

# Revisiting Reliability of Quantified Perineal Ultrasound: Bland and Altman Analysis of a New Protocol for the Rectangular Coordinate Method

S.M. Armstrong,<sup>1</sup> J.M. Miller,<sup>1,2\*</sup> K. Benson,<sup>1</sup> S. Jain,<sup>1</sup> K. Panagopoulos,<sup>1</sup>  
J.O.L. DeLancey,<sup>2</sup> and C.M. Sampselle<sup>1,2</sup>

<sup>1</sup>University of Michigan School of Nursing, Ann Arbor, Michigan

<sup>2</sup>Department of Obstetrics and Gynecology, University of Michigan Medical School, Ann Arbor, Michigan

**Aims:** This study tested the reliability of a new protocol for the rectangular coordinate method of quantifying perineal ultrasound. **Methods:** Representative scans of healthy primiparous females were quantified by positioning a pubic bone template, drawn onto an acetate sheet containing x–y axes, over scans, by aligning the x-axis with the pubic bone central axis. Values for x ( $D_x$ ) and y ( $D_y$ ) located the urethrovesical junction (UVJ) at Rest, and at maximal Valsalva and Kegel. Range of motion (V–K) was calculated. Bland and Altman analysis, correlations, and *t*-tests determined intra- and inter-rater reliability, and variance due to designation of the pubic bone central axis (template control). **Results:** Correlations averaged 0.72, 0.70, and 0.92 for intra-rater, inter-rater, and template control experiments.  $D_x$  Rest,  $D_x$  Kegel, and V–K were reliable in all experiments. First and second measures for inter-rater  $D_y$  Rest and  $D_y$  Kegel, and template control  $D_y$  Valsalva were significantly different. Bland and Altman analysis showed  $D_y$  Rest,  $D_y$  Kegel, and  $D_x$  and  $D_y$  Valsalva for both reliability experiments to have limits of agreement (LOAs) large enough to explain  $\geq 50\%$  of the actual value ranges. Template control LOAs explained  $\leq 30\%$  of the actual value ranges. **Conclusions:** The reliability of this protocol varied according to the conditions analyzed; accurate reliability assessment of all conditions required Bland and Altman analysis; and the designation of the pubic bone central axis remained a source of variance between investigators. Our results suggest Bland and Altman analysis be used with each study that quantifies perineal ultrasound. *Neurourology. Urodynam.* 25:731–738, 2006. © 2006 Wiley-Liss, Inc.

**Key words:** bladder neck position; incontinence; perineal ultrasound; reliability

## INTRODUCTION

Perineal ultrasound is currently used to investigate risk factors and treatment modalities for stress urinary incontinence. As such, several measurement systems have been devised to quantify bladder neck position and mobility on ultrasound scans [Mouritsen and Rasmussen, 1993; Schaer et al., 1995; Chen et al., 1997; Reddy et al., 2001; Pregazzi et al., 2002; Yalcin et al., 2002]. Only one of the systems measuring bladder neck position, however, has been tested for and been reported to have good inter-rater reliability [Schaer et al., 1995].

Schaer's method quantifies bladder neck position by locating the position of the urethrovesical junction (UVJ) on a rectangular coordinate system, constructed by drawing a line through the proximal and distal tips of the pubic symphysis to use as an x-axis, and dropping a y-axis at the proximal tip of the bone. This method which was expanded to quantify mobility [Peschers et al., 1996], was endorsed twice by the German Association of Urogynecology [Schaer et al., 1996; Tunn et al., 2005], and has been frequently used [Schaer et al., 1999; Miller et al., 2001; Peschers et al., 2001; Troeger et al., 2003; Reed et al., 2004].

In the hands of our research group, however, the rectangular coordinate method, while more reliable than other methods attempted, has repeatedly proved to be problematic. Most notably, we have encountered several inherent sources of variance when attempting to draw the central axis of the pubic bone on a scan. First, the pubic bone is often asymmetrical, providing no objective and obvious central axis. Second, at times only the proximal half of the bone appears on a scan, requiring conjecture in the creation of the central axis. Third,

The authors could not be contacted about potential conflicts of interest. Abbreviations used:  $D_x$ , value of x;  $D_y$ , value of y; LOAs, limits of agreement; SUI, stress urinary incontinence; UVJ, urethrovesical junction; V–K, range of motion.

Grant sponsor: National Institutes of Health-National Center for Nursing Research; Grant number: RO1 NR04007.

\*Correspondence to: J.M. Miller, PhD, Research Assistant Professor, 400 N. Ingalls, Room 3247, University of Michigan School of Nursing/Division II, Ann Arbor, MI 48109-0482. E-mail: janismm@umich.edu

Received 6 January 2006; Accepted 12 May 2006  
Published online 8 August 2006 in Wiley InterScience

(www.interscience.wiley.com)

DOI 10.1002/nau.20299

the appearance of the shape of the bone may change across frames of a scan according to how the probe is positioned, how the bladder is being maneuvered, and how the adjacent connective tissue shifts, sometimes obscuring the actual bony edge. These changes in the appearance of the bone contribute to variance in the assignment of the x-axis.

