

## ORIGINAL RESEARCH ARTICLE

# Three-dimensional endovaginal ultrasound assessment using the levator ani deficiency score in primiparas: A replication study

Emilia Rotstein<sup>1,2</sup>  | Vilhelmina Ullemar<sup>1,2</sup> | Marianne Starck<sup>3</sup> | Gunilla Tegerstedt<sup>1,2</sup>

<sup>1</sup>Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden

<sup>2</sup>Department of Women's Health, Karolinska University Hospital, Stockholm, Sweden

<sup>3</sup>Department of Surgery, Pelvic Floor Center, Skåne University Hospital, Malmö, Sweden

**Correspondence**

Emilia Rotstein, Karolinska Pelvic Floor Center-B79, Department of Gynecology and Reproductive Medicine, Karolinska University Hospital, 141 86 Stockholm, Sweden.

Email: [emilia.rotstein@ki.se](mailto:emilia.rotstein@ki.se)

**Funding information**

Karolinska Institutet

**Abstract**

**Introduction:** It is essential to assess the levator ani properly as part of clinical care in patients presenting with pelvic floor dysfunction. The levator ani deficiency scoring system is a previously published method to assess levator ani defects with three-dimensional endovaginal ultrasound. The primary aim of this study was to determine the intra- and interrater reliability of the levator ani deficiency score in a cohort of non-instrumentally delivered primiparas.

**Material and methods:** Primiparas ( $n = 141$ ) were examined at least 1 year after vaginal birth. Three-dimensional endovaginal ultrasound volumes were acquired by a single examiner using two different automated ultrasound probes. The volumes were analyzed by two separate raters who were blinded to each other's assessments. Descriptive statistics were calculated for levator ani deficiency score and categorized into three levels (mild, moderate, severe). Kendall's tau-b was calculated for intra- and interrater comparisons.

**Results:** Intrarater comparisons of levator ani deficiency score and levator ani deficiency category were high (Kendall's tau-b  $\geq 0.80$  for Rater 1;  $> 0.79$  for Rater 2). Interrater comparisons of levator ani deficiency score and levator ani deficiency category were also high (Kendall's tau-b  $> 0.9$  for assessment 1 and  $> 0.78$  for assessment 2). Varying by rater, probe and assessment, 75.9%–80.1% of the study population had no/mild deficiency, 6.4%–9.2% had moderate deficiency, and 4.3%–6.4% had severe levator ani deficiency.

**Conclusions:** The levator ani deficiency scoring system is a feasible method to assess defects of the levator ani muscle and can be reproduced with high intra- and interrater correlations. Using the scoring system in clinical practice may facilitate concordant assessment between different examiners. However, the system should be used to support clinical findings and symptomatology and not as a screening tool, as the score is lacking the category of no levator ani deficiency.

**Abbreviations:** 3D, three-dimensional; EVUS, endovaginal ultrasonography; LAD, levator ani deficiency; LAM, levator ani muscle; MRI, magnetic resonance imaging.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Acta Obstetrica et Gynecologica Scandinavica* published by John Wiley & Sons Ltd on behalf of Nordic Federation of Societies of Obstetrics and Gynecology (NFOG).

**KEYWORDS**

levator ani avulsion, levator ani muscle, pelvic floor dysfunction, systematic scoring, ultrasound, vaginal birth

## 1 | INTRODUCTION

Vaginal childbirth affects the levator ani muscle (LAM) and avulsion of the puborectal and pubo-/iliococcygeal muscles from the pubic bone is found in 10%–35% of women after vaginal delivery. This may subsequently lead to pelvic floor dysfunction such as pelvic organ prolapse and urinary and/or fecal incontinence.<sup>1–3</sup> In addition, the presence of levator ani deficiency (LAD) such as avulsion impairs results of prolapse surgery with increased risk of recurrence, and it is also associated with colorectal conditions such as intussusception.<sup>4–6</sup> Muscle defects diagnosed in magnetic resonance imaging (MRI) or ultrasound have mainly been attributed to obstetric history rather than being related to age or hormonal changes.<sup>7</sup> Using the term deficiency rather than avulsion or defect suggests that muscle impairment should be assessed and measured on a gradient rather than applying dichotomous terminology.<sup>8</sup>

