

·Clinical Experience·

Inter/intra investigator variation in orchidometric measurements of testicular volume by ten investigators from five institutions

Shinobu Tatsunami¹, Kiyomi Matsumiya², Akira Tsujimura², Naoki Itoh³, Takumi Sasao³, Eitetsu Koh⁴, Yuuji Maeda⁴, Jiro Eguchi⁵, Kousuke Takehara⁵, Takayasu Nishida⁶, Satetsu Miyano⁶, Chisato Tabata⁷, Teruaki Iwamoto⁶

¹Unit of Medical Statistics, Faculty of Medical Education and Culture, St. Marianna University School of Medicine, Kawasaki 216-8511, Japan

²Department of Urology, Osaka University, Osaka 565-0871, Japan

³Department of Urology, Sapporo Medical University, Sapporo 060-8543, Japan

⁴Department of Urology, Kanazawa University, Kanazawa 920-8641, Japan

⁵Department of Urology, Nagasaki University, Nagasaki 852-8501, Japan

⁶Department of Urology, ⁷Ultrasound Center, St. Marianna University School of Medicine Hospital, Kawasaki 216-8511, Japan

Abstract

Aim: To perform quality control studies on testicular volume measurements for a multi-center epidemiological study of male reproductive function. **Methods:** We constructed a data matrix with a balanced assignment for 2 consecutive days by ten investigators (andrological career: 4–21 years) from five institutions and 12 male volunteers aged 20–26 years. Testicular volume was measured by Prader's orchidometer. A skilled technician also performed an ultrasound estimate of testicular volume. **Results:** A statistically significant inter-investigator variation was found for both testes ($P < 0.05$). In addition, there was a statistically significant investigator-by-volunteer interaction in testicular volume measurement ($P < 0.01$). However, there was no statistically significant difference in the two measurements performed on consecutive days for either testis. The testicular volumes for both the right and left testes as estimated by ultrasonography were smaller than results using the orchidometer. However, there was no statistical significance ($P > 0.05$). The difference in experiences of the investigators did not significantly correlate with accuracy of measurements in either testis. **Conclusion:** The present study revealed significant differences in the results of estimation of testicular volume among the ten investigators, but intra-investigator variation was not considerable. Improved training and proper standardization of the measurement will be necessary before starting a multi-center study based on an andrological examination. (*Asian J Androl* 2006 May; 8: 373–378)

Keywords: testicular volume; orchidometer; quality control

Correspondence to: Dr Teruaki Iwamoto, Department of Urology, St. Marianna University School of Medicine Hospital, 2-16-1 Sugao, Miyamae-ku, Kawasaki 216-8511, Japan.
Tel: +81-44-977-8111, Fax: +81-44-977-0415
E-mail: t4iwa@marianna-u.ac.jp
Received 2005-06-29 Accepted 2006-01-12

1 Introduction

Testicular volume is indicative of the status of spermatogenesis [1–2]. As the seminiferous tubules comprise 70–80% of the testicular mass, testicular size strongly

correlates with sperm count [3]. Therefore, accurate measurement of testicular size in male infertility clinics is important. Generally, testicular size has been calculated by measuring length, width and/or depth diameter of the testes using calipers or has been measured using an orchidometer [4]. Takihara *et al.* [5] developed punched-out elliptical rings with the volume of the ellipsoids indicated on each ring; correlation with actual testicular size determined by water displacement of testes after castration was better than data obtained using calipers. Although ultrasound measurement is generally believed to provide a reliable estimation of testicular size [6–8], measurement by Prader's orchidometer [4] has been one of the most frequently-used conventional methods in andrologic clinics. However, few studies have compared intra/inter-investigator measurements as well as inter-institutional values. We participated in an international epidemiological study of male reproductive function that was organized by Skakkebeak at Copenhagen University [9]. As a part of this study, we determined the current status and regional differences in young men's reproductive function in five cities in Japan. Before beginning this research, we performed a quality control study of testicular volume measurements.

We constructed a data matrix with a balanced assignment for 2 consecutive days using ten investigators from five institutions and 12 volunteers. We analyzed inter-investigator as well as intra-investigator variations in the measurements of testicular volume.