The purpose of this study was to devise a protocol to measure bladder neck position and mobility that would minimize the above-mentioned sources of variance, and achieve measures with acceptable levels of intra- and inter-rater reliability. This protocol involves tracing the edges and all landmarks of the pubic bone onto a sheet of acetate film containing an x-y coordinate system. Before tracing, this acetate is positioned along the central axis of the bone according to specific guidelines, and, once drawn, it is used as a template for all measures in the same subject. Here, we present the protocol, and, using Bland and Altman analysis [Bland and Altman, 1986], report on the repeatability and reliability of this protocol as well as on variance due to the designation of the central axis of the pubic bone.

## MATERIALS AND METHODS

### Population

Ultrasound data used to develop and test this method for quantification of bladder neck position and mobility was collected between August 2000 and July 2003 as part of an ongoing research project addressing pelvic floor changes across the birth year in primiparous women. Subjects included healthy women over 18 years old, who were less than 20 weeks pregnant, and had no history of urinary incontinence or any other pelvic floor anomaly. All experimental procedures were approved by the Institutional Review Board at the University of Michigan.

### Perineal Ultrasound

Perineal ultrasound was performed on women at 20 weeks gestation, and at 6 weeks, 6 months, and 12 months postpartum. The examination was conducted using the Siemens Sonoline SI-400 with a 5.0 MHz curved array probe (Erlangen, Germany) and a 12-French single-sensor Gaeltec (Medical Measurements, Inc., Hackensack, NJ) intravaginal catheter. A Mini DV Handycam Camcorder (Sony Corp., Tokyo, Japan) was attached to the Sonoline SI-400 to record ultrasound sessions for later analysis. Ultrasound gel (MediChoice (Owens & Minor, Inc.), Richmond, VA) was inserted into a latex-free probe cover (Sheathing Technologies, Inc., Morgan Hill, CA), the probe was inserted into the cover, and more gel was applied to the outside of the upright probe. Ultrasound was performed with the probe positioned against the vulva in the mid-sagittal plane of subjects in the standing position, and with at least 100 cc of fluid in the bladder. (Minimum rather than constant bladder volume was used in order to avoid the potential risk and additional discomfort of urethral

catheterization during pregnancy). Once an image of acceptable quality was obtained on the screen, the ultrasound and intravaginal pressure were recorded with the subject at rest, and while performing kegel, cough, kegel-cough, and valsalva maneuvers. After the ultrasound session, scans were quantified on Dell personal computers (PC) using Adobe Premier Pro video-editing software. Rest, kegel, and valsalva maneuvers were evaluated for the purposes of this study, according to the protocol below.

### Protocol for Image Analysis

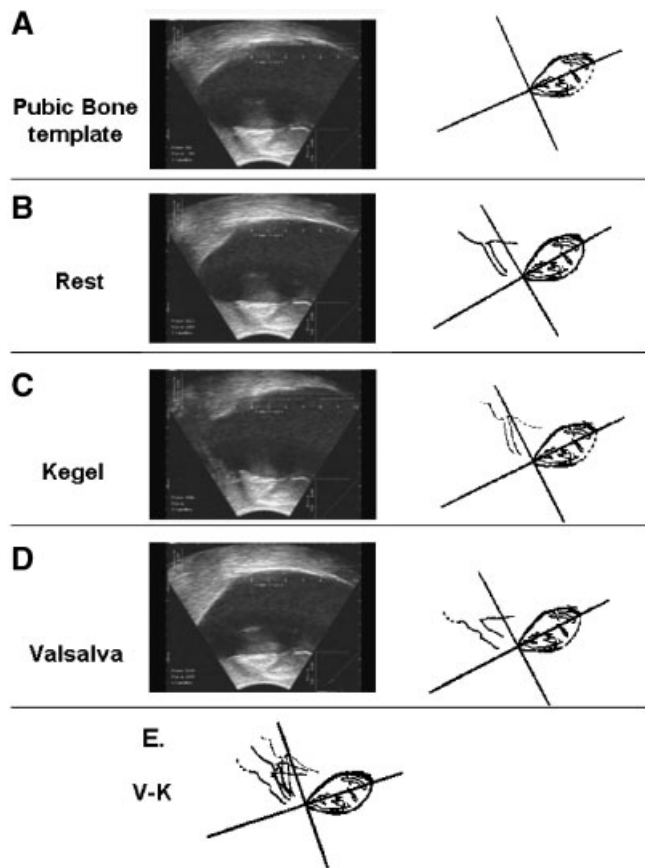
This protocol for quantifying UVJ position and mobility was conducted according to Schaer et al. [1995], locating the UVJ in the pelvic floor according to a rectangular coordinate system. Several modifications of the method, however, were specified in this protocol to improve the reliability and repeatability of these measures in our hands.

**Drawing of the template.** To have a consistent reference from which to measure changes in the pelvic floor, a template of the pubic bone in relation to an x-y coordinate system was drawn for each subject. All film clips from a subject were scanned to locate a frame that represented the pubic bone clearly and comprehensively. A transparent acetate (4" × 5.5") containing standard perpendicular x-y coordinate axes that had been Xeroxed onto it, was positioned over the ultrasound scan so that the origin of the axes was at the proximal tip of the pubic symphysis, and the positive x-axis dissected the central length of the pubic bone (Fig. 1A).

