Sultan Abdul H (Orcid ID: 0000-0001-8979-2304)
van Gruting Isabelle Magda Aurora (Orcid ID: 0000-0002-9470-1291)

Diagnostic test accuracy of MRI and pelvic floor ultrasound for diagnosis of levator ani muscle avulsion

I.M.A. van Gruting[1], A. Stankiewicz[2,3], K. van Delft[4], S.K. Doumouchtsis[5,6], J. IntHout[7], A.H. Sultan[1,6], R. Thakar[1,6]

1 Department of Obstetrics and Gynaecology, Croydon University Hospital, London, United Kingdom

2 Department of Radiology, Croydon University Hospital, London, United Kingdom

3 Department of Radiology, University Hospitals of North Midlands NHS Trust, Stoke-on-Trent, United Kingdom

4 Department of Obstetrics and Gynaecology, Jeroen Bosch Ziekenhuis, Den Bosch, the Netherlands.

5 Department of Obstetrics and Gynaecology, Epsom & St Helier University Hospitals, London, United Kingdom

6 Department of Obstetrics and Gynaecology, St George's University of London, London, UK

7 Radboud Institute for Health Sciences, section of Biostatistics, Radboud university medical center, Nijmegen, The Netherlands

Corresponding author:

Ranee Thakar

Croydon University Hospital

530 London Road, Thornton Heath

CR7 7 YE, Surrey, United Kingdom

E-mail address: ranee.thakar@nhs.net

Keywords: Diagnostic test accuracy, sensitivity and specificity, levator ani muscle, magnetic resonance imaging, ultrasound

Short title: DTA of MRI and ultrasound for LAM avulsion

Contribution

What are the novel findings of this work?

Pelvic floor ultrasound has a lower sensitivity, but a higher specificity for the detection of levator ani muscle avulsion compared to MRI. MRI and EVUS have a good predictive ability for major avulsion and correlate to anterior vaginal wall prolapse. Agreement for avulsion and hiatal measurements was highest between MRI and EVUS.

What are the clinical implications of this work?

Pelvic floor ultrasound can be implemented as a triage test to assess parous women for levator ani muscle avulsion but cannot substitute MRI. The predictive ability of ultrasound is moderate for presence and very good for absence of LAM avulsion. A positive test should be re-confirmed by a different observer or imaging technique.

Abstract

Objective: To estimate diagnostic test accuracy of magnetic resonance imaging (MRI) and pelvic floor ultrasound for diagnosing levator ani muscle (LAM) avulsion in a general population, with a view to establish if ultrasound could substitute MRI to diagnose LAM avulsion.

Methods: Cross-sectional study on 135 women four years after their first delivery. Signs and symptoms were assessed using validated methods. All underwent 4D transperineal ultrasound (TPUS), 3D endovaginal ultrasound (EVUS) and MRI. Images were acquired at rest, squeeze and Valsalva and analysed by two blinded observers. Pre-defined cut-off values were used to diagnose LAM avulsion. In the absence of a reference standard, latent class analysis (LCA) was used to establish diagnostic test characteristics for LAM avulsion as the primary outcome measure. Secondary outcomes were kappa agreement between imaging techniques, intra-class correlation (ICC) for hiatal measurements at rest, squeeze and Valsalva, and correlation with signs and symptoms of pelvic floor dysfunction.

Results: Prevalence of LAM avulsion with MRI was 22.9%, TPUS 11.1%, and EVUS 17.8%. Prevalence based on LCA was 15.7%. Sensitivity of TPUS (71%; CI 50-90) and EVUS (91%; CI 74-100) for LAM avulsion is lower than MRI (100%; CI 84-100). Specificity of TPUS (100%; CI 97-100) and EVUS (95%; CI 91-99) is higher than MRI (91%; CI 85-97). MRI (PPV 95% and NPV 100%) and EVUS (PPV 100% and NPV 98%) have high predictive values for assessment of major LAM avulsion and correlate to anterior vaginal wall prolapse. TPUS has high predictive values for minor LAM avulsion (PPV 100% and NPV 95%). Agreement for diagnosis of LAM avulsion (κ 0.69) and hiatal measurements (ICC 0.60-0.81) was highest between MRI and EVUS.

Conclusion: Pelvic floor ultrasound can be implemented as a triage test to assess parous women for LAM avulsion because of its high specificity. Ultrasound cannot substitute MRI because of a lower sensitivity. The predictive ability of ultrasound is moderate for presence and very good for absence of LAM avulsion. A positive test should be re-confirmed by a different observer or imaging technique.

Introduction

During childbirth the levator ani muscle (LAM) stretches up to 3.3 times its initial length to allow passage of the fetal head[1]. This could cause stretch injury to or avulsion of the muscle from the pubic bone. LAM avulsion occurs in 13-36% of women during first vaginal delivery[2][3]. Women with an enlarged hiatal area or LAM avulsion have an increased risk of developing pelvic organ prolapse (POP)[4][5] and recurrence of cystocele after surgical repair[6][7]. Early detection of LAM avulsion could help to select women who are at risk to develop pelvic floor dysfunction (PFD) and prevent symptoms by early implementation of pelvic floor muscle training[8]. Identification of LAM avulsion in women wíth symptoms of PFD could change treatment strategy[9][10]. LAM avulsion is not easily demonstrable on clinical examination, hence additional imaging is required[11][12]. Currently, no imaging method is implemented in clinical practice to assess women for LAM avulsion.

Magnetic resonance imaging (MRI) was the first imaging technique to demonstrate LAM avulsion and considered the reference standard. Ultrasound is less expensive and more widely available, allowing an easier implementation in clinical practice after establishing its diagnostic test accuracy and agreement of measurements. For detection of LAM avulsion, transperineal ultrasound (TPUS) shows reasonable agreement with MRI in women who are postpartum[13] and in those with POP[14][15], and endovaginal ultrasound (EVUS) is comparable to MRI in women with PFD[16]. Agreement of hiatal measurements between TPUS and MRI was assessed in asymptomatic nullipara[17][18][19] and in symptomatic women at rest[20]. Diagnostic test accuracy (DTA) comparing all three imaging techniques has not been established, as well as agreement of measurements at rest, squeeze and Valsalva in parous women.

The primary objective was to establish the DTA of MRI, TPUS and EVUS for diagnosing LAM avulsion in a general population, using latent class analysis in the absence of a reference standard. Secondly, to assess agreement between imaging techniques for LAM avulsion and hiatal measurements at rest, squeeze and Valsalva, and to correlate LAM avulsion to signs and symptoms of PFD. The aim of this study was to establish if ultrasound could substitute MRI to diagnose LAM avulsion.

Methods

Patient recruitment

This is a cross-sectional study on women four years after their first delivery, who were recruited as part of a prospective longitudinal study with the aim to establish the prevalence of LAM avulsion during childbirth[2]. Between January 2011 and May 2012, 269 nulliparous women were recruited antenatally at Croydon University Hospital, Croydon, United Kingdom. All nulliparous women were invited to participate to create a sample representative of the general population. Inclusion criteria were a singleton pregnancy, no previous pregnancies beyond 20 weeks of gestation, age 18 years or above, and being able to read and understand English. The study was approved by the National Research Ethics Service South West London Committee (REC10/H0806/87). All study participants gave written informed consent for the prospective study and separately for the MRI in this cross-sectional study.

Data collection

The appointment four years following first delivery was performed by one investigator (IvG). Patient characteristics and delivery details were collected from patient's history and hospital records. Symptoms were assessed using validated questionnaires. The International Consultation on Incontinence Questionnaire - Short Form (ICIQ-SF) to assess the frequency, amount, type and bothersomeness of urinary incontinence [21]. The International Consultation on Incontinence Questionnaire – Vaginal Symptoms (ICIQ-VS) to assess symptoms of POP and sexual dysfunction[22]; the presence of a lump or bulge inside or outside of the vagina, as well as the interference and bothersomeness of vaginal symptoms in a sexual relationship. The St Mark's score was used for faecal and flatal (anal) incontinence[23].

