

Anthony G. Visco · John T. Wei · Leslie Ain McClure
Victoria L. Handa · Ingrid E. Nygaard

Effects of examination technique modifications on pelvic organ prolapse quantification (POP-Q) results

Received: 22 April 2002 / Accepted: 29 October 2002 / Published online: 28 March 2003

© International Urogynecological Association 2003

Abstract The pelvic organ prolapse quantification system (POP-Q) is currently the most quantitative, site-specific system for describing pelvic organ prolapse. To ensure that anatomic outcomes can be optimally assessed, investigators in the Pelvic Floor Disorders Network evaluated the impact of specific technique variations on POP-Q measurements performed on 133 patients by 16 examiners at seven sites. Values for genital hiatus and perineal body were higher when measured with maximal strain than on resting. With the exception of TVL, internal points did not differ significantly when measured with or without a speculum. The maximum extent of prolapse was best seen with the patient standing. These results suggest that genital hiatus and perineal body should be measured at rest and during straining, as the measurements may assess different

aspects of pelvic floor function, and that internal points can be measured with or without a speculum. They also emphasize the value of the standing examination to observe the maximum extent of pelvic organ prolapse.

Keywords Pelvic anatomy · Pelvic floor dysfunction · Pelvic organ prolapse

Abbreviations *POP* pelvic organ prolapse · *GH* genital hiatus · *PB* perineal body · *TVL* total vaginal length

For the Pelvic Floor Disorders Network (PFDN)
Supported by grants from the National Institute of Child Health and Human Development (U01HD41249, U10 HD41268, U10 HD41248, U10 HD41250, U10 HD41261, U10 HD41263, U10 HD41269, and U10 HD41267).

A. G. Visco
Department of Obstetrics and Gynecology,
University of North Carolina at Chapel Hill,
4036 Old Clinic Building, Chapel Hill, NC 27599-7590, USA

J. T. Wei (✉)
University of Michigan, 1500 E. Medical Center Drive,
Taubman Center Box 0330, Ann Arbor, MI 48109-0330, USA
E-mail: jtwei@umich.edu
Tel.: +1-734-6153040
Fax: +1-734-9369127

L. A. McClure
Department of Biostatistics, School of Public Health II,
University of Michigan, 1420 Washington Heights,
Room M4507, Ann Arbor, MI 48109-2029, USA

V. L. Handa
Johns Hopkins School of Medicine, 600 N. Wolfe Street,
Harvey 319, Baltimore, MD 21287, USA

I. E. Nygaard
Department of Obstetrics and Gynecology,
University of Iowa College of Medicine,
200 Hawkins Drive, Iowa City, IA 52242, USA

Introduction

The POP-Q is currently the most objective, site-specific system for quantifying and describing pelvic organ prolapse [1]. The POP-Q system describes support to the anterior, posterior and apical aspects of the vagina by measuring the distance between the most dependent part of each point and the hymeneal ring. Points above the hymen are assigned a negative value, whereas points below the hymen are positive. Two external measurements, the genital hiatus and the perineal body, complete the evaluation.

After the introduction of the POP-Q, investigators reported good reproducibility of the measures [2, 3]. Patient position did, however, affect reproducibility in that the degree of pelvic organ prolapse was higher when women were examined in a birthing chair at a 45° angle, rather than in the dorsal lithotomy position [4]. Although the original description of the POP-Q system recommended that the type of vaginal retractor be recorded and that the patient position be specified, the effect of variation in these techniques has not been previously studied.

The National Institute of Child and Human Development (NICHD)-funded Pelvic Floor Disorders Network (PFDN) is a cooperative network of investigators from seven clinical academic centers and a Data Coordinating Center (DCC). The primary goal of the PFDN is to improve the level of knowledge about pelvic floor

disorders in women, including pelvic organ prolapse, through clinical trials. The POP-Q was selected as the measurement tool to evaluate baseline and post-treatment anatomical findings. An initial survey of members of the PFDN revealed that although all were of the opinion that they were performing the POP-Q in a standardized fashion, significant variability existed in potentially important aspects of the examination, including patient position, measurement of genital hiatus and perineal body at rest or strain, and use or not of a speculum. To ensure that anatomic outcomes are assessed in a consistent manner, and to improve the reliability of POP-Q measurements, the PFDN sought to evaluate the impact of specific technique variations on POP-Q measurements. The objectives of this study were to assess: 1) whether there were differences in the POP-Q measurements of Aa, Ba, C, D, Ap, Bp and TVL obtained with and without a speculum; 2) whether there were differences in GH and PB between rest and maximum strain; and 3) whether the point of maximum prolapse was different in the lithotomy position compared to standing.