2 Materials and methods

2.1 Subjects

Twelve healthy young male volunteers, university students aged from 20 to 26 years, participated in the present study. The volunteers had already participated in an international epidemiological study on male reproductive function in healthy young men. Four men had no varicocele, 2 had Grade 1 varicocele, 3 had Grade 2 varicocele and 3 had Grade 3 varicocele according to the Dubin and Amelar classification [10]. An experienced urologist (Dr Teruaki Iwamoto) determined the grade. All varicoceles were unilateral and left-sided. Ten urologists from five institutions having a mean of 11 years (4–21 years) of clinical experience in an andrology clinic served as investigators in this quality control study of testicular volume measurement. There were two urologists from each of the five institutions.

2.2 Measurements of testicular volume

Six volunteers were examined simultaneously in separate consultation rooms. Each of the ten investigators circulated between the rooms with 6 min allotted for each measurement. Then another 6 volunteers were also examined in the same way. Testicular size was estimated using Prader's orchidometer [4] with graded sizes (1-2-3-4-5-6-8-10-12-15-20-25 mL). The investigators were allowed to interpolate and extrapolate between the different sizes of the ellipsoids. The measurements were repeated using the identical procedure on the following day. Examinations were performed blindly so that each investigator did not know the results obtained by other investigators. The consultation rooms were kept at 28°C on both days without air conditioning.

2.3 Ultrasound measurement of testicular volume

Measurement of testicular volume by ultrasonography was performed by a skilled technician on the first day of the study using a combex scanner (SSD-2000; Aloka Co. Ltd., Tokyo, Japan) with a 7.5 MHz transducer. Scanning was performed in the longitudinal and transverse plane and testicular volume (V) was calculated after measuring the longitudinal axis (A) and two perpendicular dimensions (B , C) using the following formula [11–12]:

$$V = 0.71 \times A \times B \times C \quad (1).$$

2.4 Statistical methods

2.4.1 Comparisons of testicular volume in 12 volunteers measured by ten investigators

Measured volumes of left and right testes were compared using both paired and unpaired t -tests. In the paired t -test, left and right measurements were matched with each other in the same volunteers, whereas right and left datasets were treated independently in the unpaired t -test.

Intra-investigator variation between measurements performed on consecutive days and inter-investigator variation were evaluated by analysis of variance (ANOVA). To determine to what extent the variation is a result of investigators, institutions, day of measurements, testis sides, as well as volunteers, the total dataset was subjected to a $5 \times 10 \times 2 \times 2 \times 12$ (institutions \times investigators \times days \times sides [left and right] \times volunteers) mixed-model ANOVA with institutions, investigators and volunteers as random effects and days and sides as fixed

effects, and investigators nested within institutions [13]. Institutions and volunteers were cross-classified random effects, as were investigators and volunteers. The two fixed effects were also modeled as crossed. The model included interactions of all non-nested factors. Variance components for random factors and variance because of fixed factors were calculated by setting the observed means squares equal to the expected mean squares.

2.4.2 Relation between clinical experience of investigators and measurement skill

To evaluate the relation between the clinical experience of investigators and measurement skill using the orchidometer, we computed the deviation of estimation of testicular volume in the *j*-th investigator (*j* = 1, 2, ..., 10) from the results of ultrasonography. Therefore, we computed the sum of squares (SSQ_{*j*}) of the deviations defined as follows:

$$SSQ_j = \sum_k (u_i - V_{ij})^2 \quad (2),$$

where *u_i* is the testicular volume from ultrasonography for the *i*-th volunteer (*i* = 1, 2, ..., 12), and *V_{ij}* is the estimated volume for the *i*-th volunteer by the *j*-th investigator using the orchidometer. The summation suffix *k* is the number of volunteer, but it varies from 1 to 24 in Eqn 2, corresponding to a total of 24 measurements for 12 volunteers with a repetition. Then, the relationship between years of clinical experience of the *j*-th investigator and SSQ_{*j*} was evaluated by linear regression analysis for both left and right testes.