A systematic, repeatable scoring system increases the quality of information given to patients and is paramount to optimal care.<sup>9</sup> Several scoring systems for levator assessment have been validated over the last two decades; however, none used three-dimensional (3D) endovaginal ultrasound (EVUS). DeLancey et al. described levator ani defects using MRI and scored left and right muscle defects separately, focusing specifically on the severity of a defect.<sup>10</sup> In addition, Dietz et al. described a scoring system using 3D transperineal ultrasound where defects were evaluated by the number of tomographic slices with muscle discontinuity.<sup>11</sup> The LAD scoring system was described by Rostaminia et al. as an assessment of the appearance of the LAM subdivisions to score LAD by 3D EVUS.<sup>12,13</sup> To date, the LAD score is the only structured system for endovaginal 3D ultrasound assessment of the levator ani muscle. Moreover, 3D EVUS is valuable when assessing LAM in women with pelvic floor dysfunction to help patients make an informed decision regarding risks and benefits of suggested treatment.<sup>14</sup>

The purpose of this study was to evaluate the reproducibility of the LAD score in a cohort of primiparas with non-instrumentally assisted deliveries and to assess the intra- and interrater reliability.

## 2 | MATERIAL AND METHODS

### 2.1 | Study population

The study population was recruited as part of an experimental cohort study (MIMA, the Midwives' Management during the Second Stage of Labor in Relation to Second-Degree Tears – An Experimental Study) where an intervention in delivery strategy was compared with standard delivery practice at two delivery clinics

### Key message

The levator ani deficiency score is easy to reproduce with high correlations between different raters allowing for improved clinical communication.

in Stockholm.<sup>15</sup> Inclusion criteria were Swedish-speaking women at gestational age  $\geq 37$  full weeks of pregnancy, with spontaneous onset or induction of labor. All women included in the original study ( $n = 597$ ) were invited to a 3D EVUS 1 year after delivery.

### 2.2 | Ultrasound technique

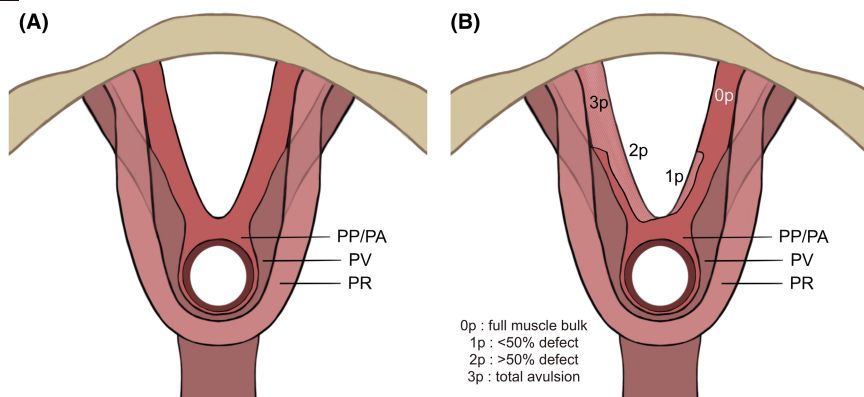
All participants underwent gynecological examination including a 360° 3D high-resolution EVUS performed with two different probes by the same urogynecologist (Rater 1). Examinations were performed in an office setting, with the patient in dorsal lithotomy position, hips flexed and abducted. The probe was inserted in a neutral position to minimize pressure on surrounding structures. Images were rendered during rest and stored digitally for analysis. A BKmedical Flexfocus machine was used.

In order to test for equal reliability of the LAD score regardless of probe, the 3D EVUS was performed with two separate probes that were chosen because they are the most commonly used in our clinical settings. Probe 1 (BK 2052 6–16 MHz) has an internal automated motorized system allowing an acquisition of 60 mm consisting of 300 transaxially aligned 2D images of 0.2 mm each. Probe 2 (BK 8838 6–16 MHz) has a built-in automatic linear array 360° acquisition of 1440 2D images of 0.25° each. Both transducers allow 3D acquisition of 2D images without any movement of the probe within the cavity. A set-up of 9 MHz was used for both probes.

### 2.3 | Scoring of LAM

Volumes were assessed offline and independently by two assessors; ER (Rater 1) with 4 years of experience in ultrasound assessment, and MS (Rater 2) with 15 years of experience. Both raters were blinded to patient history, clinical data and one another's assessments.