On clinical evaluation, pelvic floor muscle strength (PFMS) was assessed by digital palpation using the Modified Oxford Score (MOS) on a six-point scale[24] and POP using the validated International Continence Society POP quantification (ICS-POP-Q) staging method[25]. Pelvic floor ultrasound consisting of EVUS and TPUS was performed during the same consultation. The MRI appointment was either before or after the ultrasound appointment and where possible on the same day. Time between MRI and ultrasonography was kept as short as possible.

Image acquisition

All imaging techniques were performed with the women in the supine position with the knees semi-flexed and an empty bladder. MRI was conducted by radiographers who were specifically trained for this study by an experienced radiologist (ASt). Pelvic floor ultrasound was performed by an experienced research fellow (IvG) rigorously trained by the principal investigator (RT).

Magnetic resonance imaging was performed with a closed 1.5 Tesla magnet scanner (Siemens Avanto). The patients were instructed on how to correctly perform pelvic floor muscle contraction and maximum Valsalva. Initially, T2W HASTE sagittal images were acquired at rest for an overview of the pelvic anatomy (Field of View (FoV) 250 mm, slice thickness 4 mm, TR 5180ms, TE 128 ms) and high resolution 3-dimensional T2W sequence SPACE (FoV 256 mm, slice thickness 1 mm, TR 1600 ms, TE 90 ms) was used for more detailed assessment of the LAM. Secondly, T2W TRU FIST fast dynamic sequences were performed as cineloops in the sagittal plane during pelvic floor muscle contraction and maximum Valsalva (FoV 350 mm, slice thickness 5 mm, TR 3.57 ms, TE 1.79 ms). The FoV was corrected according to degree of pelvic floor lifting during contraction and descent during Valsalva. Finally, women were asked to maintain constant squeeze for 16 seconds, when 5 axial images of pelvic floor level and puborectalis muscle were taken (T2W FSE, FoV 256 mm, slice thickness 5 mm, TR 3100ms, TE 128 ms). Following this, women were asked to maintain constant maximal Valsalva for 16 seconds and images were obtained using the same protocol. All images were stored and analysed using Sectra PACS and diagnostic radiological workstations.

Four-dimensional (4D) TPUS was performed using the Voluson I system (GE Medical Systems, Zipf, Austria) with a 4–8 MHz curved array volume transducer (acquisition angle 85 degrees) positioned on the perineum between the mons pubis and the anal margin, with slight pressure and good tissue contact[26]. Images were acquired at rest, maximum pelvic floor contraction and maximum Valsalva. Three cineloops were taken and the best was used for analysis. Offline analysis was carried out using 4D View software (version 10.2; GE Medical Systems).

Three-dimensional (3D) EVUS was performed using the Profocus 2022 ultrasound scanner (BK-Medical, Herlev, Denmark). A high-resolution linear array transducer type 8838 (6–12 MHz, focal range 3–60 mm, contact surface 65 mm) was placed into the vagina providing a 2D image in the mid-sagittal plane. A 3D volume was acquired by 360 degrees rotation of the build-in array in 60 seconds with intervals of 0.2 mm. Images were acquired at rest only, as dynamic studies are currently not possible using 3D EVUS. Offline analysis was performed using the 3D viewer software (version 5.19; BK Medical).

Imaging analysis

On all three imaging techniques the LAM was assessed for avulsion in the plane of the minimal hiatal dimensions. The mid-sagittal plane was used to identify the minimal antero-posterior diameter of the levator hiatus, from the posterior margin of the symphysis pubis to the anterior margin of the LAM where it defines the anorectal angle[5][27][28].

On MRI, LAM attachment to the pubic bone was assessed on the axial images using a previously validated scoring system and confirmed in the coronal plane[29][30]. The left and the right side were scored separately. A score of 0 was assigned if no damage was visible, 1 if < 50% of the muscle bulk was missing, 2 if ≥ 50% of the muscle bulk was missing and 3 if the complete muscle was detached from the pubic bone. The bilateral score, the sum of both sides ranging from 0 to 6, was categorised as follows: 0 = no avulsion or intact muscle; 1–3 = minor avulsion; 4–6 or unilateral 3 = major avulsion.

On 4D TPUS, LAM avulsion was assessed using Tomographic Ultrasound Imaging (TUI) creating a set of eight slices in the axial plane with 2.5 mm slice interval from 5 mm caudal to 12.5 mm cranial of the plane of the minimal hiatal dimensions[31]. The three central slices (presenting the plane of the minimal dimensions as well as the two slices cranial of that plane) were assessed for absent connection between the LAM and inferior pubic ramus using direct visualisation, scoring left and right side separately. The final unilateral score ranged from 0 slices affected (no avulsion), 1 or 2 slices affected (partial avulsion) to 3 slices affected (complete avulsion). A summed total score was assigned for both sides, and classified as no LAM avulsion (summed score 0), minor LAM avulsion (summed score 1–3) or major LAM avulsion (summed score 4–6, or unilateral score 3)[2][32][33][34].

On 3D EVUS, LAM avulsion was assessed in the axial plane, where the pubic bone, the midurethra, the anal canal and the pubococcygeus parts of the LAM were all visualised on the same image[35]. LAM attachment to the pubic bone was assessed using the DeLancey score also described for MRI, which has been used for EVUS[12], and confirmed in the sagittal planes.

Offline imaging analysis was performed by investigators blinded to clinical details, results of the other observer and other imaging techniques. An observer assessed images of different techniques at least a month apart to overcome recollection bias. LAM avulsion on each imaging technique was assessed by two independent observers and discrepancies were resolved by a third observer. A discrepancy meeting was held if there was a difference of at least 2 points on the 7-point scale (bilateral sum score ranging from 0 to 6) or in case there was a difference in category: no, minor avulsion, or major avulsion. MRI was analysed by IvG and ASt (3 and 12 years of experience respectively) and SD as third observer (5 years of experience). TPUS was assessed by IvG and KvD (3 and 6 years of experience respectively) and RT as third observer (10 years of experience). EVUS was analysed by IvG and ASt and discrepancies were resolved by RT. IvG was rigorously trained in assessment of LAM avulsion by the senior authors (AS, ASt, RT).

Hiatal measurements were performed on all three imaging techniques in the axial plane, in the plane of the minimal hiatal dimensions, at rest, squeeze and Valsalva: hiatal transverse diameter, hiatal antero-posterior diameter and hiatal area[35]. Measurements on all three techniques were performed by one observer (IvG).

Reference standard

MRI has previously been considered as reference standard for diagnosis of LAM avulsion. The normal anatomy of the LAM on MRI correlates well with cadaveric sections[36], however LAM avulsion on MRI has not been correlated histologically. A method that could be used in the absence of a reference standard is latent class analysis (LCA). This is a statistical model which aims to identify the true patient's status by combining the results of the three imaging techniques [37]. The actual results of the imaging techniques are assumed to be imperfect observations of the true unobserved status of the patients, i.e. latent classes 'healthy' and 'diseased'. Other methods dealing with the absence of a reference standard are composite reference standard and consensus diagnosis[38]. A composite reference standard could not be used because no a priori consensus exists about what combination of tests would be a suitable reference standard, and consensus diagnosis might be subjective, less reproducible and is labour intensive[39].

Pre-defined outcome measures

It is important to minimise the number of false positives (e.g. high specificity), because costs (financial and discomfort to the patient) of subsequent tests are high in relation to the potential benefit of early treatment. It is less important to minimise the number of false negatives (e.g. high sensitivity), because LAM avulsion is not a life-threatening condition and does not require urgent treatment as it does not progress quickly. Hence for ultrasound to be able to be a triage test it should have a similar specificity compared to MRI and it should have a similar sensitivity and specificity to be able to substitute MRI.

