Levator ani defect scores and pelvic organ prolapse: is there a threshold effect?

Mitchell B. Berger · Daniel M. Morgan ·
John O. DeLancey

Received: 13 January 2014 / Accepted: 30 March 2014
© The International Urogynecological Association 2014

Abstract

Introduction and hypothesis The goal of this study was to use a well-described system of quantifying levator ani defect (LAD) severity using magnetic resonance imaging (MRI) to examine the relationship between defect severity and the presence or absence of prolapse.

Methods This is a secondary analysis of two case-control studies comparing 284 cases (with prolapse) to 219 controls (normal support) defined by using Pelvic Organ Prolapse Quantification (POP-Q) exams. LAD were assessed on MRI, with scores from 0 (no defects) to 6 (complete, bilateral defects). The number of cases and controls at each score were compared. Logistic regression and receiver operating characteristic (ROC) analyses were used to quantify relationships between LAD and prolapse.

Results The proportion of cases exceeds the overall prolapse rate in this study at LAD scores ≥ 3 , with higher rates of prolapse at higher LAD scores ($p < 0.0000001$). Prolapse risk stratifies into low risk at LAD scores 0–2, moderate at 3–5, and high at 6. ROC analysis for classification of prolapse based on LAD scores has an area under the curve of 69.9 % ($p < 0.001$), suggesting LAD alone can discriminate between normal support and prolapse for nearly 70 % of patients. Logistic regression identified higher parity and higher LAD scores as independent predictors of prolapse.

Conclusions There are three clusters of prolapse risk: low (0–2), moderate (3–5), and high (6). Although LAD have a dose-response-like effect for prolapse, other factors are clearly involved.

Keywords Levator ani defects · Magnetic resonance imaging · Pelvic organ prolapse · Threshold effect

Abbreviations

PFD	Pelvic floor disorders
POP	Pelvic organ prolapse
LAD	Levator ani defects
MRI	Magnetic resonance imaging
ROC	Receiver operating characteristic
HRT	Hormone replacement therapy

Introduction

Pelvic floor disorders (PFD), including urinary incontinence, fecal incontinence, and pelvic organ prolapse (POP), affect almost a quarter of all adult women living in the United States [1]. These disorders are associated with considerable social, quality of life (QoL), and financial burdens [2]. As the prevalence of PFDs increases with older age and the general population in the USA is aging, estimates suggest that the prevalence of these disorders will increase substantially in the near future [3].

Strong evidence exists associating PFDs and levator ani defects (LAD), and levator injury correlates most strongly with POP [4]. These levator defects occur in the pubovisceral portion of the muscle and can be detected using magnetic resonance imaging (MRI) or 3D ultrasound (US). Theoretically, there can be different degrees of injury, ranging from loss of a few fascicles to disruption of the entire muscle. It can be expected that minor injuries might be compensated for by

Presented as a poster at the 33rd Annual Scientific Meeting of the American Urogynecologic Society (3–6 October 2012, Chicago, IL USA).

M. B. Berger · D. M. Morgan · J. O. DeLancey
Pelvic Floor Research Group, Department of Obstetrics and
Gynecology, University of Michigan, Ann Arbor, MI, USA

M. B. Berger (✉)
L4100 Women's Hospital, 1500 E. Medical Center Drive, SPC 5276,
Ann Arbor, MI 48109-5276, USA
e-mail: mitchberg@umich.edu

surrounding muscle, but more major injuries may exceed the threshold beyond which sufficient remaining muscle cannot replace lost contractile force.

Several systems exist to quantify the degree of levator ani injury [4–6]. The first of these was based on an empirical system derived from a study of 29 women with pubovisceral defects after vaginal delivery [7]. It was used in a case–control study of prolapse and showed that defects involving more than half of the muscle were more likely to be found in women with prolapse [8]. However, this sample was not large enough to examine the relationships between different levels of LAD and the associated presence or absence of prolapse. The goals of this study were to use a larger sample to: (1) examine the proportion of women with POP at each LAD score, and (2) identify trends in the relationship between the presence of prolapse and LAD scores.