To maximize the consistency and objectivity of the positioning of the template axes, the following guidelines were observed:

- (a) The positive x-axis was positioned to dissect the length of the pubic bone from the proximal tip through the midpoint of the widest visible segment of the bone. This stipulation standardized the positioning of axes on pubic bones that were asymmetrical, and in those in which only part of the bone was visible on the ultrasound scan.
- (b) Before tracing the pubic symphysis on the chosen film frame, the bone was viewed during the kegel and valsalva maneuvers in order to detect movement of connective tissue at the borders and tip of the bone, and thereby distinguish the true bony edges of the symphysis. This proved to be especially important when distinguishing the bony tip from cartilaginous tissue that frequently followed the contour of the tip, and in recognizing clitoral and connective tissue lying adjacent to the inferior border of the symphysis. This connective tissue was subsequently excluded from the outline of the bone.

Once the axes were correctly positioned over the bone, the acetate was attached directly to the computer screen with



**Fig. 1.** Quantification of UVJ position on ultrasound scans. Clear acetate sheets (right) are placed over scans on computer screen (left), and urethrovesical junction (UVJ) position is traced relative to the pubic bone. Illustrated here: **(A)** Template of pubic bone with landmarks; **(B)** Template + UVJ at rest (on 2nd acetate); **(C)** Template + UVJ at height of kegel; **(D)** Template + UVJ at depth of valsalva; **(E)** Overlay of kegel and valsalva acetates to measure V-K.

removable “Post-its” (3M, St. Paul, MN) and, using a vis-à-vis Extra Fine Point Erasable Marker, the investigator created a template of the pubic bone, tracing the edges and all noticeable landmarks of the bone (a minimum of two bright white lines and several areas of variable opaqueness) onto the acetate. This template was then used on other film clips of this subject to provide an x-y coordinate system that could be consistently oriented with the pubic bone.

**Tracing the UVJ at rest, and at maximal displacement during kegel and valsalva maneuvers.** To test our method on the full range of motion of the UVJ, UVJ position was measured at rest, and at maximal displacement during kegel and valsalva maneuvers. (The UVJ was defined as the point where the ventral wall of the urethra met the ventral bladder floor and formed an apex. In the case that distension of the neck of the bladder occurred on a valsalva maneuver (sometimes referred to as “funneling”), the UVJ was specified as the last angle the urethra made with the bladder floor.) Films were reviewed to

find a high quality picture of the pelvic floor at rest, and to determine the frames in which maximal displacement of the bladder neck occurred during kegel and valsalva maneuvers. On each of these frames, the template of the pubic bone was positioned with a post-it on the computer screen over the scan, using the landmarks as well as the traced edges of the bone to align the template with the current scan. Once the template was attached, a second acetate, identical to the one used to make the template, was positioned over the template by alignment of the x-y axes, and the bladder neck and urethra were traced onto the 2nd acetate.

**Calculating  $D_x$ ,  $D_y$ , and Range of Motion (Valsalva to Kegel (V-K)).** Acetate sheets were individually placed over axes on graph paper in millimeter denomination, and the location of the UVJ at rest, at maximal kegel, and at maximal valsalva, was specified by the distance of x ( $D_x$ ), and the distance of y ( $D_y$ ) (in mm). The range of motion (V-K) was defined as the distance between the UVJ at maximal kegel and at maximal valsalva, and was calculated by overlaying the acetates for these two maneuvers, and measuring the distance between these two points (see Fig. 1E).

#### Data Selection

All images were previewed, and were excluded from analysis if any of the following structures were not visible: a significant portion of the pubic bone (at least the posterior half and the tip); the bladder; the urethra; and/or the UVJ. Technical flaws (such as shifts in the ultrasound gel and contrast resolution problems) resulted in the exclusion of approximately 28% of collected data. The remaining scans were studied and graded according to the quality of the image. Subject data was chosen for the experiments in this study to represent the quality of our viable dataset, and as such, we included low and middle, as well as high quality images.

Also of note, while these data of pregnant females were used because they comprised a large, complete data set, the experiments in this study were designed to test only intra- and inter-rater reliability of the protocol on a constant data set. Consequently, the following experiments are neither indicative of nor affected by potential though minimal changes to the pubic bone that might occur as a result of pregnancy [Kelly, 1979; Ortega et al., 2003].

**Experiment I (R1/R1).** To test intra-rater reliability, Rater 1 (R1:SA) rated, and 1 month later re-rated, images from 12 ultrasound sessions; 3 from each of the 4 time periods across the birth year.

**Experiment II (R1/R2).** To test inter-rater reliability, R1 and Rater 2 (R2:JM) rated images from 24 ultrasound sessions; 6 from each of the 4 time periods across the birth year.

**Experiment III (R1/R3).** To test whether drawing of the template—thus, identifying a central axis of the bone, and positioning the x-axis of the acetate along that axis before drawing the template—contributed variance to the reliability of this protocol, R1 rated images from 24 ultrasound sessions, and Rater 3 (R3:KB) then re-rated these images, but used the same template drawn by R1. All scans were from the 6-week post-partum ultrasound sessions.

**Statistical Analysis**

To determine repeatability and reliability of this protocol for quantifying bladder neck position and range of motion (V–K), Student’s paired *t*-tests, and Bland and Altman Reliability Analyses [Bland and Altman, 1986] were run, and Pearson correlations calculated between 1st and 2nd measurements of each condition within each experiment. *P* values < 0.05 were considered statistically significant. Scatter plots of actual values were drawn to assist in interpretation. For Bland and Altman analyses, differences between 1st and 2nd measures were plotted against the mean values, and plots were examined for bias, data distribution, and determination of 95% limits of agreement (LOA). The LOA are defined as the mean difference ( $\bar{d}$ ) +/– 2SD.