Statistical analysis

Agreement between imaging techniques for diagnosis of LAM avulsion at rest, squeeze and Valsalva was evaluated using weighted kappa, which enables the assessment of agreement between ordinal variables, corrected for chance findings[40]. The weighted difference between no and minor, and between minor and major avulsion were set at 1, and the weighted difference between no and major avulsion was set at 2. A kappa ≤ 0 refers to a less-than-chance agreement, 0.01–0.20 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, 0.81–0.99 almost perfect agreement and 1.00 perfect agreement[41].

The diagnostic test accuracy of MRI, TPUS and EVUS for diagnosis of LAM avulsion was assessed for the methods that are most commonly used in current practice as they seem to best diagnose LAM avulsion; i.e. MRI at rest, TPUS at maximum pelvic floor contraction and EVUS at rest. Because of the absence of a reference standard, diagnostic test characteristics were estimated using LCA. The latent classes were estimated with an expectation maximisation algorithm assuming conditional independence between the diagnostic test results [42], as the images of each of the three imaging techniques were analysed blinded (i.e. the outcomes are not related to each other). Different starting values were used to ensure a global maximum. We assessed model fit by checking the goodness of fit statistics (AIC and BIC), by comparing observed versus expected frequencies, by evaluating the plausibility of the assignments of the individual patterns, and by comparing the median bootstrap results with the point estimates from the original data. BIC and AIC indicated a two-classes model when using the original data. Sensitivity, specificity, area under the curve (AUC), positive and negative predictive values (PPV/NPV), and positive and negative likelihood ratio's (LR+/LR-) with 95% bootstrap confidence intervals were estimated based on the observed three categories (no/minor/major) to calculate results for any (minor/major vs no), minor (minor vs no/major) and major (major vs minor/no) avulsion. Bootstrapping based on 500 runs was used to calculate the 95% confidence intervals. In case confidence intervals ranged from 1 to 1 with a point estimate of 1, the exact binomial distribution (based on the number of women who had an avulsion according to LCA with >50% probability as cut-off) was used to calculate a more appropriate confidence interval. For example, if 10 women had a major avulsion according to LCA, the confidence interval was based on the exact binomial distribution for 10 events out of 10 women. If needed, for the LR+, the lower limits of confidence intervals with value infinity were replaced by the point estimate of the sensitivity divided by the lower limit of the CI for the specificity. Similar for the LR-, upper limits of confidence intervals with value 0 were replaced by (1 - lower limit of CI for sensitivity) divided by the point estimate of the specificity. Test accuracy was defined as follows: AUC 0.9–1.0 excellent, 0.8–0.9 good, 0.7–0.8 moderate, 0.6–0.7 fair, 0.5–0.6 poor, and below 0.5 as a useless test.

Agreement between imaging techniques for LAM biometry measurements at rest, squeeze and Valsalva was assessed using Bland Altman plots[43] to detect systematic bias between the imaging techniques. The mean difference (δ) and standard deviation of the differences (SDd) were calculated. Limits of agreement (LOAs) were constructed as δ ± 1.96 x SDd. Intra-class correlation coefficients (ICCs), using a two-way mixed model with test as fixed effect and rater as random effect, were calculated to measure repeatability between imaging techniques for hiatal measurements (<0.0, no repeatability; 0.0–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; and 0.81–1.00, excellent repeatability).

The association of LAM avulsion to symptoms and signs of PFD was assessed with the Mann-Whitney U non-parametric test when data were not normally distributed and with the independent samples t-test when normally distributed. For the assessment of urinary incontinence, the ICIQ-SF total score was used (range 0-21) and for anal incontinence the St Mark's total score (range 0-24)[21][23]. Symptoms of POP were evaluated using ICIQ-VS questionnaire question 5 and 6 (range 0-28) and for symptoms of sexual dysfunction ICIQ-VS questionnaire questions 11-13 (range 0-36)[22]. POP-Q points Ba, C and Bp were respectively taken for assessment of anterior, middle and posterior compartment prolapse[25].

The initial power calculation was based on the incidence of LAM avulsion after first delivery[2]. For this cross-sectional study on diagnostic test accuracy, no separate power calculation was performed. Statistical analyses were performed using SPSS 26 and R 3.6.2 (package poLCA version 1.4.1 for polytomous variables)[44]. Two-sided P values <.05 were considered statistically significant.

Results

Basic characteristics

Of the 269 women who were recruited antenatally, 147 returned at the four-year follow-up and 135 agreed for an MRI scan. All women who underwent ultrasound and MRI were included in the analysis (Figure 1). The mean age was 34.9 (SD 5.6) years, BMI 26.5 (SD 6.0) kg/m^2 and parity 1.6 (SD 0.6). Ethnicity comprised of 79 (58.5%) Caucasian, 26 (19.3%) Black, 20 (14.8%) Asian, 8 (5.9%) mixed and 2 (1.5%) any other ethnicity. The women were assessed at a mean of 45 months (SD 4.9) after their first delivery; 32 (23.7%) had one or more caesarean sections and 103 (76.3%) had one or more vaginal deliveries. Most women (100/135; 74%) had ultrasound and MRI examinations on the same day. The median time difference between ultrasound and MRI was 0 days (range 0-212 days). No adverse events from performing these imaging techniques were observed during this clinical trial. For MRI, two scans at squeeze and three at Valsalva were not analysable, and for TPUS at Valsalva one scan was not analysable. In the discrepancy meeting, the disagreements in diagnosis between two observers were resolved; 21 MRI scans at rest, 25 at squeeze and 19 at Valsalva were re-assessed by the third observer. For TPUS 22 at rest, 14 at squeeze and 14 at Valsalva, and for EVUS 35 scans were re-assessed. The inter-observer agreement (weighted kappa) for diagnosis of LAM avulsion was moderate (κ 0.46 – 0.62), except for TPUS at Valsalva (κ 0.25) (Supplemental digital content Table 1).

LAM avulsion

The prevalence of minor and major LAM avulsion on each imaging technique at rest, squeeze and Valsalva are presented in Table 1. Prevalence of LAM avulsion assessed at Valsalva was lower than at rest or squeeze. Agreement between imaging techniques for diagnosis of minor and major LAM avulsion at rest, squeeze and Valsalva is presented in Table 2. Agreement between MRI and TPUS assessed at the same condition (e.g. both rest, squeeze and Valsalva) was moderate (κ 0.45-0.47). EVUS at rest correlated best with TPUS squeeze (κ 0.60) and MRI at rest (κ 0.69). Agreement for diagnosis of LAM avulsion on MRI at rest compared to MRI performed at squeeze and Valsalva was moderate (both κ 0.58). Agreement for diagnosis of LAM avulsion on TPUS at rest compared to squeeze and Valsalva was moderate (both κ 0.53).

The diagnostic test accuracy of MRI, TPUS and EVUS for diagnosis of LAM avulsion was assessed for the methods that are most commonly used in current practice as they seem to best diagnose LAM avulsion; i.e. MRI at rest, TPUS at maximum pelvic floor contraction and EVUS at rest (Figure 2). Prevalence of any minor or major LAM avulsion was for MRI 22.9%, TPUS 11.1% and EVUS 17.8%. Diagnostic test characteristics for diagnosis of LAM avulsion of all three techniques based on LCA are presented in Table 3. The prevalence of any minor or major avulsion estimated with LCA was 15.7% (CI 10.1 – 22.1), of which 9.2% (CI 4.4 – 14.1) were major and 6.5% (CI 3.0 – 12.3) minor; corresponding with 21 women with any LAM avulsion; 12 major and 9 minor (by illustration in round numbers). When using LCA as reference standard, MRI correctly diagnosed any minor or major LAM avulsion in all 21 cases and did not underdiagnose. MRI overdiagnosed 9 minor and 1 major avulsions. TPUS correctly diagnosed any minor or major LAM avulsion in 15 out of 21 cases and underdiagnosed 2 minor and 4 major avulsions. TPUS did not overdiagnose any minor or major avulsion, but overdiagnosed 5 major avulsions that were actually minor avulsions. EVUS correctly diagnosed any minor or major LAM avulsion in 19 out of 21 cases and underdiagnosed in 1 minor and 1 major avulsions. EVUS overdiagnosed 5 minor avulsions and no major avulsions.

