Ultrasound assessment of testicular volume -An interobserver variability study

Abstract

Background

Ultrasonography measurement of the testicles and subsequent calculation of the testicular volume is recommended as a part of a standard scrotal ultrasound examination. The interobserver variability of testicular volume measurement has implications for surgical recommendations. Therefore, this study aimed to investigate the interobserver variability in the measurement of testicular volume.

Methods

Interobserver variability was established by comparing testicular measurements performed by two observers on the same patient during the same clinical appointment. The observers were blinded to each other's measurements. Testicular volume was calculated using the Lambert formula: length x width x height x 0.71. A total of three observers, A, B and C, participated in the study. The observers had between 4 to 20 years' experience with scrotal ultrasound examinations.

Results

In total, 24 patients' were included (48 testicles). The patient´s mean age was 43 years (range 19-75 years). The overall mean right testicular volume was 19.8 ml (range 7.3 – 31.6 ml), and the left was 20.1 ml (range 7.1-36.1 ml). The interclass correlation coefficient (ICC) between observer A and C was excellent (ICC= 0.98, CI:0.92-0.99), between observer A and C, was excellent (ICC=0.91, CI: 0.77-0.97) and between B and C good (ICC=0.82, CI:0.51-0.93).

Conclusion

Variability in estimating testicular volume is low, with interobserver agreement ranging from good to excellent. Ultrasound provides a highly reproducible tool to determine testicular volume.

Keywords: Ultrasound, testis, observer variability, testicular volume

Introduction

The testicle is an oval-shaped organ found extra-abdominally inside the scrotum. Testicular volume (TV) is associated positively with testosterone level, fertility, and risk of testicular cancer (1, 2). TV is used to evaluate testicular condition such as prepuberty, varicoceles, testicular function, testicular cancer, and atrophy (3-7). TV estimation may be performed using an orchidometer; however this method has been shown to overestimate the TV (8) and has a high interobserver variance (9, 10). The superficial location of the testicles is ideal for high-frequency ultrasound examination. Measuring TV during ultrasonography is more accurate than an orchidometer (9, 11, 12) and is now considered the gold standard (13, 14). Therefore, TV should be part of a standard scrotal ultrasound examination during primary investigation for testicular cancer (15, 16). A high degree of accuracy and reproducibility most be the standard when measing and hence calculation the TV.

The typical size of the testicle is up to 5 cm in length, 3 cm in width, and 2-3cm in height (17). Ultrasound measurement evaluation is performed with three maximum diameters: anterior-posterior (hight), transverse (width) diameter in transverse scan, and longitudinal diameter (length) in longitudinal scan. Calculation of TV is performed using a mathematical equation. Currently, there is no standard formula when calculating TV. Lambert´s formula is recommended by the European Society of Urogenital Radiology (ESUR) (18, 19) and uses length x width x height x 0.71 (4, 12, 20-23), while European Academy of Andrology (EAA) prefers the formula length x height x width x 0.52 for both TV and prostate volume (24). Furthermore, some ultrasound machines can today measure TV but it is not always possible to have available information on type of equation the manufacture have used per standard. Recently the Lambert formula length x width x height x 0.71 is preferred (25-27) to calculate the TV, while others prefer formulas such as π / 6 × length × height × width (5), or length x height x width x 0.52 (7, 28). One study investigated 30 testicles using three different volume formulas and found the formula of Lambert to be the optimal choice in clinical practice (21). Alternative testicular size assessment can be performed by orchiodometry; which is a tool placed directly on each testicle to measure height, length, and width on the widest part. However, this method is a rough estimate and tends to overestimate TV (26, 29).

The ESUR guideline established atrophic testicles to be less than 12 ml, and atrophic testicles may be a risk factor for testicular cancer (30). Additionally, the relation between atrophic testicles and testicular microlithiasis is debated (5). Studies on intra- and interobserver variability on ultrasound TV measurements are limited. Sorokin et al. compared inter-institutional and inter-observer variability between two settings and found TV to vary significantly (31). Welliver et al. found a 20% difference in paediatric TV measurements (32). Conversely, Oehme et al. found no significant mean interobserver differences in ultrasound-measured TV (12).

Testicular ultrasound is the gold standard modality. Urological clinical guidelines advocate scrotal ultrasound to include TV measurement (3). Scrotal ultrasound investigation for example measurement of varicocele, measurement of testicular tumor all requires accurate measurement before surgical intervention. Few studies have established the ultrasound interobserver agreement, but it is important to provide clinicians with information about the measurement variation. Therefore, this study aimed to investigate the interobserver variability of ultrasound testicular volume in patients referred from general practice to scrotal ultrasound examination.