**RESULTS**

**Descriptive Statistics**

Table I includes the means and ranges of the actual values determined within each experimental condition. Column 1 shows average values ( $\bar{x}$ ) of repeated measures by R1 (subject *n* = 12); column 2, average values of measures taken by R1 and R2 (subject *n* = 24); and, column 3, average values of measures taken by R1 and R3, who both used the template drawn by R1 (subject *n* = 24). Values across column 1 and 2 are highly similar, as expected, since data in column 1 is a subset of that

in column 2, and both groups are comprised of an equal number of subjects from each of four time points across the birth year. Of note, all  $D_x$  measures and V–K are nearly identical between columns 1 and 2 (differences ≤ 2 mm).  $D_y$  values between column 1 and 2 differ, on average, by 4 mm. Means (and ranges accordingly) of column 3 differ from those of the first 2 columns by 0–7 mm, also expected, since this dataset was drawn solely from the 6-week postpartum time point.

**Correlations Between 1st and 2nd UVJ Measures Within Each Experimental Condition**

Pearson correlations and scatterplots demonstrate the relationship between measures in each of the three experimental conditions (Fig. 2), and present patterns within the data that offer insight into our method. First, the correlations for intra-rater measures (column 1: R1/R1) range from 0.59 ( $D_y$  Valsalva) to 0.90 ( $D_x$  Kegel) ( $\bar{r}$  = 0.72), and those for inter-rater (column 2: R1/R2) from 0.52 ( $D_y$  Valsalva) to 0.87 ( $D_x$  Kegel) ( $\bar{r}$  = 0.70), with a high similarity in *r* values between column 1 and 2 for each condition. Second, while data in most (10 out of 14) of the panels in columns 1 and 2 have regression lines that coincide with the line of perfect agreement (the dotted diagonal line), intra-rater  $D_x$  for Kegel and Valsalva (panels g and m), and inter-rater  $D_y$  for Rest and Kegel (panels e and k) consist of data that are weighted on one side of the diagonal line, suggesting potential bias in these measures. Third, the data in column 3 (R1/R3), produced when R1 and R3 used the same template, have *r* values that range from 0.89 ( $D_x$  Valsalva) to 0.96 ( $D_x$  Kegel) ( $\bar{r}$  = 0.92), and, as such, are more highly correlated overall, as well as for each condition measured, than are the measures in column 1 or 2. Furthermore, data for all conditions in column 3, with the possible exception of  $D_y$  Valsalva, regress toward the line of perfect agreement (*x* = *y*). Finally, in these scatterplots, of the four indications of UVJ position or mobility (R, K, V, or V–K), V–K appears to be the only condition measured that is consistently distributed around the line of perfect agreement across all three experimental conditions. Of note, this condition was the only one whose repeated measures would be unaffected by variance in how the pubic bone for a subject was aligned on the coordinate system.