Overall, the estimated diagnostic test accuracy for any, minor or major LAM avulsion was for MRI AUC 95.7 (CI 92.4 – 98.3), TPUS AUC 85.7 (CI 75.0 – 95.1) and EVUS AUC 92.8 (CI 84.4 – 98.6). The point estimates for sensitivity of TPUS (71%; CI 50-90) and EVUS (91%; CI 74-100) for any LAM avulsion were lower compared to MRI (100%; CI 84-100). The point estimates for specificity of TPUS (100%; CI 97-100) and EVUS (95%: CI 91-99) were higher compared to MRI (91%; CI 85-97). In a general population with a low prevalence of LAM avulsion, when a major LAM avulsion is present on MRI (PPV 95%; CI 69-100) or EVUS (PPV 100%; CI 69-100), the probability of this diagnosis is quite likely, however when a minor LAM avulsion is present on MRI (PPV 49%; CI 23-80) or EVUS (PPV 54%; CI 26-82) this diagnosis remains uncertain. When a minor LAM avulsion is present on TPUS (PPV 100%; CI 15-100) the probability of this diagnosis is likely, however when a major LAM avulsion is visible on TPUS (PPV 62%; CI 38-90) this diagnosis remains uncertain. When no minor or major LAM avulsion are seen on any of the three imaging techniques (NPV 94.8%-100%), the absence of a LAM avulsion is nearly certain.

Hiatal measurements

Results of hiatal measurements performed on each imaging technique at rest, squeeze and Valsalva are provided in Table 4, together with Bland-Altman limits of agreement (LOA) and intra-class correlation coefficient (ICC) for differences in measurements between imaging techniques. Bland-Altman plots are provided in Supplementary digital content Figure 1. In parous women with and without LAM avulsion, the mean hiatal measurements on TPUS were all larger than on MRI and EVUS when performed at rest. At squeeze and Valsalva, TPUS transverse diameter was larger than MRI, however anterior-posterior diameter and hiatal area were smaller than MRI. Differences in hiatal measurements between EVUS and MRI at rest were smallest, with the highest ICC (0.60-0.81).

Signs and symptoms of PFD

LAM avulsions diagnosed on best practice imaging methods were correlated to signs and symptoms of PFD to evaluate if these imaging techniques are able to identify clinically significant avulsions. Results comparing women with minor or major LAM avulsion to women with no LAM avulsion are provided in Table 5. Women with major avulsion on all three imaging techniques did not significantly have more symptoms of POP. Women with a minor avulsion on EVUS (4.8 vs 1.1; p=0.001) and TPUS (10 vs 1.6; p=0.002) had more symptoms of POP than women with no LAM avulsion. Presence of major LAM avulsion did not correlate to symptoms of urinary/faecal incontinence and sexual dysfunction. Pelvic floor muscle strength was reduced in women with either minor or major avulsion diagnosed on all three imaging techniques (1.0-2.4 vs 3.4-3.5; p=0.000 – 0.039). Anterior vaginal wall prolapse was worse in women with major LAM avulsion on MRI (Ba -1.9 vs -2.6; p=0.017) and EVUS (Ba -1.7 vs -2.6; p=0.003), but not on TPUS (-2.2 vs -2.5; p=0.469). Anterior compartment prolapse stage 2 or higher was more often present in women with major LAM avulsion on MRI (3/13 (23%) vs 3/104 (2.9%); p=0.018) and EVUS (3/10 (30%) vs 3/111 (2.7%); p=0.007), but not on TPUS (1/13 (7.7%) vs 6/120 (5%); p=0.522). Middle (POP-Q C) and posterior (POP-Q Bp) compartment prolapse was not significantly different between women with or without LAM avulsion. Women with a major avulsion had a significantly larger hiatal area on EVUS (15.9 vs 11.6; p = 0.005) and MRI at rest (17.3 vs 11.7 cm2; p<0.001), squeeze (15.4 vs 11.1 cm2; p<0.001) and Valsalva (27.5 vs 21.3 cm2; p=0.036), this correlation was not found for minor avulsion. On TPUS, a larger hiatal area in women with major LAM avulsion was found at pelvic floor squeeze (14.3 vs 10.9 cm2; p= 0.008) and Valsalva (24.2 vs 19.8 cm2; p=0.020), but not at rest.

Discussion

Main findings

In a random sample of 135 women four years after their first delivery, the prevalence of any minor or major LAM avulsion on MRI was 22.9%, TPUS 11.1%, EVUS 17.8%. Prevalence estimated with LCA was 15.7%. Sensitivity of pelvic floor ultrasound for LAM avulsion was lower, but specificity was higher compared to MRI. According the pre-established definitions, both TPUS and EVUS can become a triage test, but will not be able to substitute MRI. MRI and EVUS have a good predictive ability for major LAM avulsion and correlate to anterior vaginal wall prolapse. TPUS has a good predictive ability for minor LAM avulsion. Agreement for diagnosis of LAM avulsion and hiatal measurements was highest between MRI and EVUS.

Interpretation and correlation to literature

All three imaging techniques are well able to identify an intact LAM (high specificity and NPV). However, diagnosis of LAM avulsion appeared to be more difficult (moderate sensitivity and PPV). Differences in interpretation of presence of LAM avulsion could be caused by air or movement *artefacts*, however the investigators tried to keep this to a minimum by repeating acquisition/sequences. A small *deviation of the plane of the minimal hiatal dimensions* may place a defect just inside or outside the assessed image, which occurs easily as there is no fixed reference point[13][14]. *Muscle thickness* varies widely among unaffected women. When the LAM appearance is symmetrical, it is hard to distinguish between bilateral partial avulsion and muscle atrophy, as the amount of muscle missing is related to the normal diameter of the contralateral muscle[15][45]. The use of different *classification systems* could affect diagnosis; the TPUS classification measures the defect depth[31][34], whereas the MRI and EVUS classifications measure the width of the muscle damage[12][30]. Although, good agreement was found between them when both were applied on TPUS[15][45]. Finally, image *resolution* could affect interpretation. MRI and EVUS are more likely to overdiagnose minor avulsion and did not overdiagnose major avulsion; in the detailed high resolution images a minor asymmetry could be picked up as minor avulsion and the small difference between nearly complete and complete avulsion is well visible. TPUS is more likely to overdiagnose major avulsion and did not overdiagnose minor avulsion; due to the lower resolution it could be more difficult to determine whether the complete or nearly complete muscle bulk was missing unilaterally and minor asymmetry is less visible.

Notten et al. found a sensitivity of 78% (CI 65-91%) and a specificity of 86% (CI 79-93%) to diagnose major LAM avulsion with TPUS compared to MRI as a reference standard[14]. In a similar sample size we found a lower sensitivity (65%) and higher specificity (96%) when considering none of the tests superior. Avulsions present on TPUS and absent on MRI could also be true positives based on the assumption that MRI is not a gold standard; not considering those as false positives increases specificity of TPUS. PPV increases and NPV decreases with raising prevalence, Notten et al. found a higher PPV (71%) and lower NPV (90%) among women with stage 2 cystocele (prevalence 30%).

Hiatal measurements on TPUS were larger compared to MRI in some[18] and smaller in other reports[17][46], refuting the previously suggested systemic bias caused by differences in slice thickness between TPUS in the rendered format (2 cm) and MRI (1 mm). Interestingly, limits of agreement between TPUS and MRI in our study were similar or wider than in asymptomatic nulliparous women[17][18][19], but smaller than in women with symptoms of POP [20][46]. This suggests that agreement could depend on the degree of pelvic floor damage rather than the imaging method itself. Similar to others, hiatal area at rest showed higher agreement than hiatal area at Valsalva, explained by difficulties in obtaining equivalent Valsalva[17][46].