Materials and methods

This is a secondary analysis of two case–control studies approved by the University of Michigan Institutional Review Board comparing women with POP to asymptomatic volunteers with normal pelvic support (IRBMED #1999-0395 and 2001-0475). Study participants self-reported obstetric and gynecologic history, including gravidity and parity, history of forceps-assisted vaginal deliveries, and history of hysterectomy. All underwent clinical examination, including measurement of height and weight and Pelvic Organ Prolapse Quantification (POP-Q) examination in a semirecumbent lithotomy position [9]. Participants were classified as having prolapse (cases) if any POP-Q point was at least 1 cm beyond the hymen; controls had all POP-Q points at least 1 cm above the hymen. Controls were recruited to maintain demographic matching by age, race, and hysterectomy status; they were selected to be within 5 years of age and of the same race as cases. Attempts were made to match for parity as well, with the goal being to recruit controls having similar parity (within one delivery) with a case. However, we had difficulty recruiting high-parity controls who met our POP-Q definition. Therefore, strict matching for parity between cases and controls was not possible. Patients whose most dependent POP-Q points were <1 cm from the hymen were excluded from this secondary analysis in order to maintain strict definitions of prolapse and normal vaginal support, in accordance with previously published work [8]. Individuals were also excluded if magnetic resonance imaging (MRIs) of sufficient quality to grade pubovisceral defects were not available.

Pelvic MRIs were obtained for all individuals. Complete details of MRI acquisition were published previously [8] and included 5-mm-spaced fast-spin proton-density images in the axial, sagittal, and coronal planes using a 1.5- or 3-T superconducting magnet. LADs were graded on MR scans by two independent investigators blinded to case–control

status [5]. The right and left pubovisceral muscles were each graded separately, with possible scores of 0 (no defect), 1 (less than half of the muscle bulk missing), 2 (more than half of the muscle missing but not a complete defect), or 3 (complete loss of muscle). The total calculated levator defect score was the sum of scores from each side, with a possible range of 0–6. Individuals with a total LAD score of 0 were classified as having no defects. Women with an LAD score of 1–3, except those with a unilateral grade 3 defect were considered to have minor defects, and those with scores of 4–6 or unilateral grade 3 defects were classified as having major defects [5].

Continuous variables were compared using *t* tests or Kruskal–Wallis tests, as appropriate. Categorical variables were compared using chi-squared tests. Testing for a nonlinear trend suggestive of a dose–response effect was performed using the extended Mantel–Haenszel chi-squared analysis. Receiver operating characteristic (ROC) analysis was used to model the discriminatory performance for predicting POP. Binary logistic regression was used to generate a model predicting POP based on demographics, obstetric and gynecologic history, and MRI-derived LAD scores. Multivariate logistic regression was then used to generate the most parsimonious model for predicting POP, starting with the variables significantly associated with prolapse in bivariate regression analysis. Extended Mantel–Haenszel chi-squared analysis was performed with the OpenEpi statistical package [10]. All other statistical analyses were performed using SPSS Statistics, version 20 (IBM, Armonk, NY, USA). *P* values < 0.05 were considered statistically significant.

Results

MR images from 503 individuals were reviewed. There were 219 controls (43.5 %) with normal pelvic support and 284

Table 1 Demographic characteristics

Characteristic	Normal support (<i>n</i> =219)	Prolapse (<i>n</i> =284)	<i>P</i> value
Age (years)	56.3±11.2	56.9±12.2	0.60
Gravidity	3.2±1.9	3.5±2.0	0.06
Parity	2.5±1.5	2.9±1.7	0.005
BMI (kg/m ²)	26.6±5.0	27.0±5.3	0.40
Caucasian (%)	89.4 (193/216)	90.5 (256/283)	0.76
Postmenopausal (%)	74.8 (154/206)	70.0 (196/280)	0.26
FAVD (%)	26.2 (51/195)	44.3 (116/262)	< 0.001
Hysterectomy (%)	19.5 (42/215)	21.1 (60/284)	0.74
Using HRT (%)	13.7 (29/212)	22.0 (62/282)	0.02

Data are mean ± standard deviation or percentage (*n/n*).

