width.</li> <li>Data are used in the ellipsoid equation.</li> </ul>	
Scale	Beads of 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 25 mL and intermediate values between non-consecutive beads.	Oval beads of 2, 5, 10, 15, 15, 20, 25 mL and their intermediate values. TVs < 2 mL were assumed to be equal to 1 mL.	It depends on the measurement of the testicular width: since it has no maximum limit of TV, it can accurately measure large testes.	
Intrapolation	When the volume of a testis was between 2 oval beads, the intermediate value was considered. For example: between 15 and 20 mL, the value was recorded as 17.5 mL.		Not applicable.	
Extrapolation	When the TV was >25 mL, it was recorded as 25 mL, since it cannot be reliably extrapolated.		Not applicable.	
Cost*	\$1000-1300	Printing and lamination of the card \$ 100-120.	Ruler \$ 30-40.	

TABLE 1. Characteristics of three clinical orchidometric techniques

TV: testicular volume.

\*Costs expressed in Argentine pesos, estimated as of August 2020. *Source*: based on data obtained from references 1, 2, 11-14.

deforming the testis.

The statistical analysis was blinded to make up for the lack of blinding in the TV measurements by the principal investigator.

To evaluate the reproducibility of the methods among the different observers, the intra- and inter-observer agreement was estimated among 13 participants who agreed to be assessed by a second observer during the first 3 months of the study. The inter-observer comparison was made between the principal investigator (a pediatrician with 10 years of experience in adolescent care) and a general practitioner with 1 year of experience in adolescent care. Each observer blindly assessed pubertal stage and TV twice, as described under the *Procedures* section.

Data analysis and processing. Quantitative outcome measures were described with measures

of central distribution, standard deviation (SD), and interguartile range. Since the TV did not adjust to a normal distribution (Kolmogorov-Smirnov test), Pearson's coefficient and the Wilcoxon signed-rank sum test were used for related samples, for its correlation and comparison, respectively. Concordance between categorical outcome measures (onset of puberty) was compared with Cohen's kappa coefficient; and the intraclass correlation coefficient (ICC) and Bland-Altman (B&A) plots were used for quantitative outcome measures (TV in mL). The ICC is used to measure the concordance between 2 or more quantitative assessments obtained with various instruments or by different observers; values can range from 0 (no concordance) to 1 (absolute concordance). B&A plots allow comparing measurement techniques on the same

FIGURE 2. Flowchart of study adolescents



quantitative outcome measure and quantifying the mean difference between the 2 methods and the confidence limits, which include the 95 % of the differences between one measurement technique and the other. If both methods agree, the mean difference will be close to 0. If it deviates from this value, it means that both methods produce different results and one of them under or overestimates the value of the reference method. Data were processed with Epidat 4.2<sup>®</sup> (Galicia-OPS, 2016) and SPSS 20<sup>®</sup> (IBM, 2011) packages.

## **Ethical considerations**

The study was approved by the hospital's Research Ethics Committee on March 19<sup>th</sup>, 2018. Parents and patients older than 14 years of age were requested to sign an informed consent. Participation was voluntary, anonymous, and confidential. Each assessment was carried out with respect to maintain adolescents' privacy.

TABLE 2. Intra- and inter-observer concordance in testicular measurement with three orchidometric techniques (n = 13)

Orchidometric	Concordance					
techniques	<b>Intra-observer</b> : comparison of an observer against himself or herself in 2 blind evaluations		Inter-observer: comparison between 2 observers in 2 blind assessments			
	Observers	ICC (95 % CI)	Assessments	ICC (95 % CI)		
Prader	Principal investigator (PI)	0.970 (0.928-0.987)	1 <sup>st</sup> assessment by PI-SO	0.998 (0.995-0.999)		
	Second observer (SO)	0.873 (0.736-0.941)	2 <sup>nd</sup> assessment by PI-SO	0.888 (0.767-0.948)		
Chipkevitch	PI	0.996 (0.991-0.998)	1 <sup>st</sup> assessment by PI-SO	0.964 (0.917-0.984)		
	SO	0.990 (0.975-0.996)	2 <sup>nd</sup> assessment by PI-SO	0.961 (0.915-0.982)		
Sotos	PI	0.957 (0.897-0.981)	1 <sup>st</sup> assessment by PI-SO	0.666 (0.333-0.843)		
	SO	0.946 (0.885-0.9756)	2 <sup>nd</sup> assessment by PI-SO	0.768 (0.451-0.900)		

ICC: intraclass correlation coefficient, CI: confidence interval.