Major LAM avulsion correlated to signs and not to symptoms of PFD, most probably because it takes more than four years for the larger hiatal area and weaker pelvic muscles to become symptomatic [47]. Contrary to MRI and EVUS, major LAM avulsion on TPUS did not correlate to anterior vaginal wall prolapse and hiatal area at rest. This might be because TPUS overdiagnosed five and underdiagnosed four major avulsions; excluding diseased and including healthy women could cause insignificant findings. Contrary to EVUS and TPUS, minor LAM avulsion on MRI did not correlate to POP symptoms, which could be explained by MRI overdiagnosing nine minor avulsions. Zhuang et al found higher number of major avulsions on TPUS and higher number of minor avulsions on MRI[15], what they attributed to LAM avulsions being better visualised at contraction[48], however we did not find improved correlation between TPUS and MRI at maximum squeeze. These numbers are better explained by the tendency of TPUS to overdiagnose major and MRI to overdiagnose minor avulsions. Some studies claim minor LAM avulsion not to be of clinal relevance[30][34], however since women with minor LAM avulsion show symptoms of POP and reduced PFMS, this diagnosis should not be ignored.

Implication for practice

Pelvic floor ultrasound can be a triage test for diagnosis of LAM avulsion because of its high specificity, however does not have the same discriminative ability as MRI (ability of a test to recognise diseased and not-diseased patients, e.g. high sensitivity and specificity). As MRI is more expensive, time consuming and has contraindications to its use, the non-invasive ultrasound can be implemented in clinical practice to assess parous women for LAM avulsion. The predictive ability (ability of a test to predict the probability of disease or non-disease in an individual once the test result is known) of ultrasound is moderate for presence (PPV 54-100%) and high for absence of LAM avulsion (NPV 95-99%). When in doubt of presence of an avulsion, the women could be assessed by a different observer or imaging technique. Both ultrasound and MRI require a substantial learning curve, training and consolidation of expertise. Future research should focus on methods to increase agreement between observers and imaging techniques, for instance by developing an easier to apply classification system that is suitable for all three imaging methods that assesses the LAM in the same way.

Strength and Limitations

A prospectively collected random sample was enrolled, thresholds for test positivity were pre-specified and images were analysed blinded, although observers could not be blinded for presence of prolapse on imaging. Test accuracy for best practice methods was established using LCA rather than an imperfect reference standard. The application of LCA might be less stable considering the small number of patients, number of avulsions and number of tests; however this study is the first to compare the three available imaging techniques, is performed in a large sample size considering the number of MRI scans, and was designed to establish DTA in the general population, in which the prevalence is low, and where it may be more difficult to discriminate between cases with and without avulsion. For the first time high resolution 3D SPACE sequences were used at rest, allowing assessment in any desirable plane. This is unlike traditional T2W MRI sequences performed in axial, coronal and sagittal planes separately. MRI images at squeeze and Valsalva were acquired in a pre-planned plane; occasionally the LAM attachment zone was not visible because no tilting was possible. The limitation of TPUS is the low resolution caused by the convex abdominal probe with a low frequency and relatively far from the area of interest. The 3D EVUS cube can be precisely tilted in every plane, however a limitation is that EVUS could only be performed at rest and not at pelvic floor squeeze and Valsalva.

Conclusion

Pelvic floor ultrasound can be implemented as a triage test to assess parous women for LAM avulsion, because of its high specificity. Ultrasound cannot substitute MRI because of a lower sensitivity. The predictive ability of ultrasound is moderate for presence and very good for absence of LAM avulsion. A positive test should be re-confirmed by a different observer or imaging technique.

Ethics Approval:

This study was approved by the National Research Ethics Service South West London committee (REC 10/H0806/87) on 17 November 2010. *Clinical trial registration:* NCT01310660.

Funding: IMA van Gruting was funded by the Mayday Childbirth Charity Fund

References

1. Lien KC, Mooney B, DeLancey JO, Ashton-Miller JA. Levator ani muscle stretch induced by simulated vaginal birth. Obstet Gynecol. 2004 Jan;103(1):31-40.

2. van Delft K, Thakar R, Sultan AH, Schwertner-Tiepelmann N, Kluivers K. Levator ani muscle avulsion during childbirth: a risk prediction model. BJOG. 2014 Aug;121(9):1155-63; discussion 1163.

3 Schwertner-Tiepelmann N, Thakar R, Sultan AH and Tunn R. Obstetric levator ani muscle injuries: current status. Ultrasound Obstet Gynecol 2012; 39: 372–383

4. Dietz HP, Simpson JM. Levator trauma is associated with pelvic organ prolapse. BJOG 2008;115:979–984.

5. Dietz HP, Shek C, Clarke B. Biometry of the pubovisceral muscle and levator hiatus by three dimensional pelvic floor ultrasound. Ultrasound Obstet Gynecol 2005; 25: 580-585

6. Dietz HP, Chantarasorn V, Shek KL. Levator avulsion is a risk factor for cystocele recurrence. Ultrasound Obstet Gynecol 2010;36:76–80.

7. Vergeldt TF, Notten KJ, Weemhoff M, van Kuijk SM, Mulder FE, Beets-Tan RG, Vliegen RF, Gondrie ET, Bergmans MG, Roovers JP, Kluivers KB. Levator hiatal area as a risk factor for cystocele recurrence after surgery: a prospective study. BJOG 2015 Jul;122(8):1130-7.

8. Boyle R, Hay-Smith EJ, Cody JD, Mørkved S. Pelvic floor muscle training for prevention and treatment of urinary and faecal incontinence in antenatal and postnatal women. Cochrane Database of Systematic Reviews 2012, Issue 10.

9. Cheung RYK, Lee JHS, Lee LL, Chung TKH, Chan SSC. Levator ani muscle avulsion is a risk factor for expulsion within 1 year of vaginal pessary placed for pelvic organ prolapse. Ultrasound Obstet Gynecol. 2017 Dec;50(6):776-780.

10. Wong NKL, Cheung RYK, Lee LL, Wan OYK, Choy KW, Chan SSC. Women with advanced pelvic organ prolapse and levator ani muscle avulsion would significantly benefit from mesh repair surgery. Ultrasound Obstet Gynecol. 2021 Apr;57(4):631-638.

11. Kearney R, Miller JM, Delancey JO. Interrater reliability and physical examination of the pubovisceral portion of the levator ani muscle, validity comparisons using MR imaging. Neurourol Urodyn. 2006;25(1):50-4.

12. 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 Jan;26(1):33-9.

13. Calderwood CS, Thurmond A, Holland A, Osmundsen B, Gregory WT. Comparing 3-Dimensional Ultrasound to 3-Dimensional Magnetic Resonance Imaging in the Detection of Levator Ani Defects. Female Pelvic Med Reconstr Surg. 2018 Jul/Aug;24(4):295-300.

14. Notten KJ, Kluivers KB, Fütterer JJ, Schweitzer KJ, Stoker J, Mulder FE, Beets-Tan RG, Vliegen RFA, Bossuyt PM, Kruitwagen RFPM, Roovers JPRW, Weemhoff M. Translabial three-dimensional ultrasonography compared with magnetic resonance imaging in detecting levator ani defects. Obstetrics and Gynecology 2014;124(6):1190–7.

15. Zhuang RR, Song YF, Chen ZQ, Ma M, Huang HJ, Chen JH, Li YM. Levator avulsion using a tomographic ultrasound and magnetic resonance-based model. Am J Obstet Gynecol 2011;205:232.e1–232.e8.

16. Javadian P, O'Leary D, Rostaminia G, North J, Wagner J, Quiroz LH, Shobeiri SA. How does 3D endovaginal ultrasound compare to magnetic resonance imaging in the evaluation of levator ani anatomy? Neurourol Urodyn. 2017 Feb;36(2):409-413.

17. Kruger JA, Heap SW, Murphy BA, Dietz HP. Pelvic floor function in nulliparous women using three-dimensional ultrasound and magnetic resonance imaging. Obstet Gynecol. 2008 Mar;111(3):631-8.

18. Nardos R, Thurmond A, Holland A, Gregory WT. Pelvic floor levator hiatus measurements: MRI versus ultrasound. Female Pelvic Med Reconstr Surg. 2014 Jul-Aug;20(4):216-21.

