Journal Pre-proof

Noninvasive Diagnosis of Adenomyosis: A Structured Review and Meta-Analysis of Diagnostic Accuracy in Imaging

Tina Tellum M.D. , Staale Nygaard Ph.D. , Marit Lieng M.D., Ph.D.

PII: S1553-4650(19)31287-7
DOI: https://doi.org/10.1016/j.jmig.2019.11.001
Reference: JMIG 3997

To appear in: *The Journal of Minimally Invasive Gynecology*

Received date: 26 June 2019
Revised date: 6 October 2019
Accepted date: 4 November 2019

Please cite this article as: Tina Tellum M.D. , Staale Nygaard Ph.D. , Marit Lieng M.D., Ph.D. , Noninvasive Diagnosis of Adenomyosis: A Structured Review and Meta-Analysis of Diagnostic Accuracy in Imaging, *The Journal of Minimally Invasive Gynecology* (2019), doi: https://doi.org/10.1016/j.jmig.2019.11.001

Systematic Review

Noninvasive Diagnosis of Adenomyosis: A Structured Review and Meta-Analysis of Diagnostic Accuracy in Imaging

Tina Tellum, M.D. [a], Staale Nygaard, Ph.D. [b], Marit Lieng, M.D., Ph.D. [a,c]

[a] Department of Gynecology, Oslo University Hospital, Oslo, Norway
[d] Department of Informatics, The Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway
[c] Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Norway

Corresponding author:
Tina Tellum, MD, Department of Gynecology, Oslo University Hospital, Ullevål, PB 4950 Nydalen, 0424 Oslo; Norway. E-mail: tina.tellum@gmail.com; Mobile phone: +47-97793704

Disclosure statement:
All authors are also authors of two of the studies that were included in the meta-analysis.

Funding information:
The authors did not receive any funding for their work related to this article.

Prior presentation:
This work was not published or presented before.

Institutional Review Board approval:
Institutional review board approval was not necessary for this study.

Trial registration:
PROSPERO database (protocol number CRD42019125405)

Word count: 2963

Abstract

Objective: To conduct a systematic review and meta-analysis to evaluate imaging methods used to diagnose adenomyosis.

Data Sources: A thourough search was completed through the Cochrane Central Register of Controlled Trials, EMBASE, and PubMed/MEDLINE databases from January 2000 to June 2019.

Methods of Study Selection: Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool, studies reporting the diagnostic accuracy of an imaging method and histopathology as a reference standard were selected and prospecticvely registered in the International Prospective Register of Systematic Reviews (PROSPERO). Statistical analysis was performed using the R-package Meta-Analysis of Diagnostic Accuracy (mada).

Tabulation, Integration, and Results: Of 1,168 records identified, 10 studies were included in the meta-analysis and included 827 patients undergoing 2- or 3-dimensional (2D, 3D) transvaginal ultrasound (TVUS) and 317 patients undergoing magnetic resonance imaging (MRI). The risk of bias was “low” or “unclear” for all QUADAS domains. The overall prevalence of adenomyosis was 40%. Overall MRI, 2D-TVUS, 3D-TVUS, and TVUS had the following aggregated diagnostic qualities (95% confidence interval): sensitivity 78% (70%–84%), 74% (68%–79%), 84% (77%–89%), 78% (73%–82%), respectively; specificity 88% (83%–92%), 76% (71%–79%), 84% (77%–89%), 78% (74%–81%), respectively; positive likelihood ratio 6.8 (4.5%–10%), 3 (2.5%–3.7%), 5.2 (3.6%–7.4%), 3.5 (3%–4.2%), respectively; negative likelihood ratio 0.25 (0.18%–0.35%), 0.34 (0.27%–0.43%), 0.19 (0.13%–0.28%), 0.28 (0.23%–0.34%), respectively; pooled area under the operator curve 0.77, 0.7, 0.83, 0.73, respectively. The pooled area under the operator curve for all modalities was not significantly different (all: $P \geq .4$).

Conclusion: As a result of the systemic review and meta-analysis, we identified TVUS and

MRI as good and comparable non-invasive imaging methods for diagnosisng adenomayosis, leading us to recommend TVUS as the first-line diagnostic imagin method and MRI as second if TVUS is inconclusive.