Methods

The National Data Protection Agency and the Hospital Institutional Review Board approved this prospective study (Identification number 18/62866). Informed consent was mandatory, and patient participation was voluntary.

Consecutive patients undergoing scrotal ultrasonography at the Department of Radiology were invited to participate. Patients were referred by their general practitioner and were eligible if they were above the age of 18 years. Testicular cancer was an exclusion criterion. Indications for the scrotal examination were recorded.

A Hitachi HI VISION Ascendus (Hitachi Medical Systems GmbH, Wiesbaden, Germany) was used. The scrotal examinations was performed with one of the following linear Hitachi transducers; EUP-L74M or EUP-L75, both with broadband operating frequency ranges of 5.0 to 13.0 or 5.0 to 18.0 MHz, respectively. The procedure was performed with sufficient privacy, and the patient was supine. Adequate amounts of gel were applied to the scrotal area. When obtaining the testicular measurements, the minimum pressure was applied to avoid distorting the shape of the testicle. First, a longitudinal view was performed, measuring the length (L) of the testicle followed by the height (H). Subsequently, the transducer was turned 90 degrees to assess the testicle in the transverse plane and measure the width (W). Testicular measurements were performed on both testicles. Still images and cine loops were recorded and stored in the Picture Achieve Communication System (PACS, Easyviz Impax Workstation, Medical Insight Valby, Denmark) Department of Radiology. TV was calculated using the Lambert formula: length x width x height x 0.71 (20, 21). Volume calculations were performed in Microsoft Excel (version 16.75).

Three observers (observer A, B and C) performed scrotal investigation independently. The interobserver variability was assessed between observer A and B, observer B and C, and observer A and C. All observers had between 4 to 20 years of experience in scrotal ultrasound examinations. All the patients had a standard scrotal ultrasound investigation performed independently by two of the three observers, including measurements of the testicle's length, width, and height of both testicles. First, one of the observers performed the clinical ultrasound examination, including length, width, and height measures. Then the first observer left the ultrasound room, and another observer entered and performed the same measurements, blinded to the results of the previous observer. A template was provided to note the scrotal measurements. An example of ultrasound measurement is shown in Figure 1.

Statistical analysis

Data were analyzed using IMB SPSS statistics software version 27.0 (IMB cooperation, Armonk, New York, US). Continuous variables were described by their mean and standard deviation (SD) and categorical variables by their number and percentage. Interobserver agreement was calculated using the interclass correlation coefficient (ICC) and interpreted as poor (<0.5), fair (0.51-0.75), good (0.76-0.90) and excellent (>0.90) (33). Bland-Altman plots were conducted to assess the limits of agreement between the observers (34, 35) by plotting the difference between the TV and the mean TV of each testicle. Bland Altman plots are a simple method to accurately visualize agreement between two observers. A strength of this method may be to help in clinical decision-making or compare methods against another or a reference. Bland-Altman plots was set at 95% confidence internal.

Results

A total of 24 patients were included in the study. The mean age was 43 years ranging from 19 to 75 years. Table 1 shows the indication for ultrasound examination.

The mean volume of the right testicle was 19.8 ml ranging from 7.3 ml to 31.6 ml. The mean left TV was 20.1 ml ranging from 7.1 to 36.1 ml.

The overall mean values for observer A compared with observer B were 18.7 (SD=6.1) versus 18.7 ml (SD=5.7), respectively. The average mean value for observer B compared with observer C was 22.4 (SD=4.1) versus 22.6 ml (SD=3.9), and for observer A compared with observer C 17.7 (SD=5.3) versus 18.9 ml (SD=6.6). A difference between the ultrasonographer of more than 20% occurred in 12.5 % of the cases (6 out of 48 testicles).

Interobserver agreement between observer A and B (n=12 testicles) was excellent, with an ICC of 0.98 (95% CI: 0.92-0.99). Interobserver agreement between B and C (n=18 testicles) was good, with an ICC of 0.82 (95% CI: 0.51-0.93). Interobserver agreement between A and C (n=18 testicles) was excellent, with an ICC of 0.91 (95% CI: 0.77-0.97). The Bland-Altman plots (Figure 2, 3 and 4) showed low variability between the observers, e.g., A and B had an excellent degree of agreement (between <4 mm up to >-4 mm).