The intra- and inter-observer agreement or concordance with the Prader and Chipkevitch techniques was excellent (close to 1). For the Sotos method, the intra-observer agreement was excellent, while the inter-observer agreement was good in the first assessment and improved in the second one. All measurements were statistically significant (p < 0.001).

FIGURE 3. Testicular volume obtained with three orchidometric techniques (Chipkevitch, Prader and Sotos) in 367 male adolescents



MTV: mean testicular volume, RT: right testis, LT: left testis.

Descriptive statistics. Chipkevitch orchidometer: MTV:  $11.5 \pm 7.9$  mL; median: 12.5; interquartile range: 5-15 mL; minimum: 1 and maximum: 25. Prader orchidometer: MTV:  $11.7 \pm 6.9$  mL; median: 12; interquartile range: 6-15 mL; minimum: 1 and maximum: 25. Sotos method: MTV:  $3.9 \pm 3$  mL; median: 3.6; interquartile range: 2-5 mL; minimum: 0.4 and maximum: 13.75.

#### RESULTS

The most frequent cause of exclusion or withdrawal related to a urogenital condition was varicocele, with a total of 24 cases. Male adolescents with a urogenital condition were referred for medical follow-up.

A total of 367 male adolescents were assessed with 3 orchidometric techniques (*Figure 2* and *Table 2*). Their average age was  $13.8 \pm 2.5$  years.

**Testicular volume (TV).** The following was observed with the 3 methods: 1) the distribution of TV was not parametric (*Figure* 3); 2) there is a high correlation between the size of the right testis (RT) and left testis (LT) (Pearson's coefficient: 0.988 [Prader]; 0.991 [Chipkevitch], and 0.967 [Sotos]; p < 0.001); however, the comparison of median values showed that, overall, the RT was larger than the LT (Wilcoxon signed-rank sum test for related samples; p < 0.001); and 3) orchidometric techniques had a high correlation in the measurement of RT (Pearson's

coefficient: 0.993 [Prader-Chipkevitch] and 0.882 [Prader-Sotos]; p < 0.001) and of LT (Pearson's coefficient: 0.992 [Prader-Chipkevitch] and 0.889 [Prader-Sotos]; p < 0.001).

With the Prader and Chipkevitch orchidometry, 10 adolescents with TV > 25 mL were identified, all in stage G5. Similarly, the Sotos orchidometry detected 8 male adolescents in stage G5 with extreme values in terms of TV distribution.

**Concordance among orchidometric techniques.** The Prader-Chipkevitch comparison of the RT yielded an ICC of 0.993; a 95 % CI of 0.991-0.994; and a p < 0.001. The Prader-Chipkevitch comparison of the LT yielded an ICC of 0.992; a 95 % CI of 0.990-0.994; and a p < 0.001. The Prader-Sotos comparison of the RT yielded an ICC of 0.320; a 95 % CI of -0.095-0.642; and a p < 0.001; while for the LT it yielded an ICC of 0.305; a 95 % CI of -0.095-0.642; and a p < 0.001. The B&A plots showed a mean difference between Prader and Chipkevitch close

FIGURE 4. Bland-Altman plots: Concordance among orchidometric techniques in 367 male adolescents



The plot shows, as an example, the average between the right and left testis, given that their individual values are similar. TV: testicular volume, ULA: upper limit of agreement, LLA: lower limit of agreement. Bland-Altman plots: The mean of the differences and the limits of agreement on the differences, upper limit of agreement (ULA) and lower limit of agreement (LLA), are shown as horizontal lines. Y = 0 is the line of perfect average concordance. If both methods agree, then the mean difference will be 0 or close to 0. If it deviates from this value, it means that both methods produce different results. Chipkevitch-Sotos concordance is not shown in the plot; values were as follows: mean difference: 8.2 mL (95% CI: 7.6-8.7); SD: 4.5 mL; LLA: -0.6 mL; ULA: 16.9 mL. to 0 for both the RT and the LT; on the contrary, the mean difference between Prader and Sotos was close to 8 mL for both testes (*Figure* 4).

