Abstract

Aims: We examined comparative accuracy of the portable ultrasound bladder scanner, Liliumα-200, and conventional ultrasonography (CUS) in bladder volume measurement. We also examined factors that could lead to measurement errors.

Methods: Post void residual (PVR) volume was measured by Liliumα-200 and CUS with catheterized volume as comparator in 224 consecutive men, of which 109 were also measured for the serially inflated bladder with saline. The measurement accuracy with respect to the actual volume was evaluated by calculating the error volume, % error volume (EV), and their absolute values. The absolute %EV of ≤ 20% has been designated as non-error. The measurement of prostate volume, abdominal thickness, and pelvimetry was performed on MRI images.

Results: PVR volumes measured by CUS are better correlated with actual volumes (r=0.779) than those of Liliumα-200 (r=0.606). When the measurement accuracy indicated by absolute values of EV and %EV, CUS provided a more accurate estimate (21±21ml, 60±42%) than Liliumα-200 (32±45ml, 91±142%). The frequency of error was significantly increased at lower bladder volumes.

Overestimation was associated with larger prostate size for Liliumα-200, while underestimation was associated with greater bladder flattening for both methods.

Conclusions: PVR volumes measured by Liliumα-200 were fairly correlated with actual volumes. However, their relative errors were too large to correctly predict the actual volume. Flattened bladder and a large prostate may hinder accurate measurements. Consequently, Liliumα-200 is not superior to CUS and its feasibility is limited to when precise measurement is not required.

Keywords: Bladder Scanner, Residual Urine Volume, Ultrasonography.

1 INTRODUCTION

Measurement of post-void residual urine volume (PVR) is important for assessing voiding dysfunction and the therapeutic effect of certain treatments. Urethral catheterization is the most accurate procedure for measuring PVR despite its invasive nature. Alternatively, transabdominal ultrasonography has been recommended as a non-invasive method to estimate the volume of the bladder.[1,2] However, stationary general-purpose ultrasound scanners are expensive and can only be used by a trained examiner. Contrarily, in recent clinical practice, the volume of PVR is often measured by portable ultrasound (US) devices with acceptable accuracy.[3-7]

The Liliumα-200 (Lilium Otsuka, Kanagawa, Japan) is a new portable bladder scanner to periodically monitor and record the bladder volume (Fig. 1a). Kamei et al have previously reported limited feasibility of approximating the volume of the bladder using this device.[7] They mentioned that the bladder volumes measured by Liliumα-200 were strongly correlated with actual volumes. However, they showed considerable variation and may not predict actual volume accurately. We applied this device to PVR measurement to evaluate its accuracy relative to direct measurement by catheterization and compare its reliability with conventional transabdominal ultrasonography (CUS) measurement to test whether this new device can be an alternate to CUS for PVR measurement. In addition, we investigated clinical factors that may affect the accuracy of bladder volume measurement.

2 METHODS

2.1 Patient Recruitment

From April 2018 to December 2019, consecutive male patients with elevated levels of prostate-specific antigen (PSA) undergoing a prostate biopsy were included in this prospective study. Objective patients were ≥ 50 years old and had a PSA level ≥ 3.5 ng/dl with suspicious prostate cancer findings on the magnetic resonance image (MRI). In the case of many blood clots in the bladder after biopsy, they were excluded from the study. We obtained prior approval from the institutional review board (#18-004) and informed consent of all patients.