BMI body mass index, *FAVD* forceps-assisted vaginal delivery, *HRT* hormone replacement therapy

Fig. 1 **a** Number of cases (black bars) and controls (white bars) at each levator ani defect (LAD) score. **b** Proportion of cases (black bars) and controls (white bars) for each LAD score. Dashed line indicates overall proportion of individuals with prolapse at least 1 cm beyond the hymen (56.5 %)

cases (56.5 %) with prolapse at least 1 cm beyond the hymen. Demographic characteristics of both groups are presented in Table 1. Age, body mass index (BMI), gravidity, race, menopausal status, and proportion of individuals who had had a hysterectomy were similar in both groups. The mean parity of the group with prolapse was statistically significantly higher than the group with normal support, but the actual difference was minimal. A greater proportion of cases reported a history of forceps assistance during vaginal delivery than did controls. Relatively few women in each group self-reported using hormone replacement therapy (HRT), although the proportion in the prolapse group was significantly higher than that in the control group.

Women with normal pelvic support and those with prolapse at least 1 cm beyond the hymen were identified with all seven possible LAD scores, as originally proposed [7]. The number of cases and controls in each LAD score group are shown in Fig. 1a. The proportion of cases compared with controls increases as the LAD score rises (Fig. 1b), with a dose-response-like effect (the extended Mantel–Haenszel χ^2 for linear trend=69.65, $p<0.0000001$).

Receiver operating characteristic (ROC) analysis was next used to model the use of LAD scores alone for predicting POP. The ROC curve generated from our data is shown in Fig. 2. The area under the ROC (AUC) curve is

Fig. 2 Receiver operating characteristic curve for predicting prolapse using levator ani defect scores

69.9 %, with a p value <0.001 . Likelihood ratios for LAD scores in predicting prolapse were also calculated (Table 2).

Logistic regression was then performed to identify the factors independently associated with prolapse. We first used bivariate analyses to distinguish characteristics that significantly associate with the presence of prolapse, testing age, parity, BMI, race, postmenopausal status, use of HRT, history of hysterectomy, history of forceps-assisted vaginal delivery (FAVD), and total LAD score. The factors that were identified as statistically-significant included parity, use of HRT, history of FAVD, and LAD score (Table 3). Multivariate logistic regression was then performed, starting with the characteristics identified as significant in bivariate analyses. The final model for predicting prolapse includes parity and LAD score (Table 3). There was no statistically significant interaction term between parity and LA defect score ($p=0.49$). The final model was statistically significant ($\chi^2=82.42$, $p<0.001$, degrees of freedom=2) and able to distinguish between women with and without prolapse. Goodness of fit for this model, as tested by the Hosmer–Lemeshow test, was also supported ($p=0.07$).

A total LAD score of 3 can be classified as either minor or major, depending on whether it results from bilateral (LAD score 1 on one side and 2 on the other) or unilateral (LAD score 3 on one side and a normal contralateral muscle) defects, respectively [7]. We therefore separately analyzed women with LAD scores of 3, comparing the proportion with prolapse

Table 2 Likelihood ratios for levator ani defect (LAD) scores in diagnosing pelvic organ prolapse

LAD score	Likelihood ratio
0	0.5
1	0.5
2	0.9
3	2.3
4	1.7
5	2.2
6	9.3

Table 3 Logistic regression models predicting pelvic organ prolapse

Independent variable	Odds ratio	95 % CI	Regression coefficient	Standard error	P value
Univariate analyses					
Age (years)	1.00	0.99–1.02	0.004	0.008	0.60
Parity	1.19	1.05–1.33	0.17	0.06	0.007
BMI (kg/m ²)	1.02	0.98–1.05	0.02	0.02	0.40
Race (Caucasian vs. non-Caucasian)	1.13	0.63–2.03	0.12	0.30	0.68
Postmenopausal	0.79	0.53–1.18	-0.24	0.21	0.25
Use of HRT	1.78	1.10–2.88	0.58	0.25	0.02
Hysterectomy	1.10	0.71–1.72	0.10	0.23	0.66
FAVD	2.24	1.50–3.35	0.81	0.21	< 0.001
Total LAD score	1.47	1.34–1.62	0.39	0.05	< 0.001
Multivariate analysis					
Intercept	–	–	-0.96	0.22	< 0.001
Parity	1.20	1.06–1.36	0.18	0.07	0.005
Total LAD score	1.47	1.34–1.62	0.39	0.05	< 0.001

Forceps delivery is coded as 1 if the woman recalled having had at least one FAVD and 0 if she never had a forceps delivery.