Discussion

The current study showed good to excellent interobserver agreement on the TV calculation, with ICC of 0.82, 0.91 and 0.98. Variations in the placement of the cursor can affect the length, width, and height measurements and subsequently the results. A study demonstrated that a 1 mm increase in length, width and height could result in a 2-10% difference when calculating the TV (31). In this study we found 20% difference in TV occurred in 12.5% of the patients. Our study had a small number of patients. Nevertheless, we found observers to have a high degree of agreement, which is confirmed by the ICC.

We found a TV mean of 19.8 ml (SD=5.4) in the right testicle and 20.1 ml (SD=5.7) in the left testicle. In general, testicular volumes vary according to the population examined i.e., infertile people, different countries, age etc. Huang et al. found a mean right TV of 14.7 and left 14.3 ml in males being tested for infertility (36). Hoai et al. found right TV of 13.9 and left of 13.3 ml in a cohort of Vietnamese adults (37). Pozze et al. reported TV of 15.0 ml in the right, and 14.0 ml in the left testicle (6), and Lenz reported a median of 14.5 ml in the right testicle and 13.0 in the left (28). TV varies depending on ethnicity and also age may affect fertility. A study investigated white European males found TV was lower in infertile men compared to fertile (15.0 ml vs 22.5 ml) (2). Difference between testicles is a common finding for example Vaganeé et al. investigated 345 adolescents and found that one out of five had a smaller testicle (38). Furthermore, we found a good up to excellent ICC between the observers with 0.82, 0.91 and 0.98).

An interobserver study can help determine the level of expertise and training required by medical staff. By identifying the level of agreement, this study provides a reliable basis for diagnostics and follow-up for example tumours, hydrocele, and varicocele. For patients under primary investigation for testicular cancer TV is used to evaluate the size of the affected testicle compared to the contralateral. TV of the healthy testicle can potentially decrease the need for additional biopsies in the future (15). Furthermore, patient safety is expected to increase when there are performed accurate and consistent measurements and potentially better treatment planning.

Limitations

The study was conducted over a period of two years, from January 2020 to December 2021. To include patients, two of the three observers had to be present in the ultrasound clinic during the patient's scrotal examination, which was often not possible due to other work-related tasks, covid-19, and long-term illness in department staff.

A strength of the study was that all examinations were performed in the same ultrasound machine. Furthermore, it is a strength that this study was planned and executed as a prospective study, whereas most published studies are retrospective. Another strength is the real-life setting of this study, which represents clinical practice. Another strength of this study is that patients were selected on their clinical referral and presentation which replicates the realistic setting of an ultrasound department. Future studies could evaluate interobserver variance divided on various testicular conditions and including the interobserver variation of elastography measurements.

Conclusion

Variation in measurements exists, but we found a negligible interobserver variation in TV between ultrasonographers, ranging from good to excellent. Ultrasound provides a highly reproducible tool to determine TV, however variation in measurement occurs.

Declaration of competing interest

All authors declare no conflict of interest.

References

1. Bahk JY JJ, Jin LM, Min SK. Cut-off value of testes volume in young adults and correlation among testes volume, body mass index, hormonal level, and seminal profiles. Urology. 2010;75:1318-23.
2. Boeri L CP, Ventimiglia E, Cazzaniga W, Pozzi E, Belladelli F, Pederzoli F, Alfano M …Solonia A. Testicular volume in infertile versus fertile white-european men: a case-control investigation in the real-life setting. Asian J Androl. 2021;23:501-9.
3. Srinivas R, Thomas RJ, Sebastian T, JJ. K. Testicular volume in a cohort of prepubertal Indian Children. J Indian Assoc Pediatr Surg. 2019;24:192-6.
4. Christman MS ZS, Kolon T Comparison of testicular volume differential calculations in adolescents with varicoceles. J Pediatr Urol. 2014;10:396-8.
5. Pedersen MR, Osther PJS, SR R. Ultrasound evaluation of testicular volume in patients with testicular microlithiasis. Ultrasound Int Open. 2018;4:E99-103.
6. Pozza C, Kanakis G, Carlomagno F, Lemma A, Pofi R, Tenuta M, et al. Testicular ultrasound score: A new proposal for a scoring system to predict testicular function. ANDROLOGY. 2022;8:1051-63.
7. Schwen ZR, Liu JL, Gabrielson AT, Patel HD, Gupta M, Rowe ST, et al. Testicular ultrasound underestimates the size of small testicular masses: a radiologic–pathologic correlation study. World Journal of Urology. 2021;39:3399-405.
8. Mbaeri TU, Orakwe JC, Nwofor AM, Oranusi KC, OO M. Accuracy of Prader orchidometer in measuring testicular volume. Niger J Clin Pract. 2013;16:348-51.
9. Carlsen E, Andersen AG, Buchreitz L, Jørgensen N, Magnus Ø, Matulevicuus V, et al. Inter-observer variation in the results of the clinical andrological examination including estimation of testicular size. International Journal of Andrology. 2000;23:248-53.
10. Elder CJ LJ, Stanton A, De Silva S, Akbarian-Tefaghi L, Wales JKH, Wright NP A simulation study assessing the accuracy and reliability of orchidometer estimation of testicular volume,. Clin Endocrinil 2019;90:623-9.
11. Sakamoto H, Saito K, Oohta M, Inoue K, Ogawa Y, H Y. Testicular Volume Measurement: Comparison of Ultrasonography, Orchidometry, and Water Displacement,. Urology. 2007;69:152-7.
12. Oehme NHB, Roelants M, Bruserud IS, Eide GE, Bjerknes R, Rosendahl K, et al. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys — intra- and interobserver agreement and comparison with Prader orchidometry. Pediatric Radiology. 2018;48:1771-8.
13. Behre HM, Nashan D, Nieschlag E. Objective measurement of testicular volume by ultrasonography: evaluation of the technique and