2.2 Bladder Volume Measurement

Patients were told to empty the bladder just before moving to the operating room. After the prostate biopsy was performed under general anesthesia, the volume of the bladder was initially measured by the Liliumα-200. The small US plate-shaped probe placed on the suprapubic area of patients periodically measures the volume of the bladder which appears in the form of a serial bar graph and the maximum value is indicated as the estimated bladder volume (denoted LiVmax). We have also adopted the **mean** volume (denoted LiVmea) which is calculated from a series of measurements made per test over a certain time (Fig. 1b). Bladder volume was also measured by CUS using an ellipsoid formula; CUSVe = 0.52 x length x width x height and spherical formula; CUSVs = 4pi /3 x [(length + width + height) / 3]^{3} The actual bladder volume was measured by urethral catheterization. Finally, the emptied bladder by catheterization was inflated with saline to the volume of 50, 100, 150 and 200ml and the estimated volume was measured similarly. The person performing the scans was not blinded to the volume filled with the catheter and the sequence of scans was always first the liliumα-200 scan and second, the CUS. The accuracy of the two methods was evaluated not only by calculating the error volume (EV) = actual volume - CUSV or LiV and the % error volume (%EV) = EV x 100 / actual volume, but also their absolute values (AEV and A%EV, respectively). The patients were categorized by the %EV into three groups of overestimation error (%EV> 20), underestimation error (%EV <-20) and non-error (-20 ≤ %EV ≤ 20). Bladder volumes were measured twice by the same device and mean values for each measurement were used for analysis. All measurements were made by the well-trained single examiner.

2.3 Measurement of Prostate Volume, Abdominal Wall Thickness, Bladder Flattening and Pelvimetry

Bladder shape, prostate size, body mass index (BMI), abdominal wall thickness, and pelvic shape are considered factors that affect bladder volume measurement. Bladder flattening was expressed as a ratio of a maximum section of width to depth at supine position. Prostate volume was calculated using the ellipsoid formula in which each dimension was measured on MRI. The thickness of the abdominal wall was measured at 1 inch above pubic symphysis on a sagittal MRI image. The radiological measurement of the pelvis (pelvimetry) was carried out on MRI according to the previously specified criteria, whereby the length of the pelvic inlet (promontory to pubic symphysis distance), width (interichiatic spinous distance) and depth (mid-inlet length) were given.[8] Pelvic flattening was defined as a ratio of width to depth. Pelvic volume was denoted as an estimation = 0.52 x length x width x depth.

2.4 Statistical Analysis

Data were expressed as mean ± standard deviation. As the obtained data did not exhibit normal distribution, Wilcoxon's rank-sum test was used to test differences between the two values of measured volumes by the ultrasound devices and catheterized actual volumes. We compared the differences in factors associated with bladder volume measurement within one method of estimation using the Mann-Whitney U-test. The compatibility between catheterized volume and estimated volume by the ultrasound devices was tested by Bland-Altman analysis.[9] This analysis is a statistical tool to evaluate if the two methods can be considered interchangeable when their differences are not statistically significant. The limit of agreement (LOA) was defined by the lines of mean of difference ± 1.96 SD. Spearman's rank correlation coefficient was used to assess correlations between the various paired variables. The proportion of the number of cases was compared using Fisher's exact test. A p-value of <0.05 was considered statistically significant. All statistical analyses were performed using the free R statistical software (version 3.2.2, https://cran.r-project.org/).

3 RESULTS

3.1 Comparative Accuracy of PVR Volume Measurement by the Two

Ultrasound Devices

A total of 224 men with a mean age of 66 ± 7.8 participated in this study. The mean actual PVR volume was 57.6 ± 81.3 ml. CUS could not detect bladder in 64 (28.6%) patients whose PVR volume was considered as zero, although the mean actual PVR volume was 16.2 ± 18.0 ml. There were significant correlations between the actual volume and the estimated volumes indicated by CUSVe, CUSVs, LiVmea, and LiVmax (r=0.779, 0.772, 0.606, and 0.622, respectively, p<0.0001 for all). Among these correlations, the actual bladder volume was better correlated with CUSV than LiV (p<0.01). The Bland-Altman analysis revealed fixed differences between the two methods with CUSVe, CUSVs, and LiVmea measuring lower and LiVmax measuring higher compared to the actual volume. Proportional differences were also seen between the actual volume and ultrasound methods except LiVmax (Fig 2).