Keywords

diagnostic test, magnetic resonance imaging, ultrasonography, uterus, transvaginal ultrasound

Introduction

Adenomyosis is a disease of the uterus described by pathologists more than 150 years ago (1), and until recently, the ectopic endometrial glands within the uterine myometrium that define adenomyosis could only be seen through a microscope during histopathologic examination after hysterectomy. It was, therefore, the only way to diagnose this condition until the development of imaging modalities made it possible to diagnose adenomyosis without removal of the uterus. The first report of the use of imaging modalities for diagnosing adenomyosis was published in the early 1980s (2)

In addition to postmenopausal patients, it has been noted that younger females suffer from adenomyosis as well (3). Because adenomyosis impacts reproductive outcomes that lead to complications in pregnancy and during childbirth, it is crucial to diagnose adenomyosis as early as possible and doing so with imaging techniques is less invasive than surgery (4).

Few studies have reviewed the various imaging modalities that can be used for diagnosis of adenomyosis, and selection criteria for studied populations vary. With the need to identify adenomyosis in younger patients, it is of interest to investigate whether previous studies that typically included postmenopausal patients are applicable with younger females. The objective of this study was to perform a systematic review and meta-analysis to evaluate the use of imaging methods to diagnose adenomyosis.

Material and Methods

This systematic review was registered in the International Prospective Register of Systematic Reviews (PROSPERO protocol CRD42019125405) and follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) checklist (5-7). Institutional review board approval was not required owing to the nature of the retrospective literature review.

Eligibility criteria

Studies were included that used noninvasive diagnostic imaging methods for diagnosing adenomyosis, and histopathology as a reference standard obtained by hysterectomy. In addition, both prospective and retrospective studies and pre- and postmenopausal study participants irrespective of hormonal therapy or the use of gonadotropin-releasing hormone (GnRH)-agonists or antagonists were included. Also included were original research studies published as research papers. Exclusion criteria included data presented in short communications, reviews, letters to the editors, and congress abstracts, as well as the use of histopathological examinations from other uterine surgical procedures such as myometerial biopsies, myomectomies, and transcervical resections as a reference standard because these procedures do not involve the whole uterus.

The literature search was completed with support from a librarian with training in scientific and systematic literature searches. The search included the Cochrane Central Register of Controlled Trials, EMBASE and PubMed/MEDLINE databases from January 2000 to June 2019. Studies performed earlier were not included owing to limitations in earlier imaging quality. Publication language was included English, Swedish, Norwegian, Danish, and German. A combination of the following MeSH terms was used: magnetic resonance, MRI, MR imaging, ultrasonograph*, ultrasound*, ultrasonic*, sonograph*, echograph*, elasticity imaging technique*, elastograph*, sonoelastograph* (*allowing different endings to the word). The reference lists of reviews and original research publications that were identified were searched to detect publications that might have been missed.

Study selection and quality assessment

In the first step, all obtained references were independently screened based on title and abstract by TT and ML, using the Rayyan web application (8) that allows a blinded assessment. In the second step, all abstracts were transferred to the citation manager, full

texts were reviewed when necessary, and current study authors viewed the manuscripts with conflicting decisions until there was agreement. In the third step, the manuscripts that met the inclusion criteria were checked for quality of methodology, using the Quality Assessment of Diagnostic Accuracy Study (QUADAS)-2 checklist for relevant items of methodology (5). A study was considered of high quality if determined to be a low risk for bias in study participant selection, performance and evaluation of the index test (imaging method), and when the reference standard (histopathology) was performed with satisfactory reliability. If the reference test was performed with different methods in the same population (eg, hysterectomy and myometrial biopsies) and it was possible to identify the subgroups, only those that underwent hysterectomy were included in the analysis. Also, the applicability of the results of the study to a general population was considered.

Data extraction and synthesis

Numbers of study participants in each study were noted as was overall diagnostic performance, diagnostic modality, and diagnostic evidence of disease used, where possible. Data were extracted for the study participants that underwent hysterectomy, and when possible for 2D and 3D ultrasound separately.