3 Results

3.1 Mean volumes of left and right testes

The mean ± SD of the measured volumes of the left and right testes measured in the 12 volunteers by the ten investigators were (19.1 ± 5.4) mL (*n* = 240; 95% confidence interval [CI]: 18.4–19.7 mL) and (20.7 ± 5.0) mL

(*n* = 240; 95% CI: 20.1–21.3 mL), respectively (Table 1). The mean volume of right testes was 1.6 mL larger than that of left testes. Statistically significant difference (*P* < 0.001) was found using both the paired and unpaired *t*-tests. The testicular volumes for both the right and left testes as estimated by ultrasonography were smaller than results using the orchidometer. However, there was no statistical significance (*P* > 0.05) (Table 1).

3.2 Intra-investigator and inter-investigator variation

We summarized the results of ANOVA in Table 2. Between two datasets corresponding to repeated measurements performed on two consecutive days, no statistically significant difference was detected (*P* = 0.708).

The variation among investigators nested within each institution was statistically significant (*P* = 0.017); however, the variation among institutions was not followed by a statistical significance.

The variance component for volunteers was readily discriminable from zero and was larger than that for investigators nested within institutions. The variance component for the interaction of investigators and volunteers was of the same magnitude as that for investigators alone, and was likewise statistically significantly different from zero (*P* = 0.008). This indicated that testicular measurement by an investigator varies by one or more characteristics of a volunteer.

Strong statistical evidence of a difference (*P* < 0.001) between measurements of volume of left and right testes is shown in Table 2, as was consistent with the results of *t*-test.

The results of testicular measurements in the 12 volunteers are summarized in Figure 1 (left testes) and Figure 2 (right testes) by box plots. Each box represents the distribution of 20 measurements obtained by the 10 investigators in the two consecutive days. In both Figures 1 and 2, various types of distribution of testicular measurements are represented. In addition, there are several values evaluated as outliers.

Table 1. Statistical parameters from measurements of the left and right testicular volumes in mL. ^a*P* < 0.001, compared with the right side; ^b*P* > 0.05, compared with the ultrasound measurement. CI, confidence interval.

Side	Estimation by orchidometer				Ultrasound measurement		
	<i>n</i>	mean (mL)	SD	95% CI	<i>n</i>	mean (mL)	SD
Both	480	19.9 ^a	5.2	19.4–20.3	24	18.7	7.4
Left	240	19.1 ^c	5.4	18.4–19.7	12	17.0	7.1
Right	240	20.7	5.0	20.1–21.3	12	20.5	7.6

3.3 Relation between variations in estimation and testicular size

The coefficients of variation; that is, the ratio of standard deviation to the mean computed in the each box in Figures 1 and 2 ranged from 5% (Volunteer No.11, right testis) to 21% (Volunteer No.1, right testis), are illustrated in Figure 3. The coefficients of variation were significantly dependent on size for both the left ($r = -0.84, P < 0.001$) and right ($r = 0.91, P < 0.0001$) testes.

3.4 Clinical experience of investigators and reliability of testicular volume measurements

The relation between the sum of squares from Eqn 2 and clinical experience of investigators in years is illus-

Table 2. Results of ANOVA. NCV means non computable value which is so tiny close to zero that it was obtained as negative value in the numerical computation.

Source of variation	Estimated variance	Probability
Single components		
Institution	NCV	0.75
Investigator nested within institution	1.30	0.017
Day of measurement	NCV	0.708
Testis side	1.28	< 0.001
Volunteer	23.00	< 0.0001
Cross components including investigator		
Investigator × Volunteer	0.98	0.008
Investigator × Day of measurement	0.08	0.23
Investigator × Testis side	0.40	< 0.001