To compare the accuracy of the two methods, EV and %EV were calculated (Fig 3 a, b). The measured volumes indicated by CUSVe, CUSVs, and LiVmea were underestimated, while the measured volume by LiVmax was overestimated relative to the actual volume. The CUS roughly underestimated the actual volume by a mean value of - 40%, which was significantly lower than those by

Lilium α-200, showing that Lilium α-200 seems to provide more appropriate estimation than CUS. As can be seen in Bland-Altman plots, error volumes are distributed upwards and downwards across zero line, their mean value is likely to be low despite their large variation. Thus, we adopted the absolute values of EV and %EV to evaluate the exact dissociation from the actual volume (Fig 3c, d). The mean AEV and A%EV measured by CUS were significantly smaller than those by the Lilium α-200, suggesting that CUS provides closer estimation to the actual volume than the Lilium α-200, being consistent with the above-mentioned better correlation between the actual volume and CUSV than that for the LiV. Therefore, CUS may provide a more accurate estimation with smaller variation compared with the Lilium α-200 despite its likelihood of underestimation. The ellipsoid formula and spherical formula calculated close values for CUSV, although all the values indicating errors were statistically smaller in the latter formula. Therefore, CUSVs seems to estimate the bladder volume most reliably with a smaller margin of error.

Since Bland-Altman analysis showed proportional differences between the actual volume and ultrasound methods, indicating a volume dependent error, we examined the association between the PVR volume range and measurement accuracy indicated by AEV and A%EV, emerged as representative indicators of measurement accuracy. AEV was significantly smaller in the smaller PVR volume range, while its relative value to the actual volume (A%EV) was inversely proportional to PVR volume ranges with a statistical significance between the groups (Table 1). However, unbalanced patient distribution by the PVR volume (≤ 50ml: 66%, 51~100ml: 21%, >100ml 13%) in the present study may lead to an inappropriate statistical comparison. Indeed, studies regarding the measurement error by the ultrasound devices in association with bladder volume demonstrated inconsistent results.[3, 4, 10-13]

In an effort to overcome this flaw, the measurements were performed sequentially over a wider range of 50, 100, 150, and 200 ml bladder volume filled with saline (n=109, denoted as the infused subgroup) to obtain an equal number of measurements at each volume for precise comparison between the groups in the latter half of the patients. As shown in Fig.4, this validation analysis confirmed that AEV and A%EV were significantly smaller in CUSVs across the four infused volumes. Summarily, CUS more accurately estimated bladder volume than the Lilium α-200 at least within a range of 200ml or less. Moreover, estimated volumes calculated by the spherical formula may yield closer values to the actual volumes than those by the ellipsoid formula measured by CUS.

Next, we compared the frequency of non-error and error cases (overestimate and underestimate) according to the infused volume levels by Fisher's exact test (Fig.5). We found a significantly skewed distribution in measurement error as a function of infused volume and the method of measurement ($p<0.0001$). The frequencies of underestimation by CUSVe, CSUVs were inversely proportional to infused volumes. The frequencies of non-error cases by CUSVe and CSUVs were the highest at 200ml volume, while non-error rates by LiVmea and LiVmax were less than 50% throughout the infused volume because of its large variation in relative error rate.

3.2 Assessment of the Factors Associated with Measurement Errors

Several factors including bladder flattening, prostate size, BMI, abdominal wall thickness, pelvic flattening, and pelvic volume were assumed to affect bladder volume measurement. Their involvement in measurement errors was evaluated using the %EV of CUSVe, CUSVs, LiVmea, and LiVmax as objective variables in the infused subgroup. With the Spearman rank correlation coefficient, there is a weak negative correlation between %EV of any measurement method and bladder flattening. We also found weak positive correlations between prostate volume and %EV of LiVmea and LiVmax (Table 2). We compared the assumed factors among the three groups according to the %EV range: non-error, overestimate and underestimate. The values of bladder flattening of the underestimate group were significantly larger than those of the others by all the measurement methods. The prostate volumes were larger in the overestimate group by LiVmea and LiVmax compared with the others (Table 3). Namely, overestimation was associated with larger prostate size when measured by the Lilium α-200, while underestimation was associated with greater bladder flattening for both measurement methods.