LAM was divided into three subgroups based on published work by Rostaminia et al.: the puboperinealis/puboanalisis (PP/PA), the puborectalis (PR) and the pubococcygeus/iliococcygeus (PV), as shown in Figure 1A.<sup>13</sup> These were evaluated bilaterally in the specific axial plane that visualized the full length of each of muscle pair, and were



**FIGURE 1** (A) Schematic image of levator ani subsections (A) and scoring system (B). The puboperinealis/puboanal (PP/PA), puborectalis (PR), and pubococcygeus/iliococcygeus (PV) are shown respectively in (A). (B) The scoring system in one of the three subdivisions of the levator ani muscle (puboperineal/puboanal muscle). (B) Schematic image of levator ani subsections (A) and scoring system (B). The puboperinealis/puboanal (PP/PA), puborectalis (PR), and pubococcygeus/iliococcygeus (PV) are shown respectively in (A). (B) The scoring system in one of the three subdivisions of the levator ani muscle (puboperineal/puboanal muscle).

scored (0=no defect, 1=minimal defect with <50% muscle loss, 2=major defect with >50% loss, 3=total absence) on each side based on thickness and detachment from the pubic bone (Figure 1B). A maximum score of 9 points per side indicated total absence of muscle, resulting in a cumulative LAD score between 0 and 18. The LAD score was categorized into mild (0–6 points), moderate (7–12 points) and severe (13–18 points).

An EVUS volume was deemed interpretable if it included the pubic symphysis and levator ani plate in an anterior–posterior array, and all LAM subdivisions. If non-interpretable, the volume was not scored and resulted in a missing value. Examples of interpretable EVUS for both probes are shown in Figure 2.

To allow for both intra- and interrater comparisons, each rater scored the same volume on two separate occasions (assessments 1 and 2). Volumes of the same patient using different probes were scored in a random order. Time elapsed between assessments of the same volume was at least 4 weeks. A maximum of 10 examinations were scored during each sitting.

## 2.4 | Statistical analyses

LAD scores were summarized for each rater assessment and probe separately and categorized into the previously mentioned categories (mild, moderate and severe). Descriptive statistics for LAD score and LAD category were presented. Because LAD score was non-normally distributed, the mean and median along with mean absolute deviation were presented.

Intra- and interrater assessments were compared for both probes. A schematic overview is presented in Figure 3. Correlations were calculated using Kendall's tau-b, a non-parametric rank-based correlation allowing for ties between ratings (as several volumes could receive the same LAD score) and suitable for underlying ordinal data. Assuming perfect correlation, Kendall's tau-b equals 1. This occurs when all pairs are *concordant*, meaning volume x is ranked

lower than volume y by both raters. Conversely, if all pairs are *discordant* (ie Rater 1 has ranked volume x higher than volume y, but Rater 2 has ranked volume x lower than volume y), Kendall's tau-b would equal  $-1$ . In case of an equal number of discordant and concordant pairs, the value is equal to 0. We considered a value of Kendall's tau-b  $>0.8$  to be a satisfactory cut-off in clinical settings.<sup>13</sup>

A P-value of  $<0.05$  was considered statistically significant. All statistical analyses were performed using IBM SPSS Statistics, version 27.0.

## 2.5 | Ethics statement

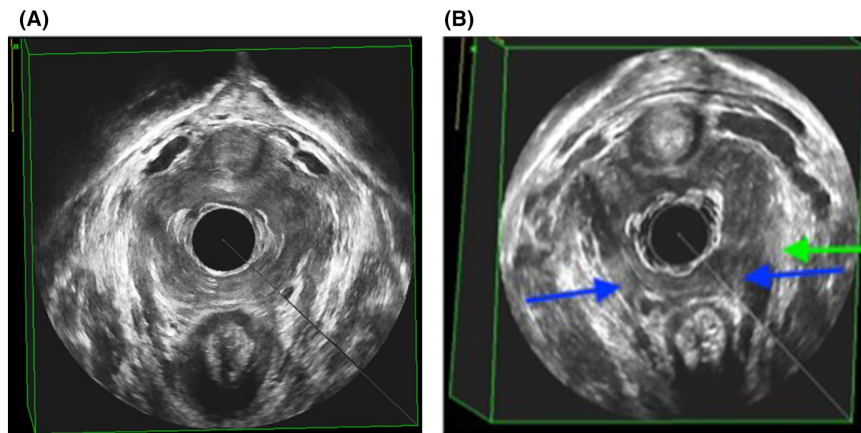
Ethical approval was obtained from the Regional Ethics Committee (Dnr: 2013/859–31/2 on June 4, 2013 and 2017/472–32 on March 3, 2017). Participants gave written informed consent and could withdraw from the study at any time.