19. Majida M, Braekken IH, Bo K, Saltyte Benth J, Engh ME. Validation of three-dimensional perineal ultrasound and magnetic resonance imaging measurements of the pubovisceral muscle at rest. Ultrasound Obstet Gynecol 2010;35:715–722.

20. Vergeldt TF, Notten KJ, Stoker J, Fütterer JJ, Beets-Tan RG, Vliegen RFA, Schweitzer KJ, Mulder FE, van Kuijk SMJ, Roovers JPRW, Kluivers KB, Weemhoff M. Comparison of translabial 3D ultrasound with MRI for measuring levator hiatus biometry at rest. Ultrasound Obstet Gynecol 2016; 47:636–641.

21. Avery K, Donovan J, Peters TJ, Shaw C, Gotoh M, Abrams P. ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. Neurourol Urodyn 2004;23:322–30.

22. Price N, Jackson SR, Avery K, Brookes ST, Abrams P. Development and psychometric evaluation of the ICIQ Vaginal Symptoms Questionnaire: the ICIQ-VS. BJOG 2006;113:700–12.

23. Vaizey CJ, Carapeti E, Cahill JA, Kamm MA. Prospective comparison of faecal incontinence grading systems. Gut 1999;44:77–80.

24. Laycock J. Clinical evaluation of the pelvic floor. In *Pelvic floor re-education, Principles and practice*, Schüssler B, Laycock J, Norton PA, Stanton SL, eds. Springer-Verlag: London, 1994; 42–48

25. Hall AF, Theofrastous JP, Cundiff GW, Harris RL, Hamilton LF, Swift SE, Bump RC. 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 1996;175:1467–70.

26. Dietz HP. Pelvic Floor ultrasound: a review. Am J Obstet Gynecol 2010; 202: 321-334

27. Shobeiri SA, Rostaminia G, White D, Quiroz LH (2013) The determinants of minimal levator hiatus and their relationship to the puborectalis muscle and the levator plate. BJOG 120:205–211

29. Lammers K, Fütterer JJ, Inthout J, Prokop M, Vierhout ME, Kluivers KB. Correlating signs and symptoms with pubovisceral muscle avulsions on magnetic resonance imaging. Am J Obstet Gynecol. 2013 Feb;208(2):148.e1-7.

29. DeLancey JO, Kearney R, Chou Q, Speights S, Binno S. The appearance of levator ani muscle abnormalities in magnetic resonance images after vaginal delivery. Obstet Gynecol. 2003 Jan;101(1):46-53.

30. DeLancey JO, Morgan DM, Fenner DE, Kearney R, Guire K, Miller JM, Hussain H, Umek W, Hsu Y, Ashton-Miller JA. Comparison of levator ani muscle defects and function in women with and without pelvic organ prolapse. Obstet Gynecol 2007; 109: 295–302.

31. Dietz HP. Quantification of major morphological abnormalities of the levator ani. Ultrasound Obstet Gynecol. 2007 Mar;29(3):329-34.

32. van Delft K, Sultan AH, Thakar R, Schwertner-Tiepelmann N, Kluivers K. The relationship between postpartum levator ani muscle avulsion and signs and symptoms of pelvic floor dysfunction. BJOG. 2014 Aug;121(9):1164-71; discussion 1172.

33. van Delft KW, Thakar R, Sultan AH, IntHout J, Kluivers KB. The natural history of levator avulsion one year following childbirth: a prospective study. BJOG. 2015 Aug;122(9):1266-73.

34. Dietz HP, Bernardo MJ, Kirby A, Shek KL. Minimal criteria for the diagnosis of avulsion of the puborectalis muscle by tomographic ultrasound. Int Urogynecol J 2011;22:699–704.

35. 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 Feb;43(2):202-9.

36. Strohbehn K, Ellis JH, Strohbehn JA, DeLancey JO. Magnetic resonance imaging of the levator ani with anatomic correlations. Obstetrics and Gynecology 1996;87: 277–8540.

37. van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KG, de Groot JA. Latent class models in diagnostic studies when there is no reference standard—a systematic review. Am J Epidemiol 2014;179:423–31.

38. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health TechnolAssess. 2007;11(50):iii, ix-51.

39. Bertens LC, Broekhuizen BD, Naaktgeboren CA, Rutten FH, Hoes AW, van Mourik Y, Moons KG, Reitsma JB. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. PLoS Med. 2013;10(10):e1001531.

40. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin. 1968 Oct;70(4):213.

41. Altman DG (1991) Practical statistics for medical research. Chapman and Hall, London

42. Pepe MS, Janes H. Insights into latent class analysis of diagnostic test performance. Biostatistics. 2007 Apr 1;8(2):474-84.

43. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol 2003; 22: 85–93.

44. Linzer, D. A., & Lewis, J. B. poLCA: An R package for polytomous variable latent class analysis. Journal of statistical software. 2011;42(10): 1-29.

45. Vergeldt TFM, Weemhoff M, Notten KJB, Kessels AGH, Kluivers KB. Comparison of two scoring systems for diagnosing levator ani muscle damage. Int Urogynecol J. 2013 Sep;24(9):1501-6.

46. Barry C, Dietz HP, Bhuta S, Lim YN, Greenland H, Reinhold M. MRI versus 3-dimensional ultrasound: a comparative study of levator hiatal dimensions in women with pelvic organ prolapse. Australian and New Zealand Continence Journal. 2006;12:61-65

47. van Gruting IMA, van Delft KWM, Sultan AH, Thakar R. Natural history of levator ani muscle avulsion 4 years following childbirth. Ultrasound Obstet Gynecol. 2021 Aug;58(2):309-317.

48. Dietz HP, Steensma AB. The prevalence of major abnormalities of the levator ani in

Figure 1 Recruitment flow chart

Figure 2a Minor LAM avulsion (arrow) in the same patient on all three imaging techniques in the plane of the minimal hiatal dimensions a) magnetic resonance imaging (rest) b) endovaginal ultrasound (rest) c) transperineal ultrasound (TUI squeeze)

Figure 2b Major LAM avulsion (arrow) in the same patient on all three imaging techniques in the plane of the minimal hiatal dimensions a) magnetic resonance imaging (rest) b) endovaginal ultrasound (rest) c) transperineal ultrasound (TUI squeeze)

Figure S1 BA plots of LAM biometry in 135 women with and without LAM avulsion

Table 1 Prevalence of LAM avulsion in 135 primiparous women on three imaging techniques

Imaging technique	*Condition*	*LAM Intact*	*Minor avulsion*	*Major avulsion*	*Any avulsion*
MRI	rest	104 (77.1)	18 (13.3)	13 (9.6)	31 (22.9)
	squeeze	103 (77.5)	19 (14.3)	11 (8.3)	30 (22.5)
	Valsalva	111 (84.1)	15 (11.4)	6 (4.5)	21 (15.9)
TPUS	rest	113 (83.7)	7 (5.2)	15 (11.1)	22 (16.3)
	squeeze	120 (88.9)	2 (1.5)	13 (9.6)	15 (11.1)
	Valsalva	123 (91.8)	3 (2.2)	8 (6.0)	11 (8.2)
EVUS	rest	111 (82.2)	14 (10.4)	10 (7.4)	24 (17.8)

All results presented as n (%)

LAM = Levator ani muscle

MRI = Magnetic Resonance Imaging

TPUS = Transperineal Ultrasound (using tomographic ultrasound imaging)

EVUS = Endovaginal Ultrasound

Table 2 Agreement between imaging techniques for diagnosis of minor and major LAM avulsions

Imaging technique	*MRI rest*	*MRI squeeze*	*MRI Valsalva*	*EVUS rest*
TPUS rest	0.456	0.390	0.436	0.524
TPUS squeeze	0.585	0.451	0.531	0.598
TPUS Valsalva	0.379	0.366	0.473	0.478
EVUS rest	0.693	0.504	0.562	n/a