Data analysis was performed via the R-package Meta-Analysis of Diagnostic Accuracy (mada, version 0.5.8) (9). To provide an overview of the heterogeneity of the studies, descriptive statistics were performed using forest plots displaying sensitivities, specificities, positive likelihood ratios (posLR), and negative likelihood ratios (negLR), all with 95% confidence intervals (CIs). A positive likelihood ratio (sensitivity/100 ≥ specificity) of 2 to 5 indicates a fair clinical test, 5 to 10 indicates a good clinical test, and >10 indicates an excellent clinical test. A negative likelihood ratio (100 ≥ sensitivity/specificity) of 0.5 to 0.2 indicates fair, 0.2 to 0.1 indicates good, and <0.1 indicates excellent (10). Next, a quantitative synthesized analysis was performed using a bivariate model (9) and implemented as the Reitsma-function in the mada package (11), a linear mixed effects model. The main output of this analysis is summarized receiver operating characteristics

(sROC) and a pooled area under the ROC-curve (pAUC) for each measurement type (aggregated over all included studies). Finally, the sROC curves were compared between the measurement types. $P \leq .05$ was considered statistically significant.

Results

Study selection

1,168 citations were identified. Figure 1 notes the process of inclusion as well as exclusion of studies. After detailed evaluation, 26 publications were evaluated (supplemental Table 1 [12-36]: 7 publications describing magnetic resonance imaging (MRI) (12-18), 15 describing ultrasound (US) (16,19-32), and 4 publications describing elastography (33-36) were assessed for quality using the QUADAS-2 tool. Ten prospective studies were included in the final meta-analysis (Tables 1 and 2) (16-20, 22, 23, 28,31,32).

Some studies appeared to have duplicate study populations based on inclusion dates and hospitals, units, and authors and only the studies with the largest number of patients were included (17,29,31). In the studies that described both MRI and US, (we chose those with the largest population for each group (16, 17). An overview of the risk of bias and concern of applicability of the included studies is provided in Figure 2 [16-20,22,23,28,31,32]. The US studies included 827 patients [16,19,20,22,23,28,31,32], the MRI studies included 317 patients [16-18], and the overall prevalence of adenomyosis for all studies was 40% [16-20,22,23,28,31,32].

MRI studies [12-15] were excluded owing to including only patients with proven histopathology for adenomyosis, comprising of patients with bulky uteri >10 weeks gestation, insufficiently describing reference standard, or unclear inclusion/exclusion criteria, 3 MRI-studies satisfied the quality criteria [16-18]. Ultrasound studies [21,24-

27,29,30] were excluded because of inclusion bias, insufficient clarification regarding the use of transvaginal (TVUS) versus transabdominal ultrasound (TAUS), how tests were performed and participants selected, and the reference standard method used. All elastography studies [33-36] were excluded owing to the reference standard used (histopathology).

Rasmussen et al [32] presented a new technique of histopathological confirmation of adenomyosis in the inner myometrium that was performed with great accuracy via 3D-TVUS, and we chose to include cases with that reference standard in a sub-analysis of diagnostic markers that only included findings of the junctional zone (JZ).

Diagnostic accuracy MRI

The aggregated sensitivity (with 95%CI in brackets) for MRI was 78% (70—84), specificity 88% (88—92), posLR 6.8 (4.54—10), negLR 0.25 (0.18—0.35), indicating an overall good test quality for MRI. The plots for each individual study, as well as the combined results, are shown in Figure 3a and 3b [16-20,22,23,28,31,32]. Each MRI study used and reported different diagnostic predictors (Table 1) [16-18], therefore an aggregated analysis of individual predictors was not possible. Two of the three included MRI studies [16,17] are from approximately 20 years ago and included a high number of postmenopausal patients; one included a significant proportion of patients with uterine cancer [17]. Only the study of Tellum et al [18] prospectively validated and reported a JZ thickness cut off of ≥12mm.

Diagnostic accuracy ultrasound

From the original 1168 citations, TAUS was determined to be a poor diagnostic modality for adenomyosis and was not further studied.

Of the included 8 TVUS studies, 4 studies evaluated 2D-TVUS only [16,20,23,31], 3 studies evaluated a combination of 2D- and 3D-TVUS [19,28,32] and one study evaluated 3D-TVUS

[22]. We grouped the data in 2D-TVUS (including also the 2D data from the 3D-TVUS studies), 3D-TVUS (the reported 3D-TVUS data, and combined 2D/3D data) and TVUS-overall, where we used the combination of 2D/3D results if reported or the largest population from each study. The sensitivity, specificity, posLR and negLR for each modality, both aggregated and individually, is shown in Figure 3a; pAUC is shown in Figure 3b [16-20,22,23,28,31,32]. In summary, 3D-TVUS resulted in a good test quality, while 2D-TVUS and TVUS-all had a fair test quality. There was no statistically significant difference between the sAUC of the different TVUS modalities (*P*= .1).