CI confidence interval, BMI body mass index, HRT hormone replacement therapy, FAVD forceps-assisted vaginal delivery, LAD levator ani defect

in the group with bilateral defects to those with unilateral defects: 70.4 % (19/27) of women with minor defects had prolapse, compared with 84.6 % (11/13) of those with a unilateral complete defect ($p=0.33$).

Discussion

Results of this analysis comparing women with clinically symptomatic POP descending at least 1 cm beyond the hymen and asymptomatic volunteer controls with normal pelvic support from case-control studies confirm that rates of prolapse increase as LAD scores worsen. The proportion of individuals with prolapse ranges from 41.5 % in the group with normal muscles to 92.3 % in the group with total LAD scores of 6. Our results indicate an LAD score of 3 as a specific threshold, given that women with scores ranging from 0 to 2 have rates of prolapse lower than expected; for those with scores ≥ 3 , the proportion with prolapse was higher than expected. Furthermore, LAD severity had a dose-response-like effect in relation to the development of POP.

Several studies have shown an association between higher rates of prolapse and worsening LADs. This analysis, using MRI, is consistent with findings of LAD assessed by endovaginal US [6, 11, 12]. This study extends the literature by quantifying the relationship between LAD degree and prolapse using a well-described technique of high inter-rater reliability [5]. Our analysis allows for a more precise prediction of whether a woman has prolapse based on her levator status [13]. For example, presuming a pretest probability of prolapse (using population-based data from the USA) of 2.9 % [1], the likelihood ratios from this study would suggest that a

woman with no LAD has a posttest probability of prolapse of just <2 %, whereas the probability of prolapse for a woman with a total LAD score of 6 is almost 30 %.

The groupings of no LAD or a normal levator ani and minor and major LADs were developed empirically to describe the severity according to MRI findings, including the ability of a total LAD score of 3 to be classified as either a major or minor defect, depending on how the defects lateralize [7]. We could not determine whether LAD scores of 3 should be categorized in this fashion due to the small number of women with total LAD scores of 3. However, whereas our sample size was not adequate to resolve this issue definitively, it does provide evidence that the language to describe the clusters of LAD scores might be best revised to reflect the risk of prolapse. Women with scores of 0–2 (normal and very minor defects) are at low risk for prolapse, those with scores 3–5 are at moderate risk, and those with a score of 6 have the highest risk. We should stress that this study was not intended to suggest that MRI be used to diagnose or predict prolapse or LADs in clinical practice. Rather, results help quantify the relationship between LADs and POP.

Although LADs are strongly associated with prolapse, there are clearly other factors involved. ROC analysis suggested that LAD scores alone explain prolapse in only 70 % of women. This emphasizes the importance of other factors, such as tissue strength and composition, genetics, non-levator-related changes after giving birth, and medical comorbidities [14, 15].

Strengths of this study are the large sample size, the validated technique of LAD scoring, assessment of pelvic support by physical examination, and use of asymptomatic, healthy volunteers as controls rather than women seeking healthcare

for other reasons. By contrast, controls in previously published studies using endovaginal US to quantify LAD all involved women seeking care for pelvic floor dysfunction [6, 11, 12].

Several factors must be kept in mind when interpreting our results. This was a case–control study with demographic group matching, not a population-based sample performed at a single time point. We are, therefore, unable to examine cause–effect relationships in these parameters or predict results from a population-based sample. We excluded women whose maximal prolapse was between 1 cm above and 1 cm below the hymen, potentially limiting the generalizability of our findings. Generalizability may be further limited by the relative homogeneity of our participants, who were recruited for two separate studies conducted at different times, so this may have introduced bias; for example, MRI acquisition with a 1.5-T magnet in the original study and a 3-T magnet in the later study. However, slice thickness was held constant during that transition in order to minimize the risk of introducing error. Finally, all studies using image interpretation are subject to inherent limitations of image analysis.