## 3 | RESULTS

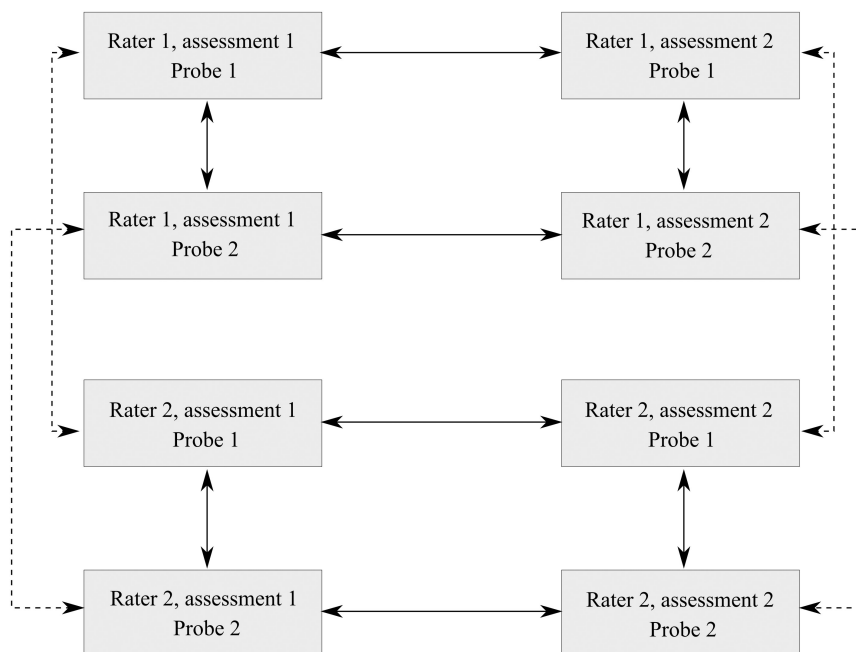
From April 2015 to June 2016, 275 women accepted the invitation. Of these, 31 were excluded due to novel pregnancies and 103 were lost to follow-up (12 due to migration; 91 after not making an appointment following two reminders), leaving 141 primiparas to be examined 1–2 years after non-instrumentally assisted vaginal deliveries.

General descriptive characteristics (age, body mass index at time of examination, time since vaginal delivery) are presented in Table 1. The variables “age” and “time since vaginal delivery” were missing in 7 (5%) and 13 (9.2%) cases, respectively. For body mass index, weight or height information was missing in 16 (11.3%). Table 2 shows the distribution of LAD score and LAD category by rater assessment and probe (intra-rater assessments). The proportion of missing ultrasound volumes ranged from 6.4% to 10.6% depending on probe and rater.

**FIGURE 2** (A) Example of an interpretable endovaginal ultrasound volume; levator ani muscle intact. (B) Example of an interpretable endovaginal ultrasound volume, LAD score 8p. Blue arrows indicating bilateral PP/PA defect (3 p each side) and green arrow indicating left-sided PV-defect (2 p).



**FIGURE 3** Schematic overview of the comparisons. Continuous lines represent intra-rater comparisons, results of which are shown in Table 3. Dashed lines represent interrater comparisons, results of which are shown in Table 4.



**TABLE 1** Background characteristics for the study population, including percentage with available data for each variable, range and mean (including standard deviation).

	n (%)	Min	Max	Mean ( $\pm$ SD)
Age	134 (95.0)	21.00	45.00	30.58 (3.91)
BMI	125 (88.7)	17.40	35.90	22.83 (3.36)
Time from delivery to examination (months)	128 (90.8)	13.09	26.45	18.82 (2.80)

The mean LAD score for the first assessment by Rater 1 using probe 1 was 1.89, with mean absolute deviation 2.97. Categorizing the scores by Rater 1, 107 (75.9%) were mild, 12 (8.5%) moderate and 7 (5%) severe. In general, the proportions of examinations divided into the three categories were similar between probes and raters. Table 3 shows the intra-rater correlations of the LAD score and LAD category.

Overall, LAD score correlations were high, with Kendall's tau-b > 0.78 for all intra-rater comparisons. Specifically, they ranged from 0.79 (for Rater 2, probe 2, assessment 1 vs 2) to 1.00 (for Rater

1, probe 1, assessment 1 vs 2). Correlations between LAD categories ranged between 0.80 (Rater 2, probe 1, assessment 1 vs 2) to 1.00 (Rater 1, probe 1, assessment 1 vs 2). All comparisons were statistically significant ( $P < 0.01$ ).