The included studies showed a variation in study population characteristics (size, pre- and postmenopausal status, age) and other baseline-markers such as the prevalence of adenomyosis and quality of the reference standard (Table 2) [16,19,20,22,23,28,31,32]. Only one study had 2 readers [23, 31] who evaluated the images, the others only one. All studies had “experienced” or “expert-examiners” but most without a definition of this criterion. An overview of the characteristics of each study including remarkable selection criteria is given in Table 2 [16,19,20,22,23,28,31,32]. The type of diagnostic predictors that were used for the diagnosis of adenomyosis was different in all studies and were defined both pre- and post-hoc, and the number of predictors used for the diagnosis for adenomyosis ranged from 1 to 9 (supplementary Table 2) [12-36]. The individual performance of each marker is shown in Figure 4 [16-20,22,23,28,31,32]. The presence of a fanshaped echo (usually caused by a circular layer of hypertrophic muscle bundles around foci of adenomyosis) and the heterogeneous echogenicity of the myometrium had the best diagnostic quality (pAUC=0.86 and 0.83), followed by “irregular appearance of the JZ” as seen in 3D TVUS (pAUC=0.81).

Comparison of imaging modalities

Figure 3b illustrates the diagnostic performance of MRI, 2D-TVUS, 3D-TVUS, and TVUS-all by sROC-curves [16-20,22,23,28,31,32]. The sensitivity and specificity of the 2D-TVUS

studies was most scattered. The pAUCs showed no statistically significant difference between the various modalities (all *P*>.11) .

Discussion

In this meta-analysis, pooled MRI had a sensitivity of 78% and a specificity of 88%, 2D-TVUS 74% and 76%, and 3D-TVUS 84% and 84% for diagnosing adenomyosis. All modalities had good or sufficient diagnostic quality for the diagnosis of adenomyosis. 3D-TVUS improved the diagnostic quality for diagnosing adenomyosis compared to 2D-TVUS and can detect changes in the JZ, which was one of the best performing diagnostic determinants. 3D-TVUS is in practice used together with real-time 2D-TVUS, and we recommend combining modalities to benefit from the strength of each. There was no statistically significant difference between the diagnostic quality of MRI and TVUS, which is in line with an earlier meta-analysis [37].

The evidence of the presented data is strong owing to the evaluation of high-quality studies with proper reference standards (histopathology obtained by hysterectomy). Though hysterectomy introduces a possible bias toward older patients who have finished childbearing and present with more advanced disease, the lack of a consensus on non-histopathologic diagnostic criteria makes it the best option so far.

The publications included in this study [16-20,22,23,28,31,32] have limitations that need consideration. The main limitation is that many diagnostic parameters were not defined before the study was commenced but determined post-hoc. This could lead to an overestimation of the diagnostic performance of those parameters as results could be over-fitted to the sample they are derived from. As a result, they that might not perform as well in a different population. The overall diagnostic accuracy used in this meta-analysis was mostly reported pooled for pre- and post-hoc determined signs, what might result in an over-estimation of the diagnostic accuracy of both MRI and TVUS. Overfitting represents a major bias, especially with numerical cut offs; for example the JZ_{max} of ≥12mm as a diagnostic cut-off for adenomyosis in MRI was determined post-hoc for the first time in a study of Reinhold

et al [38]. It has been used since that time as the main diagnostic criterion, but could not be reproduced in a recent prospective study performed on a younger, premenopausal population [18].

Furthermore, the TVUS-examination in the current included studies [16-20,22,23,28,31,32] was mostly performed by experts and only one reader, which could limit the generalizability of the diagnostic performance described as noted in Rasmussen et al [39]. Also, image resolution and automated optimalization in ultrasound and MRI has improved substantially over the years so that equipment may play a role in diagnostic accuracy. However, the experience of the examiner is most likely a key factor in diagnosing adenomyosis [40], as also studies with less sophisticated equipment yielded good results [19].

Another limitation of the current study when comparing the included studies is the heterogeneity of the study populations in age, parity, and prevalence of pre- and postmenopausal patients.

The sensitivity of a diagnostic predictor for adenomyosis depends on the extent of the condition, and as adenomyosis often progresses over time, the demographics of a study population can influence diagnostic features. We could not evaluate if the various diagnostic signs perform differently in pre- and postmenopausal patients, as study design and type of ultrasound were relevant confounders. Furthermore, 3D-TVUS was introduced at the same time as only premenopausal patients were included in studies, making it impossible to perform subgroup analysis without the risk of confounding.