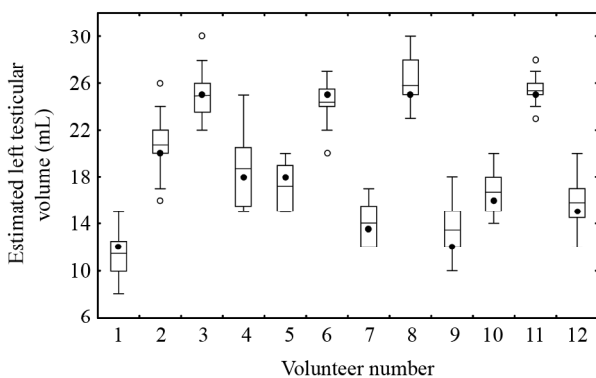


Figure 1. Box plots of estimated testicular volume of the left testis of the 12 volunteers by ten investigators. The 25th and 75th percentiles of the data define the top and bottom of the rectangles. The bullet and bar in the box indicate the median and mean, respectively. Upper and lower vertical segments express the maximum and minimum except for outliers indicated by open circles.

trated in Figure 4 for both testes. There was an identical tendency that the sum of squares slightly decreased with increases in investigator experience. However, the correlation was not statistically significant for either the left ($r = -0.28, P = 0.43$) or right ($r = -0.20, P = 0.58$) testis.

4 Discussion

In measuring testicular volume with Prader's orchidometer, which is one of the most frequently-used conventional methods, clinical experience is sometimes required to obtain satisfactory accuracy [14]. Actually, a significant variation among investigators in measurements using Prader's orchidometer was detected by Carlsen *et al.* [14]. The coefficients of variation computed for each testis ranged from 5% to 21% in our investigation

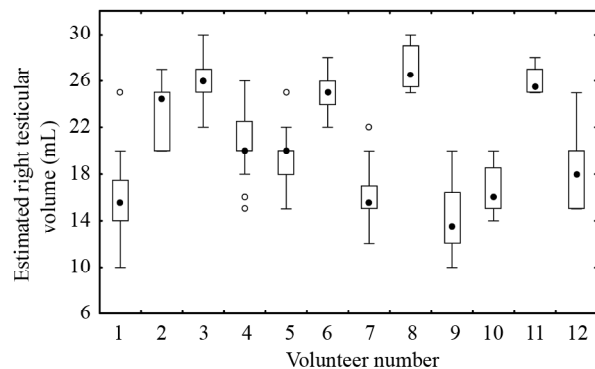


Figure 2. Box plots of estimated testicular volume of the right testis of the 12 volunteers by ten investigators.

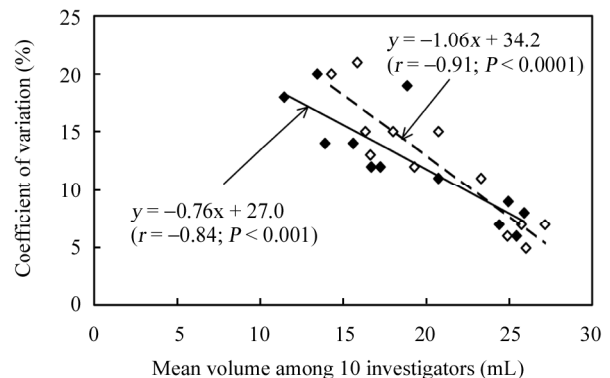


Figure 3. Plots of the coefficient of variation (%) in the estimation of testicular volume. The value of coefficient of variation among 10 investigators' estimation for each testis is expressed by closed (left testes) and open (right testes) squares with linear regression lines (solid line, left testes; dotted line, right testes).

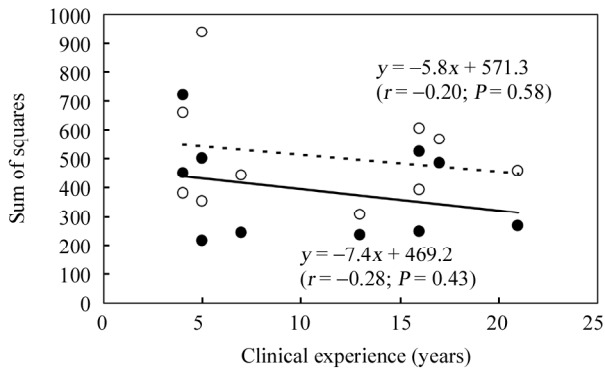


Figure 4. Plots of the sum of squares of deviation and clinical experience of the investigators in the left (closed circle) and right (open circles) testes. The sum of squares of the deviation is a measurement of deviation in estimation using the orchidometer from the results of ultrasound measurements for the each investigator. Regression lines are also indicated for left (solid line) and right (dotted line) testes.