4 DISCUSSION

In this study, we evaluated the accuracy of the new portable bladder scanner Lilium α-200 and CUS against catheterized volumes of PVR and assessed patients' factors associated with measurement errors. We adopted not only maximum values that were to be taken as measurement results by the manufacturer guide but also mean values extrapolated by serial values indicated by periodical measurements since they fluctuated even under controlled deflection and patients' breathing while operating the probe of Lilium α-200.

Bladder volumes measured by the Lilium α-200 (LiVmea and LiVmax) were significantly correlated with actual bladder volumes, however, their correlations were inferior to those by CUS. When measurement errors were indicated by the mean EV and %EV, Lilium α-200 seemed to be more accurate than CUS. However, the SD range of the former was considerable, suggesting that they might cancel the mean values by positive and negative deviation from volumes measured by catheter drainage. Thus, their absolute values (AEV and A%EV) could more accurately reflect the exact deviation from the actual volume. From this point of view, despite its risk of underestimation, CUS can provide a more accurate estimate with less deviation compared to the Liliumα-200. Although the relative error showed by A%EV was larger at smaller bladder volume, the measurement accuracy of CUS was superior to that of Lilium α-200 regardless of bladder volume of at least 200ml or less, which is the range often seen in clinical practice as PVR measurement. Insufficient accuracy of the Lilium α-200 determined in the present study agrees with a similar investigation in 15 male patients during a video-urodynamics study by Kamei et al.[7] They concluded that bladder volumes measured by this device were strongly correlated with the actual volumes. However, their relative errors were too considerable (mean %EV: 5.6 ± 62.9) to predict the actual volume accurately. These results do not agree with an accuracy of ±(15% +20ml) claimed by the manufacture based on the data obtained from the range of 100-560ml of voided urine.[14] Such discrepancy may partly lies in a different volume range of study subjects. In fact, the authors commented that this new device may have a special significance to the measurement of the bladder when the amount of urine retained is small (< 100ml), in which the measurement accuracy appears to be difficult to maintain.

The volume might affect the measurement accuracy of the bladder volume because the bladder shape depends on its volume. There have been several studies on the relation between the actual bladder volume and measurement error by ultrasound devices with mixed results. Some showed better accuracy of measurement by portable ultrasound bladder scanner at lower bladder volume,[4,5,10] while the others reported opposite results.[6,7,13] On the other hand, it was reported that CUS tends to underestimate the true volume at a volume level of 100~150ml or more.[12] Schnider et al. showed that both the bladder scanner and CUS overestimated lower filling volumes and underestimate higher filling volumes.[11] In this study, Lilium α-200, and CUS were liable to underestimate more frequently at lower infused volume.

The ellipsoid formula proposed by Simpson may be the most common way to calculate bladder volumes in CUS measurements.[15] The assumption that the bladder is ellipsoid maybe only true within a limited range of bladder volume, as the shape changes with the volume. Since the bladder is flexible and surrounded by other pelvic structures that may limit its expansion to a certain direction, we surmised that volume increase would lead to an increase in at least one out of the three dimensions. Thus, we used the mean of the three-dimensional measurements as the radius for calculating spherical volumes. The differences in AEV and A%EV between the two formulas were subtle but significantly smaller by the spherical formula, indicating its superiority to the ellipsoid formula in predicting bladder volumes.