All results presented in weighted kappa for ordinal values

MRI = Magnetic Resonance Imaging

TPUS = Transperineal Ultrasound (using tomographic ultrasound imaging (TUI))

EVUS = Endovaginal Ultrasound

Table 3 Diagnostic test characteristics for diagnosis of minor and major LAM avulsion for MRI, TPUS and EVUS using Latent Class Analysis

Imaging Technique	**Prevalence (95% CI)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**AUC (95% CI)**	**PPV (95% CI)**	**NPV (95% CI)**	**LR + (95% CI)**	**LR – (95% CI)**
Any avulsion								
MRI	15.7 (10.1 – 22.1)	100 (83.9 - 100)	91.4 (85.2 - 96.6)	95.7 (92.4 - 98.3)	68.4 (51.4 - 87.8)	100 (96.5 - 100)	11.6 (6.7 - 29.1)	0.0 (0.0 - 0.2)
TPUS		71.4 (50.0 - 90.2)	100 (96.8 - 100)	85.7 (75.0 - 95.1)	100 (95.5 - 100)	94.8 (90.3 - 98.7)	Inf (22.4 - Inf)	0.3 (0.1 - 0.5)
EVUS		90.5 (73.7 - 100)	95.1 (90.6 - 99.0)	92.8 (84.4 - 98.6)	79.9 (61.0 - 96.3)	97.5 (93.7 - 100)	18.6 (10.0 - 92.8)	0.1 (0.0 - 0.3)
Major avulsion								
MRI	9.2 (4.4 – 14.1)	100 (73.5 - 100)	99.2 (96.8 - 100)	99.6 (98.0 - 100)	95.3 (68.7 - 100)	99.7 (99.3 - 100)	200.9 (30.7 - Inf)	0.0 (0.0 - 0.3)
TPUS		64.6 (36.4 - 100)	95.9 (92.1 - 99.2)	80.3 (66.0 - 97.7)	61.5 (37.5 - 89.5)	96.4 (93.0 - 100)	15.8 (6.8 - 95.5)	0.4 (0.0 - 0.7)
EVUS		80.7 (59.9 - 100)	100 (97.0 - 100)	90.4 (79.9 - 100)	100 (69.2 - 100)	98.1 (95.3 - 100)	Inf (27.3 - Inf)	0.2 (0.0 - 0.4)
Minor avulsion								
MRI	6.5 (3.0 – 12.3)	100 (71.0 - 100)	92.9 (87.3 - 96.9)	96.4 (81.7 - 98.4)	49.0 (22.8 - 79.7)	100 (96.9 - 100)	13.7 (7.4 - 35.8)	0.0 (0.0 - 0.3)

TPUS		22.7 (0.0 - 56.2)	100 (97.1 - 100)	61.3 (50.0 - 78.1)	100 (15.8 - 100)	94.9 (89.4 - 98.2)	Inf (7.8 - Inf)	0.8 (0.4 - 1.0)
EVUS		85.7 (50.0 - 100)	95.2 (90.8 - 98.4)	90.5 (71.8 - 98.8)	54.0 (26.2 - 82.1)	99.3 (96.1 - 100)	16.8 (8.1 - 57.6)	0.2 (0.0 - 0.5)

Magnetic Resonance Imaging (MRI) was performed at rest, Transperineal Ultrasound (TPUS) at pelvic floor squeeze and Endovaginal Ultrasound (EVUS) at rest. AUC = Area under the curve; PPV = Positive predictive value; NPV = Negative predictive value; LR+ = Likelihood ratio positive; LR- = Likelihood ratio negative; CI = confidence interval; inf = infinity

Table 4 Agreement of measurements between imaging techniques in 135 primiparous women with and without LAM avulsion

Parameter	Imaging A Mean (SD)	Imaging B Mean (SD)	Mean difference (∂) (95% CI)	SDd	Lower LOA	Upper LOA	ICC# (95% CI)
MRI (A) vs TPUS (B) rest							
Hiatus TV *(cm)*	3.25 (0.56)	3.92 (0.57)	-0.66 (-0.75 to -0.58)	0.49	-1.63	0.30	0.36 (-0.09 to 0.66)
Hiatus AP *(cm)*	5.10 (0.72)	5.24 (0.75)	-0.14 (-0.23 to -0.05)	0.52	-1.15	0.87	0.74 (0.64 to 0.81)
Hiatus Area *(cm2)*	12.3 (3.4)	14.4 (3.4)	-2.11 (-2.54 to -1.67)	2.55	-7.11	2.89	0.60 (0.17 to 0.79)
MRI (A) vs EVUS (B) rest							
Hiatus TV *(cm)*	3.25 (0.56)	3.29 (0.56)	-0.04 (-0.12 to 0.05)	0.49	-1.00	0.93	0.60 (0.49 to 0.70)
Hiatus AP *(cm)*	5.10 (0.72)	4.89 (0.68)	0.21 (0.12 to 0.30)	0.53	-0.82	1.24	0.69 (0.55 to 0.79)
Hiatus Area *(cm2)*	12.3 (3.4)	12.0 (3.1)	0.26 (-0.08 to 0.60)	1.99	-3.64	4.16	0.81 (0.74 to 0.86)
TPUS (A) vs EVUS (B) rest							
Hiatus TV *(cm)*	3.92 (0.57)	3.29 (0.56)	0.63 (0.54 to 0.71)	0.50	-0.35	1.60	0.37 (-0.08 to 0.66)
Hiatus AP *(cm)*	5.24 (0.75)	4.89 (0.68)	0.35 (0.28 to 0.42)	0.40	-0.43	1.12	0.76 (0.30 to 0.89)
Hiatus Area *(cm2)*	14.4 (3.4)	12.0 (3.1)	2.37 (2.01 to 2.72)	2.07	-1.69	6.43	0.63 (0.00 to 0.84)
MRI (A) vs TPUS (B) squeeze							

Hiatus TV *(cm)*	3.43 (0.68)	3.58 (0.58)	-0.15 (-0.24 to -0.05)	0.55	-1.22	0.93	0.61 (0.49 to 0.71)
Hiatus AP *(cm)*	4.75 (0.69)	4.21 (0.64)	0.54 (0.45 to 0.63)	0.54	-0.52	1.60	0.50 (0.01 to 0.75)
Hiatus Area *(cm2)*	11.7 (3.3)	11.3 (3.1)	0.45 (0.07 to 0.84)	2.23	-3.92	4.82	0.75 (0.66 to 0.81)
MRI (A) vs TPUS (B) valsalva							
Hiatus TV *(cm)*	4.28 (0.99)	4.42 (0.77)	-0.15 (-0.31 to -0.01)	0.92	-1.95	1.65	0.46 (0.31 to 0.58)
Hiatus AP *(cm)*	6.35 (1.27)	5.91 (1.06)	0.44 (0.24 to 0.64)	1.17	-1.85	2.73	0.47 (0.30 to 0.60)
Hiatus Area *(cm2)*	21.7 (8.5)	20.4 (6.7)	1.35 (0.04 to 2.65)	7.61	-13.57	16.27	0.50 (0.36 to 0.62)

SD, standard deviation; ICC, intraclass correlation coefficient; CI, confidence interval; ∂, mean difference between measurements; SDd, standard deviation of difference between measurements; LOA, limits of agreement; TV, transverse diameter; AP, antero-posterior diameter. The mean is calculated as measurement A + B / 2 and the difference is calculated as measurement A – B. Limits of agreement are calculated as mean difference (δ) ± 1.96 x SDd.