comparison with orchidometer estimates. International Journal of Andrology. 1989;12:395-403.
14. Diamond DA, Paltiel HJ, DiCanzio J, Zurakowski D, Bauer SB, Atala A, et al. Comparative assessment of pediatric testicular volume: Orchidometer versus ultrasound. . Journal of Urology. 2000;164:1111-4.
15. Danske Multidisciplinære Cancer Gruppe. Clinical guidelines: Testikelkræft- primær screening, udredning, behandling, opfølgning og senfølger https://www.dmcg.dk/Kliniske-retningslinjer/kliniske-retningslinjer-opdelt-paa-dmcg/testiscancer/testikelkraft---primar-screening-udredning-behandling-opfolgning-og-senfolger/2023 [

16. Danish Testis Cancer database (DATECA) and Clinical guidelines. Testicular cancer - primary screening, diagnostic, treatment and follow-up (Testikel kræft - primær screening, udredning, behandling, opfølgning og senfølger). Version 3.0. 2023 [16th of July 2023]. Available from: https://www.dmcg.dk/siteassets/kliniske-retningslinjer---skabeloner-og-vejledninger/kliniske-retningslinjer-opdelt-pa-dmcg/dateca/dateca_testikel-kraft_v1_1_admgodk130121.pdf.
17. S DV, H GR, M O, J RD. Sonography of the scrotum. Radiology. 2003;227:18-36.
18. Bertolotto M, Freeman S, Richenberg J, Belfield J, Dogra V, Huang DY, et al. Ultrasound evaluation of varicoceles: systematic literature review and rationale of the ESUR-SPIWG Guidelines and Recommendations. Journal of Ultrasound. 2020;23:487-507.
19. Freeman S, Bertolotto M, Richenberg J, Belfield J, Dogra V, Huang DY, et al. Ultrasound evaluation of varicoceles: guidelines and recommendations of the European Society of Urogenital Radiology Scrotal and Penile Imaging Working Group (ESUR-SPIWG) for detection, classification, and grading. European Radiology. 2020;30:11-25.
20. Lambert B. The frequency of mumps and mumps orchitis and the consequences for sexuality and fertility. Acta Genet Stat Med. 1951;2:1-166.
21. Hsieh ML, Huang ST, Huang HC, Chen Y, YC H. The reliability of ultrasonographic measurements for testicular volume assessment: comparison of three common formulas with true testicular volume. Asian J Androl. 2009;2009:261-5.
22. Cai D, Wu S, Li Y, Q C. Validity of measurements of testicular volume obtained by a built-in software of ultrasound systems: with formula