Also, hormonal and GnRH treatment can alter the appearance of and efficiency of diagnosis of adenomyosis. As those groups were not reported individually in the included studies, it was not possible to calculate the effect of hormonal/GnRH treatment on the diagnostic quality of MRI or US. However, based on previous publications, it is important to be cautious when excluding adenomyosis in patients that are under current hormonal/GnRH therapy as imaging is less accurate in that case [41].

With the improvement of imaging, such as MRI with thin sections, it is possible to visualize

small hemorrhagic foci in 3D-reformats, and enhanced images using 3D-TVUS, we might move on to the post-histopathological era of adenomyosis diagnosis [42]. With any study of this type, there must be undisputed prestudy definitions of diagnostic criteria as proposed by the Morphological Uterus Sonographic Assessment group [43,44]. Also, it is essential to be aware of the strength and limitations of various diagnostic signs [19]. It was previously discussed that at least two diagnostic signs should be present when diagnosing adenomyosis [45]. The data in this meta-analysis do not allow for the conclusion of how many diagnostic signs should be present to confirm adenomyosis. The number of diagnostic markers that were used increased during the last two decades. Using bivariate meta-regression, we analyzed whether the number of variables that were used influenced the diagnostic quality of the studies, but could not find a significant association (*P*=.06). When also including variables that were determined post-hoc, there was a statistically significant association between the number of variables used and the diagnostic performance (*P*=.002). This association might be the result of overfitting. The optimal number most likely depends on a combination of signs and their specificity found in the individual patient. In addition, it is essential to perform a good clinical examination and history, as the symptoms presented will often help to find the correct diagnosis and determine the relevance of imaging findings [19,46,47]. Incidental findings of small foci with adenomyosis in asymptomatic females might not have clinical relevance but needs to be studied further.

Conclusion

TVUS and MRI provide a good diagnostic quality for the diagnosis of adenomyosis and show no statistically significant difference in their diagnostic ability. As it is more available, faster, and less costly, TVUS (preferably 2D- and 3D-TVUS in combination) should be the first choice when diagnosing adenomyosis and if inconclusive, MRI should be used.

References

1. Rokitansky C. Über Uterusdrüsen-Neubildung in Uterus- und Ovarial-Sarcomen. Zeitschr Ges Aerzte Wien. 1860;16:577 -81.
2. Weseley AC. The preoperative diagnosis of adenomyosis. Diagn Gynecol Obstet. 1982;4(2):105-6.
3. Dietrich JE. An update on adenomyosis in the adolescent. Curr Opin Obstet Gynecol. 2010;22(5):388-92.
4. Bruun MR, Arendt LH, Forman A, Ramlau-Hansen CH. Endometriosis and adenomyosis are associated with increased risk of preterm delivery and a small-for-gestational-age child: a systematic review and meta-analysis. Acta Obstet Gynecol Scand. 2018;97(9):1073-90.
5. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3:25.
6. Cronin P, Kelly AM, Altaee D, Foerster B, Petrou M, Dwamena BA. How to Perform a Systematic Review and Meta-analysis of Diagnostic Imaging Studies. Acad Radiol. 2018;25(5):573-93.
7. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA*. 2018;319(4):388-96.
8. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Systematic reviews*. 2016;5(1):210.
9. DOEBLER P. https://CRAN.R-project.org/package=mada [Pdf]. Philip Doebler; 2017.
10. McGee S. Simplifying likelihood ratios. *J Gen Intern Med*. 2002;17(8):646-9.
11. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* 2005;58(10):982-90.
12. Yang Q, Zhang LH, Su J, Liu J. The utility of diffusion-weighted MR imaging in differentiation of uterine adenomyosis and leiomyoma. *Eur J Radiol*. 2011;79:e47-51.
13. Stamatopoulos CP, Mikos T, Grimbizis GF, et al. Value of magnetic resonance imaging in diagnosis of adenomyosis and myomas of the uterus. *J Minim Invasive Gynecol*. 2012;19:620-626.
14. Moghadam R, Lathi RB, Shahmohamady B, et al. Predictive value of magnetic resonance imaging in differentiating between leiomyoma and adenomyosis. *JSLS*. 2006;10:216-219.
15. Kido A, Fujimoto K, Matsubara N, Kataoka M, Konishi I, Togashi K. A layer of decreased apparent diffusion coefficient at the endometrial-myometrial junction in uterine adenomyosis. *Magn Reson Med Sci*. 2016;15:220-226.
16. Dueholm M, Lundorf E, Hansen ES, Sorensen JS, Ledertoug S, Olesen F. Magnetic resonance imaging and transvaginal ultrasonography for the diagnosis of adenomyosis. *Fertil Steril*. 2001;76:588-594.
17. Bazot M, Cortez A, Darai E, Uzan S. Ultrasonography compared with magnetic resonance imaging for the diagnosis of adenomyosis: correlation with histopathology. *Hum Reprod*. 2001;16:2427-2433.
18. Tellum T, Matic GV, Dormagen JB, et al. Diagnosing adenomyosis with MRI: a prospective study revisiting the junctional zone thickness cutoff of 12 mm as a diagnostic marker. *Eur Radiol*. 2019
19. Tellum T, Nygaard S, Skovholt EK, Qvigstad E, Lieng M. Development of a clinical prediction model for diagnosing adenomyosis. *Fertil Steril*. 2018;110:957-64.e3.
20. Sun YL, Wang CB, Lee CY, et al. Transvaginal sonographic criteria for the diagnosis of adenomyosis based on histopathologic correlation. *Taiwan J Obstet Gynecol*. 2010;49:40-44.