Conclusion

LADs represent one of the strongest risk factors for POP, with higher rates of prolapse being associated with increasing severity of these muscle defects. In fact, 70 % of the time, LAD severity alone can accurately predict whether a woman has prolapse. Although LAD scores associate with prolapse in a dose–response-like fashion, the probability of prolapse can be categorized into three clusters of defect scores. Future research is needed to quantify the contribution of other risk factors to a woman's overall risk of developing POP.

Conflict of interest This research was funded by NICHD grant R01 HD38665 and ORWH grant P50 HD44406. Dr. DeLancey has consulted with Proctor & Gamble during the last year on topics not relevant to this manuscript. No conflicts of interest are declared for Drs. Berger and Morgan.

References

1. Nygaard I, Barber MD, Burgio KL, Kenton K, Meikle S, Schaffer J, Spino C, Whitehead WE, Wu J, Brody DJ (2008) Prevalence of symptomatic pelvic floor disorders in US women. *JAMA* 300: 1311–1316
2. Sung VW, Hampton BS (2009) Epidemiology of pelvic floor dysfunction. *Obstet Gynecol Clin N Am* 36:421–443
3. Wu JM, Hundley AF, Fulton RG, Myers ER (2009) Forecasting the prevalence of pelvic floor disorders in U.S. Women: 2010 to 2050. *Obstet Gynecol* 114:1278–1283
4. Lammers K, Futterer JJ, Prokop M, Vierhout ME, Kluivers KB (2012) Diagnosing pubovisceral avulsions: a systematic review of the clinical relevance of a prevalent anatomical defect. *Int Urogynecol J* 23:1653–1664
5. Morgan DM, Umek W, Stein T, Hsu Y, Guire K, DeLancey JO (2007) Interrater reliability of assessing levator ani muscle defects with magnetic resonance images. *Int Urogynecol J Pelvic Floor Dysfunct* 18:773–778
6. Dietz HP (2007) Quantification of major morphological abnormalities of the levator ani. *Ultrasound Obstet Gynecol* 29:329–334
7. Kearney R, Miller JM, Ashton-Miller JA, DeLancey JO (2006) Obstetric factors associated with levator ani muscle injury after vaginal birth. *Obstet Gynecol* 107:144–149
8. DeLancey JO, Morgan DM, Fenner DE, Kearney R, Guire K, Miller JM, Hussain H, Umek W, Hsu Y, Ashton-Miller JA (2007) Comparison of levator ani muscle defects and function in women with and without pelvic organ prolapse. *Obstet Gynecol* 109:295–302
9. Bump RC, Mattiasson A, Bo K, Brubaker LP, DeLancey JO, Klarskov P, Shull BL, Smith AR (1996) The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. *Am J Obstet Gynecol* 175:10–17
10. Dean AG, Sullivan KM, Soe MM (last updated 2013/04/06) OpenEpi: Open Source Epidemiologic Statistics for Public Health, www.OpenEpi.com. Accessed 05 November 2013
11. Dietz HP, Simpson JM (2008) Levator trauma is associated with pelvic organ prolapse. *BJOG* 115:979–984
12. Rostaminia G, White D, Hegde A, Quiroz LH, Davila GW, Shobeiri SA (2013) Levator ani deficiency and pelvic organ prolapse severity. *Obstet Gynecol* 121:1017–1024
13. Grimes DA, Schulz KF (2005) Refining clinical diagnosis with likelihood ratios. *Lancet* 365:1500–1505
14. Kerkhof MH, Hendriks L, Brolmann HA (2009) Changes in connective tissue in patients with pelvic organ prolapse—a review of the current literature. *Int Urogynecol J* 20:461–474
15. de Boer TA, Sliker-Ten Hove MC, Burger CW, Kluivers KB, Vierhout ME (2011) The prevalence and factors associated with previous surgery for pelvic organ prolapse and/or urinary incontinence in a cross-sectional study in The Netherlands. *Eur J Obstet Gynecol Reprod Biol* 158:343–349