We also investigated the factors relating to measurement error. Of the assumed factors, a flattened bladder was associated with the underestimation of bladder volume measured by both CUS and Lilium α-200. This may explain the superiority of the spherical formula to the ellipsoid formula, the former in which an error in one dimension has less effect to reduce underestimation for flattened bladder. Unexpectedly, pelvic shape indicated by pelvic flattening and estimated pelvic volume did not relate to measurement error. No correlation was found between bladder flattening and pelvic flattening or pelvic volume at any infused volume (data not shown). This may imply that bladder configuration is determined by the intrinsic plasticity of the bladder wall rather than surrounding pelvic structures. Contrarily, the Lilium α-200 was likely to overestimate bladder volumes in patients with larger prostate. Oh-ka et al. also mentioned that mistaking the prostate for the bladder was significantly high in error cases measured by a bladder scan BVI6100^{TM}.[6] Since CUS can visually distinguish the prostate from the bladder, the prostate size was not involved in measurement errors.

We should note a couple of limitations of the present study. First, since the actual volume of the bladder during the measurement can be influenced by the urine output from the kidneys, each volume of 50, 100, 150 and 200ml can deviate from the exact value. However, it only takes about 10-15 minutes to complete the measurement of the four infusion volumes, which does not seem to have much effect on the results. Actually, the final drained volumes were very close to 200ml, which corroborated negligible volume by diuresis during the measurement under fasting condition before the biopsy procedure. Second, we only evaluated the measurement accuracy of male patients and need to confirm whether the data obtained is applicable to female patients. Third, we do not know how much inter-observer variability this new device produces, because all measurements are made by a single trained examiner.

The strength of the present study should also be acknowledged. First, we set the infused subgroup for balanced distribution by bladder volumes for accurate comparison between the groups to compensate volume dependent measurement error which may affect the results of the whole study. Second, MRI images were available for all patients to help provide information about pelvic anatomy that was thought to affect bladder configuration and measurement accuracy, which has never been addressed before.

5 CONCLUSIONS

Bladder volume measured by the Lilium α-200 in male patients was fairly correlated with the actual volume, although its accuracy may not be high enough to predict bladder volume due to large variation in relative error rate. Contrarily, CUS provided a more accurate estimation compared with the Lilium α-200 despite its likelihood of underestimation. Flattened bladder shape and large prostate should be involved in measurement error by such devices. The Lilium α

-200 does not seem to be an alternative to CUS for PVR measurement and may only be feasible when the precise measurement is not required.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

1. Beacock CJ, Roberts EE, Rees RW, Buck AC. Ultrasound assessment of residual urine. A quantitative method. *Br J Urol.* 1985; 57:410–413.

2. Amole AO, Kuranga SA, Oyejola BA. Sonographic assessment of postvoid residual urine volumes in patients with benign prostatic hyperplasia. *J Natl Med Assoc*. 2004; 96:234–239.

3. Cho MK, Noh EJ, Kim CH. Accuracy and precision of a new portable ultrasound scanner, the Biocon-700, in residual urine volume measurement. *Int Urogynecol J.* 2017; 28:1057–1061.

4. Park YH, Ku JH, Oh SJ. Accuracy of post-void residual urine volume measurement using a portable ultrasound bladder scanner with real-time pre-scan imaging. *Neurourol Urodyn.* 2011;30:335–338.

5. Choe JH, Lee JY, Lee KS. Accuracy and precision of a new portable ultrasound scanner, the BME-150A, in residual urine volume measurement: a comparison with the BladderScan BVI 3000. *Int Urogynecol J Pelvic Floor Dysfunct*. 2007; 18:641–644.

6. Oh-Oka H, Fujisawa M. Study of low bladder volume measurement using 3-dimensional ultrasound scanning device: improvement in measurement

accuracy through training when bladder volume is 150 ml or less. *J Urol.* 2007;177:595–599.

7. Kamei J, Watanabe D, Homma Y, Kume H, Igawa Y. Feasibility of approximate measurement of bladder volume in male patients using the Lilium α-200 portable ultrasound bladder scanner. *Low Urin Tract Symptoms.* 2019; 11:169–173.

8. de'Angelis N, Pigneur F, Martínez-Pérez A, Vitali, GC, Landi, F, Torres-Sánchez T, et al. Predictors of surgical outcomes and survival in rectal cancer patients undergoing laparoscopic total mesorectal excision after neoadjuvant chemoradiation therapy: the interest of pelvimetry and restaging magnetic resonance imaging studies. *Oncotarget*. 2018; 9:25315–25331.

9. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015; 25:141-151.

10. Alnaif B, Drutz HP. The accuracy of portable abdominal ultrasound equipment in measuring postvoid residual volume. *Int Urogynecol J Pelvic Floor Dysfunct.* 1999; 10:215–218.

11. Schnider P, Birner P, Gendo A, Ratheiser K, Auff E. Bladder volume determination: portable 3-D versus stationary 2-D ultrasound device. *Arch Phys Med Rehabil.* 2000; 81:18–21.

12. Byun SS, Kim HH, Lee E, Paick JS, Kamg W, Oh SJ. Accuracy of bladder volume determinations by ultrasonography: are they accurate over entire bladder volume range? *Urology.* 2003; 62:656–660.

13. Hwang JY, Byun SS, Oh SJ, Kim HC. Novel algorithm for improving accuracy of ultrasound measurement of residual urine volume according to bladder shape. *Urology.* 2004;64:887–891.

14. Kodama H, Huang K, Yu J, Kuchinomachi Y, Yoshimura H, Wada M. Development of a new ultrasonic urine scanner for bladder measurement. *Trans. of the Society of Instrument and Control Engineers.* 2007; 43:156-165.

15. Luján Galán M, Paez Borda A, Martín Oses E, Ruiz i Fernández J L, Gómez Trimiño M, Llorente Abarca C, et al. Análisis de la fiabilidad de la estimacion ecográfica del residuo postmiccional [Analysis of the reliability of ultrasonic estimates of the posturination residue]. *Actas Urol Esp*. 1997;21:117 - 120.

Figure legends

Figure 1.

a. The lilium α-200, a portable ultrasound bladder scanner.

b. The numbers in the red and yellow squares represent maximum and mean values measured by the lilium α-200, respectively. The details of operating procedure is referred to the website (http://www.lilium.otsuka/en/lilium200/).

Figure 2. Bland-Altman plots for examining the level of agreement between the volumes measured by catheterization and ultrasound devices.

(a) values measured by CUS and calculated using the ellipsoid formula (CUSVe), (b) values measured by CUS and calculated using the spherical formula (CUSVs), (c) mean values measured by the Lilium α-200 (LiVmea), (d) maximum values measured by the Lilium α-200 (LiVmax).

The vertical axis indicates volume difference calculated by CSUV or LiV – catheterized volume. Each plot indicates the differences between the two methods against the mean of the two methods. The dashed lines and dotted line represent the mean difference and upper/lower limit of agreement (LOA, mean difference ±1.96SD). Bias: mean difference, 95%CI: 95% confident interval, slope: calculated by the linear regression analysis.

Figure 3. Comparison of the accuracy of measurement between CUS and the Lilium α-200.

The accuracy was indicated by (a) error volume, (b) % error volume, (c) absolute error volume, and (d) absolute % error volume. Statistical differences ($p<0.001$) were seen in all the intergroup comparisons for error volume and % error volume by Wilcoxon's rank-sum test, however, significant differences were lost between LiVmea and CUSVe or CUSVs when their absolute values were compared.

Figure 4. Comparison of the accuracy of measurement between CUS and the Lilium α-200 at each infused bladder volume.

The accuracy was indicated by (a) absolute error volume and (b) absolute error volume rate. Values were compared between the two methods of measurement at each volume. Measurement errors by CUS were smaller than those by LiV. Errors were most prominent when indicated by LiVmax. Differences between CUSVe and CUSVs were subtle but statistically significant.

Figure 5. Case distribution according to the measurement error and bladder volume for CUS and Lilium α-200.

The numbers beside the bar graph represent the number of cases. There were significantly skewed case distributions in measurement errors as a function of infused volume and the method of measurement by Fisher's exact test ($p<0.0001$).