**Onset of puberty.** With the Prader orchidometer as a gold standard, the agreement among the methods in the detection of the onset of puberty was assessed. For the Prader-Chipkevitch comparison, the kappa value was 0.93; 95 % CI: 0.87-0.98; and a p < 0.001; for the Prader-Sotos comparison, values were 0.75; 0.67-0.83; and < 0.001, respectively.

**Inter-observer agreement.** Each observer performed 26 Tanner evaluations and 26 testicular measurements with 3 orchidometric techniques in 13 adolescents. The intra- and inter-observer agreement for classifying PH was excellent (kappa: 1; p < 0.001), as was the intra- and inter-observer agreement for classifying the G stage (kappa: 1 and 0.887, respectively; p < 0.001).

**Orchidometry** ICC results are shown in *Table 2*.

# DISCUSSION

Pubertal testicular growth is an indirect indicator of gonadal function in children and adolescents, as it reflects the development of the seminiferous tubules.<sup>16,17</sup> The availability of simple and reliable orchidometry instruments could improve the assessment of pubertal development and of different conditions affecting testicular growth in children and adolescents. In this study, the concordance among 3 methods of orchidometry (Prader, Chipkevitch, and Sotos) was compared, being the Prader method the gold standard.

The Chipkevitch graphic method showed a very good correlation and concordance with the Prader orchidometry in the estimation of the TV and the onset of puberty. As per the authors' knowledge, this is the first study to evaluate the Chipkevitch visual method after its first description.<sup>2</sup>

In our population, the TV range obtained by the Sotos method was 0.4-13.75 mL, similar to that described by its author (0.5-17.32 mL).<sup>1,14</sup> Coincidentally, Koskenniemi et al. describe that a TV obtained by ultrasound is approximately 0.6 mL in prepuberty and 13 mL in postpuberty.<sup>17</sup> Compared to the Prader orchidometry, the Sotos technique had low concordance in the estimation of TV, but a good one in detecting puberty onset. The Prader orchidometry has good correlation with ultrasound, but a poor concordance<sup>15-17</sup> (the Prader orchidometry overestimates the TV compared to the ultrasound). The Sotos method was developed to overcome this obstacle and, therefore, it is expected that, like ultrasonography, it will have a good correlation but no concordance with the Prader orchidometry.

**Inter-observer agreement.** In the present study, intra-observer agreement for male pubertal markers was excellent. Inter-observer agreement was excellent for Tanner staging and the Prader and Chipkevitch orchidometry, but good for the Sotos orchidometry.

Slora et al.,<sup>18</sup> evaluated the inter-observer agreement in Tanner staging and the Prader orchidometry and obtained results similar to this study. That study involved 79 children and 16 health care providers, mostly pediatricians, who were trained in 2 phases, a theoretical one and a hands-on one. In their conclusions, they emphasized that, with proper training, primary care providers can obtain reliable values of male pubertal markers. More recently, a simulation study with 3D testis and scrotum models showed a large variability and a poor accuracy in intra- and inter-observer measurements, with a clear tendency to overestimate the TV.<sup>19</sup> This experience was conducted during a meeting of the British Society of Paediatric Endocrinology and Diabetes, with 215 participants (80 % were pediatric endocrinology consultants or fellows) and only 25 % had been formally trained in testicular measurement.