(Figure 3). This could be considered comparable to the observation by Carlsen *et al.* [14] of an inter-observer error in testis size of 16%.

The gradations of volume in the present orchidometer are 5 mL for measurements larger than 15 mL. Therefore, it might be proposed that the larger the testicular volume, the larger the variation in estimation. However, our results showed that the coefficients of variation decreased with larger testis volume. However, this finding might imply that there is a risk that the larger variation in measurements of smaller testes might result in overlooking a diagnosis of infertility by physical examination.

The mean \pm SD of the measured volumes of the left and right testes measured in the 12 volunteers by the ten investigators were 19.1 ± 5.4 mL ($n = 240$; 95% CI: 18.4–19.7 mL) and (20.7 ± 5.0) mL ($n = 240$; 95% CI: 20.1–21.3 mL), respectively as summarized in Table 1. These are within the range of normal adult testicular size from 13.8 mL to 21.4 mL that is proposed by Takihara *et al.* [5] in Japanese men.

The mean volume of right testes was significantly larger than that of left testes, with a 1.6 mL difference (Table 1). This was similar to results obtained using a wooden orchidometer by Jorgensen *et al.* [15] from Finland, Estonia and Denmark. However, there are 8 men with unilateral left varicocele among the present 12 volunteers. Therefore, there is a possibility that the difference of the testis size was affected by the varicoceles in the left testes.

The variance component for the interaction of investigators and volunteers was of the same magnitude

as that for investigators alone, as summarized in Table 2. This suggests that the variance for an investigator varied as a function of one or more characteristics of a volunteer, the most obvious candidate characteristic being mean testicular volume of the volunteer. In fact, variation in an investigator's measurement of testicular volume varied with mean testicular volume of the volunteer, as displayed in Figure 3.

Ultrasound measurement is generally believed to give a more precise estimate of testicular size than measurements using an orchidometer; also, usually smaller volumes compared to those using an orchidometer have been reported [6, 8, 14]. In agreement with those studies, we found lower estimates using ultrasound than using the orchidometer. The difference among means from ultrasonography and orchidometer in Table 1 is not statistically significant. However, the difference would become larger and statistically significant if we used the coefficients $p/6$ instead of 0.71 in Eqn 1 [8].

Fuse *et al.* [7] concluded that ultrasonography should be the preferred method when quantitative estimates of testicular volume are required. However, the usage of an orchidometer is still clinically valuable for a rapid and rough assessment of spermatogenesis and sexual maturation.

Accuracy of measurement of testicular size using an orchidometer is dependent on experience [4, 14]. Therefore, in our study, the difference among the individual investigators might have reflected differences in past experience. Using results from ultrasound measurements, we evaluated the accuracy of the estimates by the present investigators. Ultrasound can be considered a reliable means of evaluation: Tajima [16] observed precise coincidence of testicular volume by ultrasound measurement with the volume of 34 actual testes extracted from 17 Japanese men. Values of SSQ_j from Eqn 2, that is, the sum of deviations of orchidometer estimations from the results of ultrasonography, showed no statistically significant dependence on clinical experience of the investigators (Figure 4). Therefore, the present study does not show a precise association between investigator experience and the quality of measurement.

Practically, it is difficult for many investigators from various institutions to get together and diagnose an identical group of volunteers or patients at one time. Therefore, in multi-center studies, the random effect model has sometimes been used to estimate variance components from a dataset with a partial lack of measurements [14, 17, 18]. However, our data is balanced; that is, all measurement

data were completely obtained for both investigators and volunteers. In this context, the significance of the results of the present measurements is notable.

In conclusion, the present study revealed statistically significant differences in the results of testicular volume estimation among ten investigators, but intra-investigator variation was not considerable. In this context, improved training and proper standardization of measurement techniques such as periodic comparison with ultrasonography will be necessary before starting a multicenter study based on andrological examinations and clinical experience at infertility clinics.