recommended by updated guidelines as reference. J Ultrason 2020;20:e181-e4.
23. Sihag P, Tandon A, Pal R, Bhatt S, Sinha A, M S. Sonography in male infertility: a useful yet underutilized diagnostic tool. Journal of Ultrasound. 2022;25:675-85.
24. Lotti F, Frizza F, Balercia G, Barbonetti A, Behre HM, Calogero AE, et al. The European Academy of Andrology (EAA) ultrasound study on healthy, fertile men: An overview on male genital tract ultrasound reference ranges. Andrology. 2022;10 Suppl 2(Suppl 2):118-32.
25. Taskinen S, Taavitsainen M, S W. Measurement of testicular volume: Comparison of 3 different methods. J Urol. 1996;155:930-3.
26. Sotos JF, Tokar NJ. Appraisal of testicular volumes: volumes matching ultrasound values referenced to stages of genital development. Int J Pediatr Endocrinol. 2017;2017:7.
27. Liu C, Liu X, Zhang X, Yang B, Huang L, Wang H, et al. Referential Values of Testicular Volume Measured by Ultrasonography in Normal Children and Adolescents: Z-Score Establishment. Front Pediatr. 2021;9:648711.
28. Lenz S, Giwercman A, Elsborg A, Cohr KH, Jelnes JE, Carlsen E, et al. Ultrasonic testicular texture and size in 444 men from the general population: correlation to semen quality. Eur Urol. 1993;24:231-8.
29. Lin CC, Huang WJ, Chen KK. Measurement of testicular volume in smaller testes: how accurate is the conventional orchidometer? J Androl. 2009;30(6):685-9.
30. Richenberg J, Belfield J, Ramchandani P, Rocher L, Freeman S, Tsili AC ea. Testicular microlithiasis imaging and follow-up: guidelines of the ESUR scrotal imaging subcommittee. Eur Radiol. 2015;25:323-30.
31. Sorokin I, Welliver C, Elebyjian L, Feustel PJ, A M. Interinstitutional variability in testicular volumes and varicocele presence by ultrasound: Surprising discrepancies and implications for clinical decision making. Infertility. 2015;85:1079-84.
32. Welliver C, Cardona-Grau D, Elebyjian L, Feustel PJ, BA K. Surprising interobserver and intra-observer variability in pediatric testicular ultrasound volumes. J Pediatr Urol. 2019;15:386.e1-.e6.

33. Koo TK, MY L. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15:155-63.
34. Pedersen MRV, Loft MK, Dam C, Rasmussen L, Timm S. Diffusion-Weighted MRI in Patients with Testicular Tumors-Intra- and Interobserver Variability. Curr Oncol. 2022;29(2):837-47.
35. Christensen HS, Borgbjerg J, Borty L, Bogsted M. On Jones et al.'s method for extending Bland-Altman plots to limits of agreement with the mean for multiple observers. BMC Med Res Methodol. 2020;20(1):304.
36. Huang Y-P, Liu W, Lui Y_D, Wang H-X, Hu K, Chen B, et al. Right testicular volume is a dominant predictor of testicular function determined by sperm parameters and total testosterone. Andrologica. 2018;50:e12955.
37. Hoai BN, Hoang L, Tran D, Cao TN, Tien LD, Sansone A, et al. Ultrasonic testicular size of 24,440 adult Vietnamese men and the correlation with age and hormonal profiles. Andrologia. 2022;54:e14333.
38. Vaganée D, Daems F, Aerts W, Dewaide R, van den Keybus T, De Baets K DWS, et al. Testicular asymmetry in healthy adolescent boys. BJU International. 2018;4:654-66.

Figure and table legends

Table 1. Scrotal ultrasound indication.

Figure 1

Ultrasound images with red arrows indicating how to perform measurements of the testicle length (A), width (B) Height (C) in a patient aged 19 years old referred to a scrotal ultrasound due to testicular pain.

Figure 2

Figure 2. Bland-Altman plot showing interobserver agreement of testicular volume on B-mode ultrasound between Observer A and B. The x-axis shows the mean of the two measurements, and the y-axis shows the difference between the two measurements. Black line equals the mean difference between observers, and the dotted lines equals the ± 95% (+1.96 SD) limits of agreement. The circles indicate right testicular measurement and square is the left testicle.

Figure 3

Figure 3. Bland-Altman plot showing interobserver agreement of testicular volume on B-mode ultrasound between Observer B and C. The x-axis shows the mean of the two measurements, and the y-axis shows the difference between the two measurements. Black line equals the mean difference between observers, and the dotted lines equals the ± 95% (+1.96 SD) limits of agreement.

Figure 4

Figure 4. Bland-Altman plot showing interobserver agreement of testicular volume on B-mode ultrasound between Observer A and C. The x-axis shows the mean of the two measurements, and the y-axis shows the difference between the two measurements. Black line equals the mean difference between observers, and the dotted lines equals the ± 95% (+1.96 SD) limits of agreement.