**TABLE I. Means and Ranges of UVJ Position in Primiparous Women by Experiment**

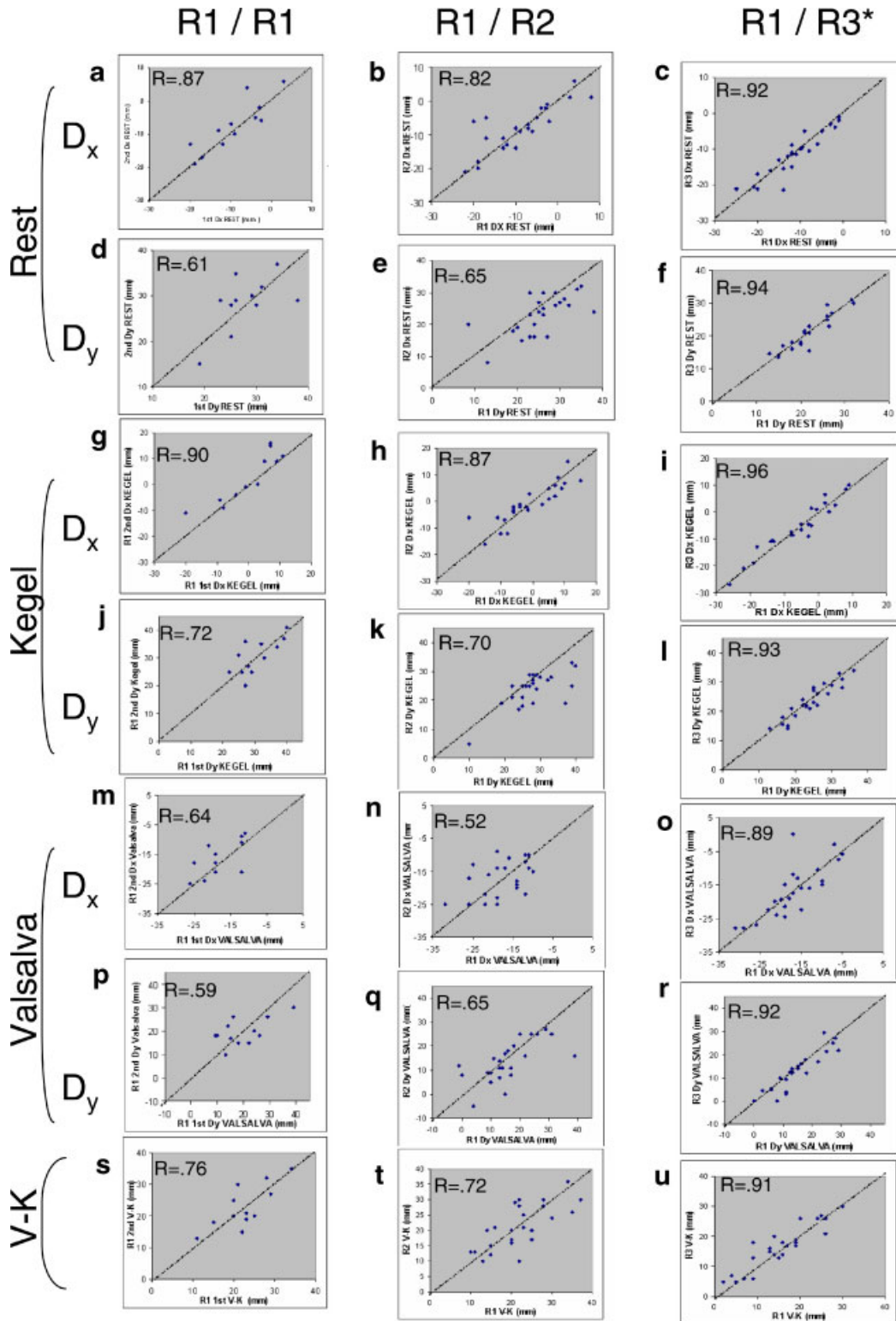
		R1/R1 $\bar{x}$ (range)	R1/R2 $\bar{x}$ (range)	R1/R3* $\bar{x}$ (range)
Rest	$D_x$	–8 (–19 → 4)	–8 (–21 → 5)	–11 (–23 → –1)
	$D_y$	28 (17 → 35)	24 (10 → 33)	21 (14 → 31)
Kegel	$D_x$	1 (–15 → 11)	–1 (–15 → 13)	–6 (–26 → 9)
	$D_y$	30 (23 → 40)	26 (7 → 36)	24 (13 → 35)
Valsalva	$D_x$	–17 (–25 → –9)	–17 (–28 → –10)	–17 (–29 → –5)
	$D_y$	20 (11 → 34)	15 (–1 → 28)	14 (0 → 27)
V–K		23 (12 → 34)	22 (11 → 35)	16 (3 → 30)

R1/R1 = repeated measures by R1 on scans from four time points across the birth year; R1/R2 = measures between R1 and R2 on scans from four time points across the birth year; R1/R3\* = measures between R1 and R3 on scans from the 6 week post-partum time point.

\*Raters used the same pubic bone template.

**Reliability and Repeatability of UVJ Measures: Bland and Altman Analysis, and Student’s *t*-Tests.**

Bland and Altman analysis [Bland and Altman, 1986] tested the reliability and repeatability of this protocol for measuring UVJ position and mobility, and Student’s paired *t*-tests screened for statistically significant differences between the 1st and 2nd measures of conditions within each experiment. Table II summarizes the results from Bland and Altman scatterplots of differences against the mean (actual scatterplots are available upon request), and indicates with asterisks experimental conditions in which the repeated measures were



**Fig. 2.** Correlations between urethrovesical junction (UVJ) position assessed by multiple raters. Scatter plots of actual data points illustrate the relationships for: repeated measures taken by R1 (1st column (R1/R1)); measurements taken by R1 and R2 (2nd column (R1/R2)); and measurements taken by R1 and R3 (3rd Column (R1/R3\*)). Labels on left border specify condition assessed in each row. Pearson correlation, *r*, is indicated in the top left corner of each panel. Diagonal dashed lines represent perfect agreement ( $r = 1; x = y$ ), and are included for reference only. \*Same pubic bone template was used by R3 and R1.

**TABLE II. Reliability Indices of Condition by Experiment**

Measure	Intra-rater			Inter-rater			Template control		
	$\bar{d}$	SD <sub>d</sub>	95% LOA	$\bar{d}$	SD <sub>d</sub>	95% LOA	$\bar{d}$	SD <sub>d</sub>	95% LOA
Rest D <sub>x</sub>	1.8	3.8	-6 → 9	-0.9	4.6	-10 → 8	0.1	2.5	-5 → 5
D <sub>y</sub>	0.7	4.8	-9 → 10	2.4*	5.4	-8 → 13	0.7	1.9	-3 → 4
Kegel D <sub>x</sub>	2.4	4.2	-6 → 11	-0.3	4.4	-9 → 8	-0.3	2.6	-5 → 5
D <sub>y</sub>	0.1	4.5	-9 → 9	4.0**	4.9	-6 → 14	0.8	2.2	-4 → 5
Vals. D <sub>x</sub>	-1.7	4.7	-11 → 7	-0.7	5.6	-12 → 10	0.6	3.2	-6 → 7
D <sub>y</sub>	0.0	7.1	-14 → 14	2.3	7.4	-12 → 17	1.5*	3.2	-5 → 8
V-K	0.3	4.6	-9 → 9	0.7	5.4	-10 → 11	-1.1	3.2	-7 → 5

$\bar{d}$ , mean difference between measurements; SD<sub>d</sub>, standard deviation of the differences; 95% LOA, 95% limits of agreement calculated as  $\bar{d} \pm 2$  SD<sub>d</sub> and rounded to nearest millimeter.