Finally, Table 4 displays a high interrater correlation of LAD score between the two raters in the first assessment (Kendall's tau-b > 0.9,  $P < 0.01$ ), with a perfect correlation (Kendall's tau-b = 1.00,  $P < 0.01$ ) for the LAD category. For the second assessment, interrater correlations for the LAD score were 0.78 for probe 1 compared with 0.81 for probe 2.

**TABLE 2** Descriptive statistics: mean, median, mean absolute deviation (MAD) and range for LAD score, percentages of LAD category, for each rater assessment and probe.

	LAD score				LAD category			
	Mean	Median	MAD	Range	Mild, n (%)	Moderate, n (%)	Severe, n (%)	Missing, n (%)
<b>Assessment 1</b>								
Rater 1, probe 1	1.89	0	2.97	0–17	107 (75.9)	12 (8.5)	7 (5.0)	15 (10.6)
Rater 1, probe 2	1.82	0	2.87	0–16	113 (80.1)	11 (7.8)	8 (5.7)	9 (6.4)
Rater 2, probe 1	1.67	0	2.77	0–16	110 (78.0)	13 (9.2)	6 (4.3)	12 (8.5)
Rater 2, probe 2	1.77	0	2.87	0–16	112 (79.4)	11 (7.8)	8 (5.7)	10 (7.1)
<b>Assessment 2</b>								
Rater 1, probe 1	1.88	0	2.94	0–17	107 (75.9)	12 (8.5)	7 (5.0)	15 (10.6)
Rater 1, probe 2	1.78	0	2.86	0–16	112 (79.4)	10 (7.1)	9 (6.4)	10 (7.1)
Rater 2, probe 1	1.55	0	2.54	0–17	113 (80.1)	9 (6.4)	8 (5.7)	11 (7.8)
Rater 2, probe 2	1.66	0	2.76	0–17	113 (80.1)	9 (6.4)	9 (6.4)	10 (7.1)

	n (% of total study population)	LAD score	P-value	LAD category	P-value
<b>Rater 1</b>					
Probe 1, assessment 1 vs 2	126 (89.4)	0.997	<0.01	1.000	<0.01
Probe 2, assessment 1 vs 2	131 (92.9)	0.944	<0.01	0.996	<0.01
Probe 1 vs 2, assessment 1	125 (88.7)	0.800	<0.01	0.840	<0.01
Probe 1 vs 2, assessment 2	124 (87.9)	0.816	<0.01	0.837	<0.01
<b>Rater 2</b>					
Probe 1, assessment 1 vs 2	122 (86.5)	0.860	<0.01	0.801	<0.01
Probe 2, assessment 1 vs 2	127 (90.1)	0.787	<0.01	0.863	<0.01
Probe 1 vs 2, assessment 1	124 (87.9)	0.832	<0.01	0.831	<0.01
Probe 1 vs 2, assessment 2	126 (89.4)	0.964	<0.01	0.848	<0.01

**TABLE 3** Kendall's tau-b correlations for intra-rater comparisons of LAD score and LAD category, by probe and rater assessment.

	n (% of total study population)	LAD score	P-value	LAD category	P-value
<b>Assessment 1</b>					
Probe 1, rater 1 vs 2	124 (87.9)	0.902	<0.01	1.000	<0.01
Probe 2, rater 1 vs 2	131 (92.9)	0.966	<0.01	1.000	<0.01
<b>Assessment 2</b>					
Probe 1, rater 1 vs 2	123 (87.2)	0.782	<0.01	0.813	<0.01
Probe 2, rater 1 vs 2	127 (90.1)	0.814	<0.01	0.860	<0.01

**TABLE 4** Kendall's tau-b correlations for interrater comparisons of LAD score and LAD category, by probe and rater assessment.

## 4 | DISCUSSION

Our results show that the LAD scoring system is reproducible and performs well in settings other than that for which it was originally created.