Methods

Members of the PFDN performed POP-Q measurements as outlined in the standardization document [1] on a convenience sample of women presenting for care over a 2-week period. Institutional review board approval or exemption was obtained from each clinical site. All patients were initially examined in the supine lithotomy position after emptying their bladders. Postvoid residual urine was not removed by catheterization. A rigid measuring device, such as a marked swab or sound calibrated in centimeters, was used. The size and type of speculum used was not standardized. All internal points were measured both with and without a speculum in place. The posterior blade of the speculum was used to measure Aa, Ba, Ap and Bp. An intact speculum, posterior blade or, rarely, a small vaginal dilator was used to measure the apical points C and D and TVL. For the points measured without a speculum, the vagina was manually depressed to provide exposure for the measurements. The point of maximal prolapse was assessed with the patient in both the supine lithotomy and the upright positions by asking the patient to perform a maximum Valsalva effort. External points were measured both with and without maximal straining. Examiners indicated when they were unable to obtain an accurate measurement. Study site and subject age, weight and height were recorded. Body mass index was calculated based on height and weight (kg/m^2). Women for whom all POP-Q points were complete except for point D were assumed to have undergone a hysterectomy. Stage of prolapse was determined based on measurements of maximal protrusion obtained without a speculum.

Univariate statistics were calculated for the demographic data. POP-Q measurements (taken with and without the use of a vaginal speculum; at rest or with maximal strain; and lithotomy or standing) were compared using Student's *t*-tests, Pearson's correlation coefficients, and repeated measures analysis of variance (ANOVA). ANOVA was used to determine the relationship between the POP-Q measurements and stage of prolapse, age and body mass index, controlling for variation between observers. In order to determine whether the effect of the vaginal speculum differed across examiners, an interaction term was included in the model. A weighted κ statistic was used to determine the level of agreement between the compartment of maximal prolapse in the lithotomy and standing positions.

Results are expressed as mean \pm SD. All *P* values less than 0.05 were considered statistically significant and all analyses were performed using SAS (SAS Institute Inc., Cary, NC).

Results

POP-Q examinations were performed on 133 patients by 16 examiners at seven clinical sites. Two subjects were excluded from all analyses because there were insufficient POP-Q data to calculate the stage of prolapse. For the remaining 131 subjects, their mean age was 60.8 ± 13.8 years (range 26–84), mean weight 73.3 ± 17.1 kg (range 42.5–135), and mean body mass index (BMI) was 28.0 ± 6.7 kg/m^2 (range 15.8–63.9). Eight subjects (6%) were classified as stage 0, 30 (23%) as stage I, 62 (47%), as stage II, 27 (21%) as stage III, and 4 (3%) were classified as stage IV. Because of the small sample sizes for stages 0 and IV, the patients were reclassified into three stage groups: stage 0/I (38, 29%), stage II (62, 47%) and stage III/IV (31, 24%).

Anterior and posterior vaginal wall measurements (Aa, Ba, Ap and Bp) did not differ when obtained with and without a speculum (Table 1). For all anterior and posterior positions, the correlations between measurements taken with and without a speculum were greater than 0.89. For both the anterior and posterior measurements, 94% of the values measured without a speculum were within 1 cm of the values measured with a speculum. Only 2.5% of the values were more than 2 cm apart, with no bias as to direction.

We found no difference in C or D measurements obtained with or without a speculum (Table 1). For both the C and D measurements, the correlations between measurements taken with and without a speculum were greater than 0.90. For the C measurements obtained with and without a speculum, 75% of values obtained were within 1 cm of each other. Three per cent of the values measured with a speculum were at least 2 cm greater than values measured without a speculum, whereas slightly over 5% of values obtained with a speculum were at least 2 cm less than values measured without a speculum.