Asterisks indicate significant differences between 1st and 2nd actual measures for each condition.

\* $P < 0.05$ .

\*\* $P < 0.001$ .

significantly different ( $P < 0.05$ ). Figure 3 shows two representative Bland and Altman scatter plots.

Seven conditions indicating UVJ position and mobility are assessed in these experiments, and the reliability of this method was analyzed for each condition. In the inter-rater experiment (R1/R2), which is most relevant in predicting reliability of this method, Bland and Altman analysis shows 1st that the mean difference ( $\bar{d}$ ) between raters is  $<2$  mm for

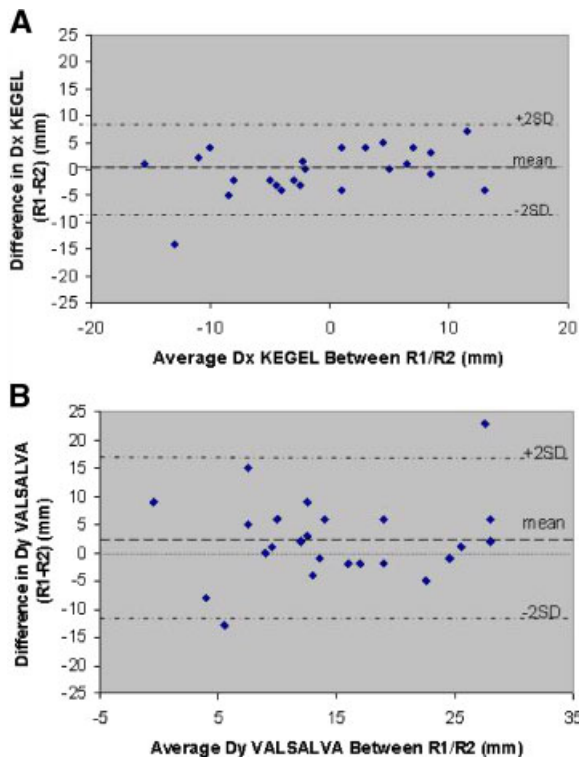
V-K and D<sub>x</sub> measures of Rest, Kegel, and Valsalva, indicating negligible bias between raters in these measures. For D<sub>y</sub> measures of Rest, Kegel, and Valsalva, however,  $\bar{d}$  was 2.4, 4.0, and 2.3, respectively, indicating a small bias in measures for each of these conditions, with R1 rating higher values each time. (Table II). (See Fig. 3 for a graphical representation of bias (B), or lack of bias (A)).

To determine agreement, for all conditions, 95% LOAs were calculated using  $\bar{d}$  and SD (LOA =  $\bar{d} \pm 2$  SD) (Table II). The LOAs predict that 95% of the differences for each measure lie between these limits. On average, the 95% LOAs were smaller for D<sub>x</sub> measures and V-K, and larger for D<sub>y</sub> measures, with D<sub>x</sub> Rest having the smallest LOA (-10 → 8) and D<sub>y</sub> Valsalva having the largest LOA (-12 → 17). Additionally, Student's *t*-test showed that measures between R1 and R2 for D<sub>y</sub> Rest and D<sub>y</sub> Kegel were significantly different at  $P < 0.05$  and  $P < 0.001$ , respectively.

For the intra-rater experiment, bias was negligible ( $\bar{d} < 2$  mm) for all measures except D<sub>x</sub> Kegel ( $\bar{d} = 2.4$ ), in which the 2nd measure tended to be higher. In all conditions, the intra-rater LOA were smaller by an average of 2 mm than those for the inter-rater experiment, and no conditions had significantly different 1st and 2nd measures. Likewise, in the template control group, for all conditions  $\bar{d}$  ( $<2$  mm) indicated negligible bias, and LOAs were smaller than inter-rater LOAs by an average of 11 mm. Student's *t*-test for the template control experiment showed D<sub>y</sub> Valsalva measures to be significantly different at  $P < 0.05$ .

## DISCUSSION

This study was conducted to test the reliability of a newly detailed protocol for the use of the rectangular coordinate system to measure bladder neck position and mobility on perineal ultrasound scans. The results of this study, conducted on data from primiparous women over the course of the birth year, offer three main findings. First, the data indicate that the reliability of this protocol varied according to the condition



**Fig. 3.** Representative Bland and Altman scatterplots. For inter-rater D<sub>x</sub> of UVJ position at maximal KEGEL (A), and D<sub>y</sub> of UVJ position at maximal VALSALVA (B), averages of R1 and R2 scores are plotted against differences in scores. Dashed line indicates mean difference, and dashed-dotted lines are 95% limits of agreement (LOA's). Dotted line indicates 0 difference between scores.



being tested. Second, all three types of statistical analysis—correlational analysis, significance tests, and Bland and Altman reliability analysis—were useful, and the latter two were required to accurately evaluate reliability of all conditions tested using this protocol. Third, the results of the template control experiment suggest that, in spite of the adjustments implemented by the protocol to eliminate variance from this measure, the identification of the central axis of the pubic bone remained a significant source of variance in determining bladder neck position, and appeared to explain much of the lack of agreement between measures in the intra- and inter-rater experiments. These three findings and their implications are discussed here in the context of the data.