As expected, the rate of LAD in our low-risk population consisting of primiparas was low. Only 6%–9% fulfilled the criteria for moderate LAD and 4%–6% for severe LAD. This is in line with previous research of the prevalence of levator ani defects after childbirth.<sup>3,16</sup> Van Delft et al. studied LAM avulsion postpartum, reporting a

prevalence of approximately 21% avulsions among 191 patients examined at 36 weeks of pregnancy and 3 months postpartum. They showed that many partial levator avulsions diagnosed in the early postpartum period displayed a regression of symptoms and ultrasound findings during the first year after childbirth, and the authors therefore advised expectancy from diagnosing avulsions by ultrasound until 1 year after delivery.<sup>17</sup>

We found high intra- and interrater correlations for all assessments and probes, using both the raw LAD score and the LAD category. The fact that we were able to achieve similar results as

Rostaminia et al. in an independent setting further strengthens the merits of the method.

In comparison, Santoro et al. presented a method of assessing LAD by ultrasound using the probe identified as probe 1 in our study. Interrater comparisons were studied across different medical specialties; overall, interrater repeatability was good to excellent for levator ani injuries.<sup>18</sup> However, in that study the measurements focused on the dimensions of the levator ani hiatus as well as ano-rectal angle and urethral thickness. The study neither assessed the muscle bulk of the LAM nor is applicable in cases of muscle avulsion, as there is no identifiable perimeter of the hiatal measurements in those cases. In general, LAD category (rather than LAD score) is used to guide the choice of intervention in clinical settings.<sup>19</sup>

Our findings draw the attention to the limitation of the LAD score in a low-risk population given that the score has only three categories, with mild LAD also covering cases with a score of 0. In our opinion, this is a weakness of the scoring system and should be acknowledged if used in a population with unknown status of pelvic floor dysfunction. We suggest the addition of a fourth category of “no LAD”, as this might improve its usability in heterogeneous samples. In a clinical setting, the LAD score should only be used in women with clinical findings of LAD to further strengthen the diagnosis. Currently, no gold standard exists for diagnosis and classification of levator ani defects across imaging modalities. Vergeldt et al. suggested focusing on correlation to clinical outcome rather than comparison of rating systems of different modalities, increasing the value for individual patients.<sup>20</sup> Nevertheless, introducing a consistent scoring method (such as the LAD score) is useful for consistent clinical care and in supporting structured learning and knowledge transfer.<sup>10</sup>

With imaging techniques becoming increasingly accessible in office settings, the challenge arises of how to interpret and compare descriptions of LAM defects in different modalities. There is a general consensus that distinguishing major defects is of the highest clinical relevance, as they are associated both with the development of pelvic organ prolapse as well as the degree of its severity.<sup>2,21,22</sup> Both MRI and transperineal ultrasound findings have been shown to correlate with clinical findings in terms of digital palpation of pelvic floor muscle contraction.<sup>23</sup> Moreover, 3D EVUS has been suggested to be comparable in assessing normal and abnormal LAM anatomy and it is arguably a superior method for assessment of the individual muscle portions.<sup>24,25</sup> This is in line with 3D endoanal ultrasound being widely considered the gold standard of diagnostics of the anal sphincter complex. The fact that it is possible to reproduce the LAD scoring system with both high intra- and interrater correlations is promising for the clinical utility of this method to assess pelvic floor deficiencies.

The main strengths of the study are the sample size (which is larger than the study on which it is based), the structured technique of scoring LAD, and two raters blinded to each other's results, which reproduced the high reliability of the LAD scoring system. Raters were also blinded to potential patient symptoms and obstetric history, with the exception of of primiparous women 1 year after vaginal delivery. This warrants assessments not influenced by the raters'

expectations due to preexisting knowledge of maternal or obstetric risk factors of LAD.<sup>26,27</sup> Not only were previous findings confirmed – we have also introduced a higher level of technical stringency by clearly separating evaluations of different probes.

We did not include data on clinical symptoms or examination beyond the ultrasound volumes, as this lay beyond the scope of the current study. Scores could therefore not be correlated to these outcomes and it will be necessary to test the method on a wider range of clinical patients, including a larger proportion of women with prior instrumental delivery and higher risk of severe injuries. Until then, there is a potential issue of generalizability to a more heterogeneous population. The endovaginal approach to 3D ultrasound has previously been shown to have good to excellent correlation with symptomatic pelvic organ prolapse and anatomical findings.<sup>8,13,28–30</sup> This is planned to be investigated in a subsequent study by our research group.