Acknowledgment

This study was supported by Health Science Research Grant 10130201, 1315050 from the Japanese Ministry of Health, Labour and Welfare. Shiari Nozawa, Mariko Nakanome and Miki Yoshiike assisted with technical arrangement for epidemiological study on male reproductive function in healthy young men.

References

- 1 Sherins RJ, Howards S. Male infertility. In: Walsh PC, Gittes RF, Perlmutter AD, Stamey TA, editors. *Campbell's Urology*, 5th edn. Philadelphia: W. B. Saunders; 1986. p640–97.
- 2 Behre HM, Yeung CH, Nieschlag E. Diagnosis of male infertility and hypogonadism. In: Nieschlag E, Behre H, editors. *Andrology*, 2nd edn. Berlin: Springer-Verlag; 2001. p87–114.
- 3 Setchell BP, Brooks DE. Anatomy, vasculature, innervation and fluids of the male reproductive tract. In: Knobil E, Neill J, editors. *The physiology of Reproduction*. New York: Raven Press; 1988. p753–836.
- 4 Prader A. Testicular Size: Assessment and clinical importance. *Triangle* 1966; 7: 240–3.
- 5 Takihara H, Sakatoku J, Fujii M, Nasu T, Cosentino MJ, Cockett AT. Significance of testicular size measurement in andrology. I. A new orchimeter and its clinical application. *Fertil Steril* 1983; 39: 836–40.
- 6 Behre HM, Nashan D, Nieschlag E. Objective measurement of testicular volume by ultrasonography: evaluation of the technique and comparison with orchidometer estimates. *Int J Androl* 1989; 12: 395–403.
- 7 Fuse H, Takahara M, Ishii H, Sumiya H, Shimazaki J. Measurement of testicular volume by ultrasonography. *Int J Androl* 1990; 13: 267–72.
- 8 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.
- 9 Baba K, Nishida T, Yoshiike M, Nozawa S, Hoshino T, Iwamoto T. Current status of reproductive function in Japanese fertile men: international collaborative project on a study of partners of pregnant women. *Int J Androl* 2000; 23 Suppl 2: 54–6.
- 10 Dubin L, Amelar RD. Varicocele size and results of varicocelectomy in selected subfertile men with varicocele. *Fertil Steril* 1970; 21: 606–9.
- 11 Schiff JD, Li PS, Goldstein M. Correlation of ultrasonographic and orchidometer measurements of testis volume in adults. *BJU Int* 2004; 93: 1015–7.
- 12 Shiraishi K, Takihara H, Kamiryo Y, Naito K. Usefulness and limitation of punched-out orchidometer in testicular volume measurement. *Asian J Androl* 2005; 7: 77–80.
- 13 Neter J, Wasserman W, Kutner MH. *Applied linear statistical models*, 2nd edn. Homewood: Richard D. Irwin; 1987. p1000–30.
- 14 Carlsen E, Andersen AG, Buchreitz L, Gensen NJ, Magnus O, Matulevicius V, *et al.* Inter-observer variation in the results of the clinical andrological examination including estimation of testicular size. *Int J Androl* 2000; 23: 248–53.
- 15 Jorgensen N, Carlsen E, Nermoen I, Punab M, Suominen J, Andersen AG, *et al.* East-West gradient in semen quality in the Nordic-Baltic area: A study of men from the general population in Denmark, Norway, Estonia and Finland. *Hum Reprod* 2002; 17: 2199–208.
- 16 Tajima M. Testicular measurement by test size orchidometer. *Acta Urol Jpn* 1988; 34: 2013–20.
- 17 Oda E, Ohashi Y, Tashiro K, Mizuno Y, Kowa H, Yanagisawa N. Reliability and factorial structure of a rating scale for amyotrophic lateral sclerosis. *Brain and Nerve* 1996; 48: 999–1007.
- 18 Yonenobu K, Abumi K, Nagata K, Taketomi E, Ueyama K. Interobserver and intraobserver reliability of the Japanese orthopaedic association scoring system for evaluation of cervical compression myelopathy. *Spine* 2001; 26: 1890–5.