Inter-rater (R1/R2) experimental results indicated that this protocol was reliable for measuring  $D_x$  Rest,  $D_x$  Kegel, and V–K, and unreliable for measuring  $D_y$  Rest and Kegel, and  $D_x$  and  $D_y$  Valsalva. For  $D_x$  Rest,  $D_x$  Kegel, and V–K, correlations between values scored by R1 and R2 were good, there were no significant differences between scores, no bias ( $\bar{d} < 1$  mm), and the LOAs were reasonable relative to the ranges of the actual values. Reliability statistics for  $D_y$  Rest,  $D_y$  Kegel, and  $D_x$  and  $D_y$  Valsalva, however, were problematic. The LOAs determined by Bland and Altman analysis for these four conditions, though only approximately 1.5 cm in either direction, were large enough to account for between 48 and 66% of the range of actual values (see Tables I and II), indicating a lack of reliability for these measures. Bias between investigators (correlation panels (Fig. 2e,k,q), and  $\bar{d}$  values in Table II), exacerbated the lack of reliability in  $D_y$  measures of Rest, Kegel, and Valsalva, and resulted in significantly different values between R1 and R2 on the  $D_y$  measure of Rest and Kegel. Careful review of the analyses revealed that the lack of reliability for measures of position on kegel and valsalva were due not only to lack of agreement in the division of the pubic bone on the template, but also to differences on the identification of the UVJ, because of urethral distortion on these maneuvers. This distortion has been previously noted [Pregazzi et al., 2002]. Taken together, these data indicate that in this population of healthy, primiparous females, this protocol for the rectangular coordinate method could be used to reliably measure some conditions, but not others.

Data from all three experiments reflect the benefit of correlational analysis, and the need for both significance testing, and particularly, Bland and Altman analysis in order to assess reliability of this protocol. Correlational analysis (Fig. 2) provided excellent, immediate insight into this protocol. It revealed, for example, the tendency toward bias in certain measures, such as Inter-rater  $D_y$  Kegel, the dramatically improved agreement between measures when investigators used the same pubic bone template (R1/R3) and, the tendency of measures for V–K across all three experiments to be distributed around the line of perfect agreement. Bland and Altman analysis, on the other hand, was essential to revealing the lack of reliability in conditions, for example,  $D_x$  and  $D_y$  Valsalva, whose measures passed significance testing, yet had LOAs that explained over 50% of the

actual range. The necessity of Bland and Altman was even more pronounced when analyzing the intra-rater results. Here, all conditions passed significance testing, however, the Bland and Altman analysis revealed that, like in the inter-rater results,  $D_x$  Rest,  $D_x$  Kegel, and V–K could be reliably measured, while  $D_y$  Rest,  $D_y$  Kegel, and  $D_x$  and  $D_y$  Valsalva had LOAs that were large enough to account for over 50% of the range of actual values; thereby rendering these four conditions unreliable. Finally, with the template control experiment, bias was minimal, and the LOAs were small enough for all conditions tested to be reliable. In this case, however, significant differences between measures by R1 and R3, again from misidentification of the UVJ due to urethral distortion, caused  $D_y$  Valsalva measures to be unreliable. Without using both Bland and Altman analysis, and significance testing on these three experiments, the lack of agreement between measures on several of these conditions would have gone undetected.

Our results, particularly those of the template control experiment, challenge the recommendations of the German Association of Urogynecology who have asserted that the central line of the symphysis provides a “reliable reference line” [Tunn et al., 2005] from which to measure bladder neck position. Prior to conducting this study we found that when using the rectangular coordinate method as described by Schaer et al. [1995], we were repeatedly unable to obtain reliable measures of bladder neck position, due primarily to an inability to identify a repeatable and reliable central axis on pubic bones, particularly those that were asymmetrical, only partially visible on a scan, or obscured by adjacent connective tissue. Despite our efforts in this protocol to address these anomalies, when the designation of the central axis of the bone was eliminated altogether as a source of variance, as it was in the template control experiment, agreement of measures between raters improved dramatically (Fig. 2, Table II), suggesting that the designation of the central axis is a considerable source of variance in this method. Conclusions about variance due to the identification of the central axis from our study are limited as different data sets were used in the inter-rater and template control experiments. Additionally, in order to avoid bias from one rater scoring the same scans twice, the second rater was different in the inter-rater (R1/R2) and template control (R1/R3) experiments, although all three raters trained extensively together on the same protocol, before collecting data. These limitations notwithstanding our data suggest that the central line of the pubic symphysis is not always a reliable reference line.

Our data indicate that this protocol for the rectangular coordinate method, which was much more detailed than that published by Schaer and colleagues in 1995, can assess bladder neck position reliably for some conditions, yet not for others. Schaer et al. [1995] asserted that the rectangular coordinate method for measuring bladder neck position was a reliable protocol. Our study can be reconciled with that of Schaer et al. in the following ways. The condition that we report as the most unreliable to quantify was  $D_y$  Kegel (with differences between raters at  $P < 0.001$ ). Schaer et al. did not test the reliability of this method

during a kegel, only during rest and valsalva maneuvers. Additionally, using the more stringent statistical analysis of Bland and Altman we determined that  $D_x$  and  $D_y$  Valsalva could not be reliably measured in our hands. Using only the test of significance, as was done by Schaer et al.,  $D_x$  and  $D_y$  Valsalva would have also been considered reliable measures in our study. Finally, we tested this method on data that represented our entire viable database, using equal parts of good, fair, and weak quality scans. Using only our best data, it is quite possible that we also may have obtained reliable measures.