In addition, the study population consisted of a homogeneous group, which implies few potential confounding factors in terms of differences in background factors such as parity and menopausal status, or in anatomical factors such as prolapse status but with only few LAD, which might result in artificially higher correlations.<sup>31</sup> Though it could be argued that, by extension, the results of this study are less generalizable to a clinical population, the purpose of this study was not to study the predictive value of LAD score in detecting pelvic floor injuries (as this has been previously established), but rather to investigate its reproducibility.<sup>8</sup>

## 5 | CONCLUSION

The LAD score is a feasible, repeatable, transferable, consistent and applicable scoring system. LAD score is not a singular method that provides a complete clinical evaluation; rather it should always be complemented by a clinical examination and assessment of the patient. It may be an important tool in the study of pelvic floor dysfunction in general and may enhance clinical evaluation.

## AUTHOR CONTRIBUTIONS

ER: Project development, data collection and management, data analysis, writing of draft. VU: Data management, data analysis, writing of draft. MS: Project development, data collection and management, editing of draft. Gunilla Tegerstedt: Project development, editing of draft. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## ACKNOWLEDGMENTS

The authors would like to thank Professor S. A. Shobeiri, Department of Obstetrics and Gynecology, INOVA Women's Hospital, Falls Church, VA, USA, for kindly contributing instrumental advice and valuable comments regarding the LAD scoring system. The authors would also like to extend their gratitude to Associate Professor G. Ajne and Dr. H. Engberg, Department of Obstetrics and Gynecology, Karolinska University Hospital, Stockholm, Sweden, for critically



revising the article and adding valuable comments. We believe this contributed greatly to improving it.

## FUNDING INFORMATION

This study was partly funded by the Karolinska Institutet Center for Medical Innovation (CIMED). The funding source was not involved in the study.

## CONFLICT OF INTEREST STATEMENT

The authors have stated explicitly that there are no conflicts of interest in connection with this article.