There were only 52 measurements of D owing to the high rate of prior hysterectomies in this population. For the D measurements obtained with and without a speculum, 89% of values obtained were within 1 cm of each other. The remaining values were equally distributed in both directions.

Several of the patients for whom the difference between C measurements was greater than 2 cm did not have measurements for D. However, for those for whom D measurements were available, those measurements also often differed by more than 2 cm.

TVL measurements performed with a speculum were statistically significantly greater than without a speculum; however, the average difference was small (0.2 cm). The difference in TVL measurements taken with and without a speculum varied by hysterectomy status

Table 1 Anterior, posterior and apical points: effect of measurements made with and without a speculum

Measurement	n	Mean \pm SD (range)		Mean difference \pm SD (range)	P value for difference ^a
		With speculum	Without speculum		
Aa	127	-0.76 \pm 1.8 (-3, +3)	-0.72 \pm 1.9 (-3, +3)	-0.035 \pm 0.74 (-3, +2)	0.59
Ba	127	-0.48 \pm 2.4 (-3, +7)	-0.50 \pm 2.4 (-3, +7)	0.020 \pm 0.85 (-4.5, +4)	0.79
Ap	123	1.45 \pm 1.6 (-3, +3)	-1.36 \pm 1.7 (-3, +3)	0.098 \pm 0.77 (-3, +2)	0.16
Bp	123	-1.22 \pm 2.2 (-3, +3)	-1.40 \pm 2.2 (-3, +7)	0.17 \pm 1.0 (-4.5, +6)	0.06
C	126	-4.8 \pm 4.6 (-14, +7)	-4.7 \pm 4.6 (-14, -7)	-0.067 \pm 1.6 (-4, -7)	0.63
D	52	-7.1 \pm 3.7 (-15, +9)	-7.0 \pm 3.6 (-15, +8.5)	-0.029 \pm 1.4 (-2, +6)	0.88
TVL	126	8.7 \pm 1.6 (5.0, 14.0)	8.5 \pm 1.7 (4.5, 13.0)	0.20 \pm 1.0 (-2, +5.5)	0.03

^aStudent's *t*-test.

Note: *P* values from the model adjusted for examiner are similar.

Total *n* is less than 131 because in certain instances the investigator was unable to obtain a given measurement.

Table 2 GH and PB measurements at rest and with maximal straining

Measurement	n	Mean \pm SD (range)		Mean difference \pm SD (range)	P value for difference ^a
		At rest	With maximal straining		
GH	131	3.5 \pm 1.3 (1.0,7.5)	4.3 \pm 1.5 (1.0,11.0)	0.79 \pm 0.78 (-0.5,4.5)	< 0.0001
PB	131	3.2 \pm 1.1 (0.5,7.0)	3.4 \pm 1.2 (0.5, 7.5)	0.19 \pm 0.43 (-1.0,2.0)	< 0.0001

^aStudent's *t*-test.

Note: *P* values from the model adjusted for examiner are similar.

Table 3 Genital hiatus (GH) measurements at rest and on maximal straining, with stage 0–I, stage II and stage III–IV pelvic organ prolapse

Stage	n	Mean \pm SD (range)		Mean difference \pm SD (range)	P value for difference ^a
		At rest	With maximal strain		
0–I	38	2.8 \pm 1.2 (1.0,6.5)	3.3 \pm 1.2 (1.0,6.5)	0.46 \pm 0.50 (-0.5,1.5)	< 0.0001
II	62	3.5 \pm 1.1 (1.5,6.0)	4.4 \pm 1.2 (2.0,7.0)	0.84 \pm 0.64 (0.2,0)	< 0.0001
III–IV	31	4.3 \pm 1.3 (2.5,7.5)	5.4 \pm 1.6 (3.0,11.0)	1.1 \pm 1.1(0, 4.5)	< 0.0001

^aStudent's *t*-test.

(*P* = 0.04). For those subjects without a uterus, TVL was longer when measured with a speculum (8.3 \pm 1.5 cm vs. 8.0 \pm 1.5 cm, *P* = 0.022), whereas for those with a uterus there was no significant difference in TVL when measured with and without a speculum (9.3 \pm 1.7 cm vs. 9.2 \pm 1.7 cm, respectively, *P* = 0.55). Correlation between TVL measurements taken with and without a speculum was 0.82.