## **Implications for practice**

Despite the widespread use of the Prader orchidometer in pubertal assessment, studies about concordance in clinical orchidometry are scarce and difficult to extrapolate to the healthy pediatric population, except for Slora's experience.<sup>19</sup> The few publications available are about studies conducted by urologists or andrologists,<sup>20,21</sup> or in patients with a previous urogenital or endocrine condition.<sup>21-23</sup>

In this study, the Prader orchidometry was compared with 2 other techniques of testicular measurement. In daily practice, the Chipkevitch method could be used interchangeably with the Prader method given its excellent concordance in measuring TV. On the other hand, as expected, the Sotos method showed a low concordance with the Prader orchidometer in estimating TV. To avoid errors in the interpretation of the results, it should be taken into account that, compared to Prader and Chipkevitch, the Sotos orchidometry always yields lower TV values since it does not include adjacent tissues in the calculation of TV. The Sociedad Argentina de Pediatría promotes evaluating the accuracy of physical growth measurements and improving it through training programs.<sup>23</sup> An orchidometry includes different methods and requires proper training to obtain a reliable measurement.<sup>5,12</sup> The authors consider that such training is necessary, both in basic and post-basic medical education, for the evaluation, diagnosis, treatment, and follow-up of children during puberty and also of children with different urogenital disorders.

#### Limitations

In the present study, no comparison was made with ultrasound examinations. It is known that the Prader and Chipkevitch methods overestimate TV compared to ultrasounds.<sup>1,5,14-17</sup> However, recent publications draw attention to the great variability and lack of consensus in testicular measurement by ultrasound.<sup>24,25</sup>

There is a risk of bias in the comparison of the 3 methods since it is not possible to obtain blinded measurements; the statistical analysis was blinded to mitigate this risk.

Participants came from a public hospital population and were assessed with a crosssectional design. Caution is advised when extrapolating these results. Large-scale studies on pubertal development of Argentine male adolescents, both cross-sectional and longitudinal, are necessary.

#### **CONCLUSIONS**

The concordance among the 3 clinical orchidometric techniques was compared, being the Prader orchidometer the gold standard. The Chipkevitch orchidometry is simple and inexpensive, and showed an excellent concordance with the Prader orchidometry; therefore, they could be used interchangeably in child and adolescent clinical care. The Sotos orchidometry, although less expensive, has a low concordance because it uses a measurement scale similar to the ultrasound, and the testicular size obtained is always smaller. ■

#### REFERENCES

- Sotos J, Tokar N. Appraisal of testicular volumes: volumes matching ultrasound values referenced to stages of genital development. *Int J Pediatr Endocrinol.* 2017; 2017:7.
   Chipkevitch E, Nishimura R, Tu D, Galea Rojas M.
- Chipkevitch E, Nishimura R, Tu D, Galea Rojas M. Clinical measurement of testicular volume in adolescents: comparison of the reliability of 5 methods. J Urol. 1996; 156(6):2050-3.
- Society for Adolescent Health and Medicine; Arik V Marcell, David L Bell, Alain Joffe, et al. The male genital examination: a position paper of the Society for Adolescent Health and Medicine. J Adolesc Health. 2012; 50(4):424-5.