## ORCID

Emilia Rotstein  <https://orcid.org/0000-0002-2529-9390>

## REFERENCES

- Ashton-Miller JA, Delancey JO. On the biomechanics of vaginal birth and common sequelae. *Annu Rev Biomed Eng.* 2009;11:163-176.
- Shek KL, Dietz HP. Intrapartum risk factors for levator trauma. *BJOG.* 2010;117:1485-1492.
- Dietz HP. Pelvic floor trauma in childbirth. *Aust N Z J Obstet Gynaecol.* 2013;53:220-230.
- Dietz HP, Chantarasorn V, Shek KL. Levator avulsion is a risk factor for cystocele recurrence. *Ultrasound Obstet Gynecol.* 2010;36:76-80.
- Chantarasorn V, Shek KL, Dietz HP. Sonographic detection of puborectalis muscle avulsion is not associated with anal incontinence. *Aust N Z J Obstet Gynaecol.* 2011;51:130-135.
- Rodrigo N, Shek KL, Dietz HP. Rectal intussusception is associated with abnormal levator ani muscle structure and morphometry. *Tech Coloproctol.* 2011;15:39-43.
- Huebner M, Margulies RU, DeLancey JO. Pelvic architectural distortion is associated with pelvic organ prolapse. *Int Urogynecol J Pelvic Floor Dysfunct.* 2008;19:863-867.
- Rostaminia G, White D, Hegde A, Quiroz LH, Davila GW, Shobeiri SA. Levator ani deficiency and pelvic organ prolapse severity. *Obstet Gynecol.* 2013;121:1017-1024.
- Siafarikas F, Staer-Jensen J, Braekken IH, Bo K, Engh ME. Learning process for performing and analyzing 3D/4D transperineal ultrasound imaging and interobserver reliability study. *Ultrasound Obstet Gynecol.* 2013;41:312-317.
- Morgan DM, Umek W, Stein T, Hsu Y, Guire K, DeLancey JO. Interrater reliability of assessing levator ani muscle defects with magnetic resonance images. *Int Urogynecol J Pelvic Floor Dysfunct.* 2007;18:773-778.
- Dietz HP, Shek KL, Chantarasorn V, Langer SE. Do women notice the effect of childbirth-related pelvic floor trauma? *Aust N Z J Obstet Gynaecol.* 2012;52:277-281.
- Shobeiri SA, LeClaire E, Nihira MA, Quiroz LH, O'Donoghue D. Appearance of the levator ani muscle subdivisions in endovaginal three-dimensional ultrasonography. *Obstet Gynecol.* 2009;114:66-72.
- Rostaminia G, Manonai J, LeClaire E, et al. Interrater reliability of assessing levator ani deficiency with 360 degrees 3D endovaginal ultrasound. *Int Urogynecol J.* 2014;25:761-766.
- Dietz HP. Clinical consequences of levator trauma. *Ultrasound Obstet Gynecol.* 2012;39:367-371.
- Edqvist M, Hildingsson I, Mollberg M, Lundgren I, Lindgren H. Midwives' management during the second stage of labor in relation to second-degree tears-an experimental study. *Birth.* 2017;44:86-94.
- Dietz HP, Lanzarone V. Levator trauma after vaginal delivery. *Obstet Gynecol.* 2005;106:707-712.
- van Delft KW, Thakar R, Sultan AH, Int'Hout J, Kluivers KB. The natural history of levator avulsion one year following childbirth: a prospective study. *BJOG.* 2015;122:1266-1273.
- Santoro GA, Wiczorek AP, Shobeiri SA, et al. Interobserver and interdisciplinary reproducibility of 3D endovaginal ultrasound assessment of pelvic floor anatomy. *Int Urogynecol J.* 2011;22:53-59.
- Yune Y, Jeong HY, Park DH, Lee JK. Three-dimensional pelvic floor ultrasound assessment of pelvic organ prolapse: minimal levator hiatus and levator ani deficiency score. *Ann Coloproctol.* 2021;37:291-297.
- Vergeldt TF, Weemhoff M, Notten KJ, Kessels AG, Kluivers KB. Comparison of two scoring systems for diagnosing levator ani muscle damage. *Int Urogynecol J.* 2013;24:1501-1506.
- DeLancey JO, Morgan DM, Fenner DE, et al. Comparison of levator ani muscle defects and function in women with and without pelvic organ prolapse. *Obstet Gynecol.* 2007;109:295-302.
- Notten KJB, Vergeldt TFM, van Kuijk SMJ, Weemhoff M, Roovers JWR. Diagnostic accuracy and clinical implications of Translabial ultrasound for the assessment of levator ani defects and levator ani biometry in women with pelvic organ prolapse: a systematic review. *Female Pelvic Med Reconstr Surg.* 2017;23:420-428.
- El-Haieg DO, Madkour NM, Basha MAA, et al. Magnetic resonance imaging and 3-dimensional transperineal ultrasound evaluation of pelvic floor dysfunction in symptomatic women: a prospective comparative study. *Ultrasonography.* 2019;38:355-364.
- Staer-Jensen J, Siafarikas F, Hilde G, Braekken IH, Bo K, Engh ME. Pelvic floor muscle injuries 6 weeks post partum-an intra- and inter-rater study. *Neurourol Urodyn.* 2013;32:993-997.
- Javadian P, O'Leary D, Rostaminia G, et al. How does 3D endovaginal ultrasound compare to magnetic resonance imaging in the evaluation of levator ani anatomy? *Neurourol Urodyn.* 2017;36:409-413.
- Blomquist JL, Munoz A, Carroll M, Handa VL. Association of delivery mode with pelvic floor disorders after childbirth. *Jama.* 2018;320:2438-2447.
- Friedman T, Eslick GD, Dietz HP. Delivery mode and the risk of levator muscle avulsion: a meta-analysis. *Int Urogynecol J.* 2019;30:901-907.
- van Delft K, Shobeiri SA, Thakar R, Schwertner-Tiepelmann N, Sultan AH. Intra- and interobserver reliability of levator ani muscle biometry and avulsion using three-dimensional endovaginal ultrasonography. *Ultrasound Obstet Gynecol.* 2014;43:202-209.
- Rostaminia G, Peck JD, Quiroz LH, Shobeiri SA. How well can levator ani muscle morphology on 3D pelvic floor ultrasound predict the levator ani muscle function? *Int Urogynecol J.* 2015;26:257-262.
- van Delft KW, Sultan AH, Thakar R, Shobeiri SA, Kluivers KB. Agreement between palpation and transperineal and endovaginal ultrasound in the diagnosis of levator ani avulsion. *Int Urogynecol J.* 2015;26:33-39.
- Wu Y, Dabhoiwala NF, Hagoort J, Tan LW, Zhang SX, Lamers WH. Architectural differences in the anterior and middle compartments of the pelvic floor of young-adult and postmenopausal females. *J Anat.* 2017;230:651-663.

**How to cite this article:** Rotstein E, Ullemer V, Starck M, Tegerstedt G. Three-dimensional endovaginal ultrasound assessment using the levator ani deficiency score in primiparas: A replication study. *Acta Obstet Gynecol Scand.* 2023;102:1236-1242. doi:[10.1111/aogs.14633](https://doi.org/10.1111/aogs.14633)