The relationships between genital hiatus (GH) and perineal body (PB) measurements performed during rest and maximal straining are summarized in Tables 2 and 3. The percentage difference between rest and maximal straining was similar for all stages (about 20%). The correlation between GH measurements performed during rest and during maximal straining was 0.86. For GH, 79% of values obtained at rest were within 1 cm of values measured with maximal straining. All the differences greater than 1 cm were in the same direction, i.e. with larger measurements at maximal strain. The correlation between PB measurements performed during rest and maximal strain was 0.93. Ninety-nine percent of

the PB values obtained at rest were within 1 cm of values measured with maximal straining.

We assessed the stage of prolapse both with and without a speculum in 127 women (Table 4). In 100 women (79%) the stage remained the same. For 26 patients (20%) their stage changed by one, with half of these subjects moving to a more severe stage group when measured with a speculum, whereas the other half

Table 4 Cross-tabulation of stage measured with and without a speculum

	Stage measured with a speculum					
	0	I	II	III	IV	Total
Stage measured without a speculum	0	5	2	0	0	7
I	1	24	4	0	0	29
II	0	9	46	7	0	62
III	0	0	3	22	0	25
IV	0	0	1	0	3	4
Total	6	35	54	29	3	127

Table 5 Maximal extent of prolapse in lithotomy and on standing

Stage	n	Mean \pm SD (range)		Mean d difference \pm SD (range)	P value for difference ^a
		Lithotomy	Standing		
0–I	22	-2.2 ± 0.50 (–3.0,–1.5)	-1.0 ± 2.0 (–5.0, +3.0)	1.2 ± 1.8 (–2.0,5.0)	0.006
II	53	0.18 ± 0.80 (–1.0,1.0)	1.0 ± 1.4 (–1.0,6.0)	0.81 ± 1.2 (–1.0,6.0)	< 0.0001
III–IV	30	4.0 ± 1.7 (1.5,9.0)	4.7 ± 1.8 (1.5,9.0)	0.77 ± 1.2 (–0.5,4.0)	0.0013
Total	105	0.78 ± 2.5 (–3.0,9.0)	1.6 ± 2.7 (–5.0,9.5)	0.88 ± 1.4 (–2.0,6.0)	< 0.0001

^aStudent's *t*-test.

moved to a less severe stage. The remaining subject moved two stage groups, from stage II to stage IV, when measured with a speculum rather than without.

Lastly, we compared the maximal degree of prolapse between standing and lithotomy positions. Mean maximal prolapse was significantly greater when standing than in lithotomy, overall and for each stage of prolapse, as shown in Table 5. The correlation between maximal prolapse measured in the standing and lithotomy positions was 0.86. The compartment that was maximally prolapsed in the supine position was the same as that maximally prolapsed standing 82% of the time ($\kappa = 0.72$, CI = 0.59, 0.84).

BMI and age were not associated with differences in measurements taken with and without either straining or speculum. The differences between measurements taken with and without a speculum differed significantly by examiner. For measurements taken with a speculum, investigators had different mean values. As the range is bounded for many of the measurements, differences for high values measured with the speculum tend to go in the opposite direction from those with low values measured with a speculum; this may lead to the investigator interaction that we observed.

Discussion

Our data suggest that the outcome of the POP-Q examination varies with certain aspects of examination technique. This examination system, initially described in 1996 [1], was proposed as a tool for quantifying and describing pelvic organ prolapse. It has since been used in clinical research [5, 6] and has demonstrated good reproducibility [2, 3]. The system does not, however, dictate all aspects of examination technique. Examples of possible variations in technique include [1]: the position of the patient, the type of table or chair used, the type of speculum used, the type and intensity of straining, the fullness of the bladder, and the contents of the rectum.