- 4. Bell D, Breland D, Ott M. Adolescent and young adult male health: a review. *Pediatrics*. 2013; 132(3):535-46.
- 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(1):30-56.
   Dorn L, Biro F. Puberty and its measurement: a decade
- Dorn L, Biro F. Puberty and its measurement: a decade review. J Res Adolesc. 2011; 21(1):189-95.
- Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb). 2015; 25(2):141-51.
- Lejarraga H, Heinrich J, Rodríguez A. Técnicas antropométricas. In: Comité Nacional de Crecimiento y desarrollo. Guía para la Evaluación del Crecimiento Físico. Ciudad Autónoma de Buenos Aires: Sociedad Argentina de Pediatría; 2013:19-24. [Accessed on: January 2<sup>nd</sup>, 2018]. Available at: http://www. sap.org.ar/docs/publicaciones/libro\_verde\_sap\_2013.pdf
- Marshall W, Tanner J. Variations in the pattern of pubertal changes in boys. *Arch Dis Child*. 1970; 45(239):13-23.
   Comité Nacional de Endocrinología. Enfoque práctico
- Comité Nacional de Endocrinología. Enfoque práctico del manejo de la ginecomastia. Seis preguntas que debe responderse el pediatra ante un paciente con ginecomastia. *Arch Argent Pediatr.* 2011; 109(4):365-8.
- Prader Å. Delayed adolescence. Clin Endocrinol Metab. 1975; 4(1):143-55.
- Prader A. Orchiometry. In: Donald RA. Endocrine Disorders: a guide to diagnosis. New York: Marcel Dekker; 1984. Pages.33-6.
- Chipkevitch E. Avaliação clínica da maturação sexual na adolescência. J Pediatr (Rio J). 2001; 77 (Suppl 2):S135-42.
- 14. Sotos J, Tokar N. Testicular volumes revisited: a proposal for a simple clinical method that can closely match the volumes obtained by ultrasound and its clinical application. *Int J Pediatr Endocrinol.* 2012; 2012(1):17.
- Joustra S, van der Plas E, Goede J, Oostdijk W, et al. New reference charts for testicular volume in Dutch children and adolescents allow the calculation of standard deviation scores. *Acta Paediatr.* 2015; 104(6):e271-8.
- Goede J, Hack W, Sijstermans K, van der Voort-Doedens L, et al. Normative values for testicular volume measured by ultrasonography in a normal population from infancy to adolescence. *Horm Res Paediatr.* 2011; 76(1):56-64.
   Tinggaard J, Grunnet Mieritz M, Sorensen K, Mouritsen
- Tinggaard J, Grunnet Mieritz M, Sorensen K, Mouritsen A, et al. The physiology and timing of male puberty. *Curr Opin Endocrinol Diabetes Obes.* 2012; 19(3):197-203.
- Slora E, Bocian A, Herman-Giddens M, Harris D, 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(4):291-9.
- Elder C, Langley J, Stanton A, De Silva S, et al. A simulation study assessing the accuracy and reliability of orchidometer estimation of testicular volume. *Clin Endocrinol (Oxf)*. 2019; 90(4):623-9.
- Carlsen E, Andersen A, Buchreitz L, Jorgensen N, et al. Interobserver variation in the results of the clinical andrological examination including estimation of testicular size. *Int J Androl.* 2000; 23(4):248-53.
- Tatsunami S, Matsumiya K, Tsujimura A, Itoh N, et al. Inter/intra investigator variation in orchidometric measurements of testicular volume by ten investigators from five institutions. *Asian J Androl.* 2006; 8(3):373-8.
- Rivkess S, Hall D, Boepple P, Crawford J. Accuracy and reproducibility of clinical measures of testicular volume. *J Pediatr.* 1987; 110(6):914-7.
- 23. Caino S. Estimación del error de medición. In: Comité Nacional de Crecimiento y Desarrollo. Guía para la evaluación del crecimiento físico. Ciudad Autónoma de Buenos Aires: Sociedad Argentina de Pediatría; 2013:25-6. [Accessed on: January 2<sup>nd</sup>, 2018]. Available at: http://www. sap.org.ar/docs/publicaciones/libro.verde. sap. 2013 pdf
- Sancessed off. January 2 7,2010. Available at http://www.sap.org.ar/docs/publicaciones/libro\_verde\_sap\_2013.pdf
   Welliver C, Cardona-Grau D, Elebyjian L, Feustel PJ, Kogan B. Surprising interobserver and intraobserver variability in pediatric testicular ultrasound volumes. J Pediatr Urol. 2019; 15(4):386.e1-6.
- In pediatric testicular turasound volunes. J ream, et et. 2019; 15(4):386.e1-6.
  Sorokin I, Welliver C, Elebyjian L, Feustel P, McCullough A. Interinstitutional variability in testicular volumes and varicocele presence by ultrasound: surprising discrepancies and implications for clinical decision making. *Urology*. 2015; 85(5):1079-84.