Table 5 LAM avulsion on MRI, EVUS and TPUS correlated to signs and symptoms of PFD

Mean (SD)	**MRI (rest)**			**EVUS (rest)**			**TPUS (squeeze)**		
	Major n=13	Minor n=18	None n=104	Major n=10	Minor n=14	None n=111	Major n=13	Minor n=2	None n=120
Symptoms of PFD									
Urinary Incontinence (range 0-21)	4.2 (4.9)	4.7 (5.1)	3.6 (4.4)	4.8 (5.2)	5.4 (6.2)	3.5 (4.3)	4.6 (4.7)	6.5 (0.7)	3.7 (4.6)
POP symptoms (range 0-28)	4.9 (8.9)	3.1 (5.5)	1.2 (3.7)	5.9 (10.0)	4.8 * (6.9)	1.1 (3.3)	2.9 (7.4)	10 * (2.8)	1.6 (4.3)
Anal incontinence (range 0-24)	0.8 (2.0)	0.5 (1.3)	0.3 (1.1)	1.0 (2.3)	0.9* (1.7)	0.2 (1.0)	1.0 (2.1)	2.5 (3.5)	0.2 (1.0)
Sexual dysfunction (range 0-36)	7.4 (9.3)	4.7 (9.6)	5.4 (9.5)	9.0 (10.3)	6.7 (13.8)	5.1 (8.7)	4.6 (8.3)	0.0 (n/a)	5.7 (9.6)
Signs of PFD on clinical examination									
Modified Oxford score (range 0-5)	2.2 (1.4)*	2.4 (1.4)*	3.5 (1.3)	2.0 (1.2)*	2.3 (1.8)*	3.5 (1.3)	1.8 (1.2)*	1.0 (1.4)*	3.4 (1.4)
POPQ Ba in cm (range -3 to +10)	-1.9 (1.1)*	-2.4 (0.8)	-2.6 (0.6)	-1.7 (1.2)*	-2.3 (0.8)	-2.6 (0.6)	-2.2 (1.1)	-2.5 (0.7)	-2.5 (0.6)
POPQ Bp in cm (range -3 to +10)	-3.0 (0.0)	-2.8 (0.4)	-2.8 (0.7)	-3.0 (0.0)	-2.9 (0.4)	-2.8 (0.7)	-3.0 (0.0)	-2.5 (0.7)	-2.8 (0.7)
POPQ C in cm (range -10 to +10)	-6.6 (1.4)	-7.2 (1.9)	-7.3 (1.4)	-6.5 (1.6)	-6.9 (1.7)	-7.3 (1.4)	-6.9 (1.3)	-7.0 (1.4)	-7.2 (1.5)
Signs of PFD on imaging									
Hiatal area in cm^2 - *rest*	17.3 (5.1)†	12.0 (3.2)	11.7 (2.5)	16.2 (5.0)†	12.5 (2.4)	11.6 (2.6)	16.0 (3.6)	18.8 (0.8)	14.1 (3.2)

Hiatal area in cm^2 - *squeeze*	15.4 (6.2)†	12.3 (3.4)	11.1 (2.3)	N/A	N/A	N/A	14.3 (5.3)†	14.2 (2.6)	10.9 (2.5)
Hiatal area in cm^2 - *Valsalva*	27.5 (12.0)†	20.2 (7.5)	21.3 (7.9)	N/A	N/A	N/A	24.2† (8.3)	33.1 (15.2)	19.8 (6.1)

* = statistically significant ($p < 0.05$) compared to women with no LAM avulsion using Mann-WU test

† = statistically significant ($p < 0.05$) compared to women with no LAM avulsion using independent T-test

MRI = Magnetic Resonance Imaging

EVUS= Endovaginal ultrasound

TPUS= Transperineal ultrasound

POP= Pelvic organ prolapse

Figure 1 Recruitment flow chart

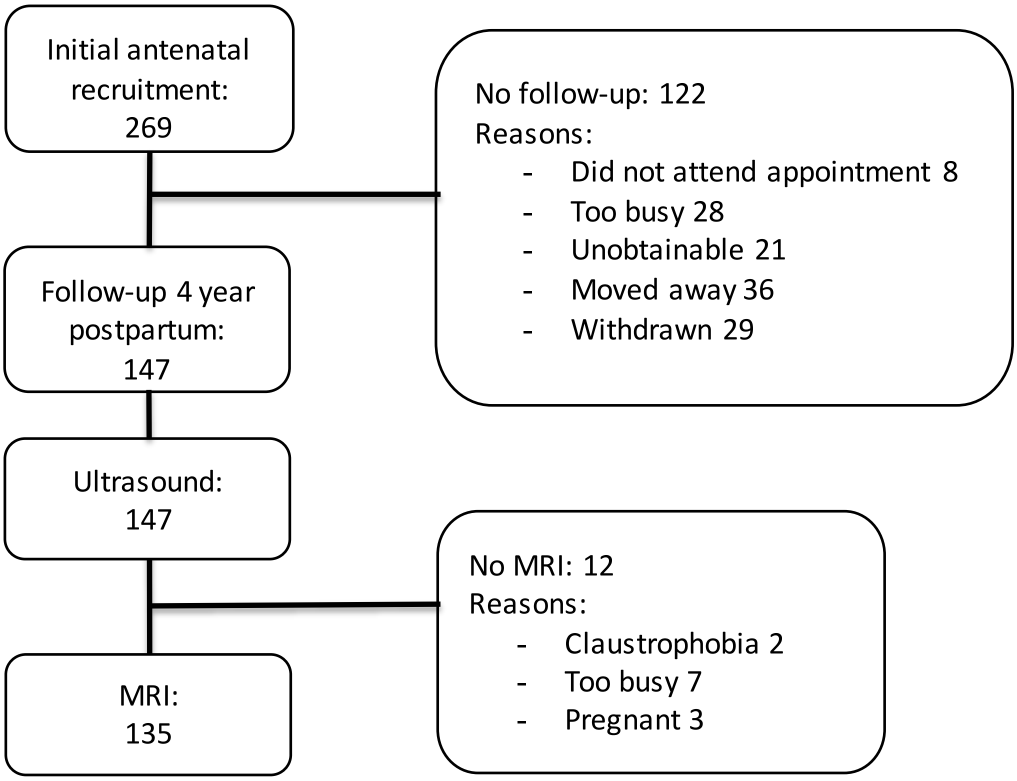

Figure 2a Minor LAM avulsion (arrow) in the same patient on all three imaging techniques in the plane of the minimal hiatal dimensions a) magnetic resonance imaging (rest) b) endovaginal ultrasound (rest) c) transperineal ultrasound (TUI squeeze)

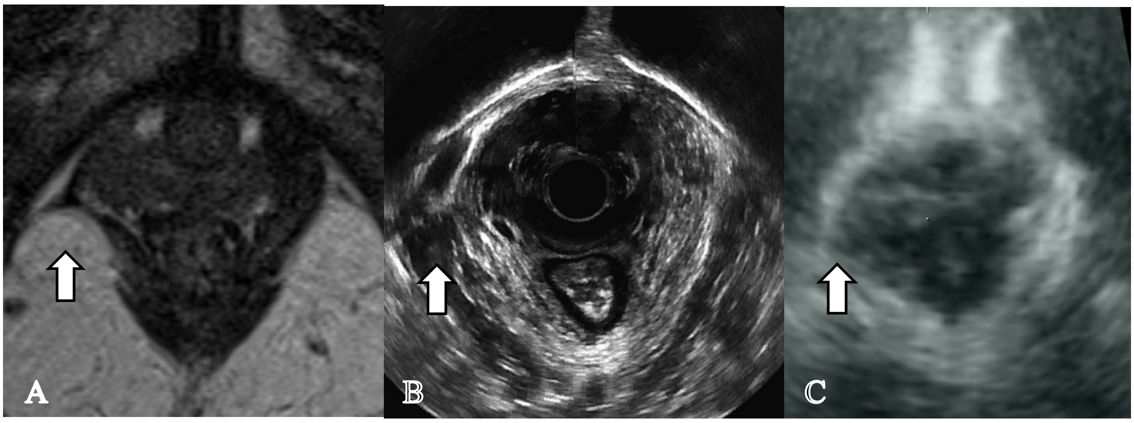

Figure 2b Major LAM avulsion (arrow) in the same patient on all three imaging techniques in the plane of the minimal hiatal dimensions a) magnetic resonance imaging (rest) b) endovaginal ultrasound (rest) c) transperineal ultrasound (TUI squeeze)