In conclusion, the reliability of this protocol and ultimately of the rectangular coordinate method for measuring bladder neck position, is dependent on several variables including the conditions being measured (can bladder neck position on conditions such as kegel and valsalva be reliably measured?), the individual raters (can measures be obtained without inter-rater bias?), and when using Bland and Altman analysis, on the population being tested (are the LOAs acceptable for the size of the range of actual values?). Because the central line of the pubic symphysis is not always a reliable reference line that can be objectively predicted by more than one investigator, because the identification of the UVJ is not always obvious, particularly when the bladder is being manipulated, and, because the quality and visibility of ultrasound scans are not equal in all studies and also affect the ability of investigators to obtain reliable measures, we suggest that each study that quantifies perineal ultrasound data should test whether its data can be reliably measured.

In light of the time and expense that would be required to regularly and accurately evaluate the reliability of studies that quantify perineal ultrasound, we suggest that the greatest strength of this minimally invasive technique to evaluate pelvic floor dysfunction may not be to provide statistically sound quantitative measures, but rather to provide qualitative information; information that can be used in the care and instruction of the incontinent individual, and in generating increased understanding of the physiology and anatomy of urinary incontinence. Perineal ultrasound, with minimal discomfort or detriment to the patient, provides a wealth of information on the functional and dysfunctional pelvic floor, and should be explored to its capacity by both clinicians and researchers to advance our ability to improve the quality of life for incontinent individuals.

#### ACKNOWLEDGMENTS

We also thank the following individuals: David Ronis for his statistical expertise; Lee Park and Meg Tolbert for their efforts in the clinic and in processing the ultrasounds; Julie

Tumbarello and Dan Morgan for helpful discussions on the quantification of ultrasound; Sandy Hines and Caroline Garcia for reviewing the manuscript; and Ruta Misiunas and Dianne Bailey for their administrative support throughout this project.

#### REFERENCES

- Bland JM, Altman DG. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–10.
- Chen GD, Su TH, Lin LY. 1997. Applicability of perineal sonography in anatomical evaluation of bladder neck in women with and without genuine stress incontinence. *J Clin Ultrasound* 25:189–4.
- Kelly MA. 1979. Parturition and pelvic changes. *Am J Phys Anthropol* 51:541–6.
- Miller JM, Perucchini D, Carchidi LT, et al. 2001. Pelvic floor muscle contraction during a cough and decreased vesical neck mobility. *Obstet Gynecol* 97:255–60.
- Mouritsen L, Rasmussen A. 1993. Bladder neck mobility evaluated by vaginal ultrasonography. *Br J Urol* 71:166–71.
- Ortega HH, Munoz-de-Toro MM, Luque EH, et al. 2003. Morphological characteristics of the interpubic joint (Symphysis pubica) of rats, guinea pigs and mice in different physiological situations. A comparative study. *Cells Tissues Organs* 173:105–14.
- Peschers U, Schaer G, Anthuber C, et al. 1996. Changes in vesical neck mobility following vaginal delivery. *Obstet Gynecol* 88:1001–6.
- Peschers UM, Fanger G, Schaer GN, et al. 2001. Bladder neck mobility in continent nulliparous women. *Bjog* 108:320–4.
- Pregazzi R, Sartore A, Bortoli P, et al. 2002. Perineal ultrasound evaluation of urethral angle and bladder neck mobility in women with stress urinary incontinence. *Bjog* 109:821–7.
- Reddy AP, DeLancey JO, Zwica LM, et al. 2001. On-screen vector-based ultrasound assessment of vesical neck movement. *Am J Obstet Gynecol* 185:65–70.
- Reed H, Freeman RM, Waterfield A, et al. 2004. Prevalence of bladder neck mobility in asymptomatic non-pregnant nulliparous volunteers. *Bjog* 111:172–5.
- Schaer GN, Koechli OR, Schuessler B, et al. 1995. Perineal ultrasound for evaluating the bladder neck in urinary stress incontinence. *Obstet Gynecol* 85:220–4.
- Schaer G, Koelbl H, Voigt R, et al. 1996. Recommendations of the German Association of Urogynecology on functional sonography of the lower female urinary tract. *Int Urogynecol J Pelvic Floor Dysfunct* 7:105–8.
- Schaer GN, Perucchini D, Munz E, et al. 1999. Sonographic evaluation of the bladder neck in continent and stress-incontinent women. *Obstet Gynecol* 93:412–6.
- Troeger C, Gugger M, Holzgreve W, et al. 2003. Correlation of perineal ultrasound and lateral chain urethrocytography in the anatomical evaluation of the bladder neck. *Int Urogynecol J Pelvic Floor Dysfunct* 14:380–4.
- Tunn R, Schaer G, Peschers U, et al. 2005. Updated recommendations on ultrasonography in urogynecology. *Int Urogynecol J Pelvic Floor Dysfunct* 16:236–1.
- Yalcin OT, Hassa H, Tanir M. 2002. A new ultrasonographic method for evaluation of the results of anti-incontinence operations. *Acta Obstet Gynecol Scand* 81:151–6.