Our results suggest that the use of specula does not affect most aspects of the examination. Only TVL differed significantly, and the difference (0.20 cm) is unlikely to be clinically significant. This effect on TVL was limited to women with a prior hysterectomy, in whom TVL appeared significantly longer when measured with a

speculum. The use of the speculum did not significantly affect POP-Q measures at other vaginal sites. The overall stage or prolapse was affected by the use of the speculum in 20% of cases, but there was no pattern in either direction. In other words, the underestimation of stage was similar either using or not using a speculum. The observed interaction between examiner and speculum may be due to the fact that the severity of prolapse differed between patients of different investigators. Alternatively, it may represent an inherent difference in examination technique among investigators. Therefore, our results suggest that for clinical studies of pelvic organ prolapse the results of the POP-Q examination are unlikely to be affected by whether or not a speculum is used. The observed variation between examiners suggests that each investigator should maintain a consistent technique when following subjects over time.

The POP-Q system does not specify whether the external measurements (GH and PB) should be made at rest or with straining. We observed statistically significant increases in GH and PB with Valsalva, which may represent widening of the levator hiatus with increased intra-abdominal pressure. It is not clear from our data whether the widening of GH and PB with Valsalva is a normal phenomenon or evidence of pelvic floor dysfunction. Because GH and PB at rest and with maximal straining may measure different aspects of pelvic floor function, we recommend that both be measured in clinical trials.

Another technical issue regarding the POP-Q examination is whether to use a standing examination to verify that the full extent of prolapse has been observed. Three criteria for the demonstration of maximal prolapse are suggested [1]: the protrusion is taut, traction causes no further descent, and/or the subject confirms that the examination has demonstrated the full extent of the prolapse. A standing examination is recommended if these criteria are not met. Patient position has been shown to affect POP-Q examination results. Specifically, examination at a 45° angle in a birthing chair demonstrates a higher degree of prolapse than a supine examination. Our results suggest that the maximum extent of prolapse was best observed with the patient standing. Mean differences were small but statistically different. In some cases, the difference between the standing and supine examination was as great as 6 cm. Such large differences could affect the management of individual

patients. More importantly, the supine and standing examinations did not always agree with respect to identification of the most dependent portion of the prolapse. This could affect the selection of surgical approach and is therefore highly relevant to clinical management. Further study is needed to compare prolapse determination between the 45° supine and the standing positions. Although this study is the first to evaluate the effects of using a speculum and straining and on POP-Q measures, other technique modifications may also be relevant. The potential effects of particular types or sizes of specula as well as rectal and bladder volume on POP-Q assessments need to be examined. Although a population-based sampling frame was not used in this study, we took advantage of the PFDN to recruit subjects quickly from a clinical sample. Clinicians and researchers alike should understand how specific technique modifications may affect their POP-Q measures. Future work should focus on the standardization of technique as a means to improve the reliability of pelvic organ prolapse measurement.

Acknowledgements The authors thank Dr Robert Park, the Chair of the PFDN Steering Committee, for his contributions to the Network.

References

1. Bump RC, Mattiasson A, Bo K et al (1996) The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. *Am J Obstet Gynecol* 175:10–17
2. Hall AF, Theofrastous JP, Cundiff GW et al (1996) Interobserver and intraobserver reliability of the proposed International Continence Society, Society of Gynecologic Surgeons, and American Urogynecologic Society pelvic organ prolapse classification system. *Am J Obstet Gynecol* 175:1467–1470; discussion 1470–1471
3. Kobak WH, Rosenberger K, Walters MD (1996) Interobserver variation in the assessment of pelvic organ prolapse. *Int Urogynecol J* 7:121–124
4. Barber MD, Lambers A, Visco AG et al (2000) Effect of patient position on clinical evaluation of pelvic organ prolapse. *Obstet Gynecol* 96:18–22
5. Swift SE (2000) The distribution of pelvic organ support in a population of female subjects seen for routine gynecologic health care. *Am J Obstet Gynecol* 183:277–285
6. Heit M, Culligan P, Rosenquist C et al (2002) Is pelvic organ prolapse a cause of pelvic or low back pain? *Obstet Gynecol* 99:23–28

Editorial comment

This is a well conceived and clearly described study by the investigators of the Pelvic Floor Dysfunction Network. It carefully evaluates whether some of the most common variations in investigator use of the POP-Q examination results in any important differences in the measured POP-Q points. The authors demonstrate that use of a speculum rather than fingers for retraction when measuring POP-Q points does not result in any frequent or any important changes in those measurements when small numbers of patients from multiple examiners are combined for analysis. They also confirm that the standing examination is probably preferable when trying to evaluate the maximum extent of prolapse.