21. Sharma K, Bora MK, Venkatesh BP, et al. Role of 3D Ultrasound and Doppler in differentiating clinically suspected cases of leiomyoma and adenomyosis of uterus. *J Clin Diagn Res*. 2015;9:QC08-12.
22. Luciano DE, Exacoustos C, Albrecht L, et al. Three-dimensional ultrasound in diagnosis of adenomyosis: histologic correlation with ultrasound targeted biopsies of the uterus. *J Minim Invasive Gynecol*. 2013;20:803-810.
23. Kepkep K, Tuncay YA, Goynumer G, Tutal E. Transvaginal sonography in the diagnosis of adenomyosis: which findings are most accurate? *Ultrasound Obstet Gynecol*. 2007;30:341-345.
24. Bozkurt DK, Bozkurt M, Cil AS, Barut MU, Ersahin AA, Caliskan E. Concomitant use of transvaginal sonography and Doppler indices improve diagnosis of adenomyosis. *J Obstet Gynaecol*. 2017;37:888-895.
25. Harmanli OH, Bevilacqua SA, Dandolu V, Chatwani AJ, Hernandez E. Adenomyosis interferes with accurate ultrasonographic detection of uterine leiomyomas. *Arch Gynecol Obstet*. 2005;273(3):146-9.
26. Hanafi M. Ultrasound diagnosis of adenomyosis, leiomyoma, or combined with histopathological correlation. *J Hum Reprod Sci*. 2013;6:189-193.
27. Ferraz Z, Nogueira-Martins N, Nogueira-Martins F. Adenomyosis: Back to the future? *Facts Views Vis ObGyn*. 2017;9:15-20.
28. Exacoustos C, Brienza L, Di Giovanni A, et al. Adenomyosis: three-dimensional sonographic findings of the junctional zone and correlation with histology. *Ultrasound Obstet Gynecol*. 2011;37:471-479.
29. ElKattan EA, Omran EF, Al Inany HG. The accuracy of transvaginal ultrasound and uterine artery Doppler in the prediction of adenomyosis. *Middle East Fertil Soc J*. 2010;15:73-78.
30. Bromley B, Shipp TD, Benacerraf B. Adenomyosis: sonographic findings and diagnostic accuracy. *J Ultrasound Med*. 2000;19:529-34.
31. Bazot M, Darai E, Rouger J, Detchev R, Cortez A, Uzan S. Limitations of transvaginal sonography for the diagnosis of adenomyosis, with histopathological correlation. *Ultrasound Obstet Gynecol*. 2002;20:605-611.
32. Rasmussen CK, Hansen ES, Ernst E, Dueholm M. Two- and three-dimensional transvaginal ultrasonography for diagnosis of adenomyosis of the inner myometrium. *Reprod Biomed Online*. 2019;38:750-760.
33. Liu X, Ding D, Ren Y, Guo SW. Transvaginal elastosonography as an imaging technique for diagnosing adenomyosis. *Reprod Sci*. 2018;25:498-514.
34. Acar S, Millar E, Mitkova M, Mitkov V. Value of ultrasound shear wave elastography in the diagnosis of adenomyosis. *Ultrasound*. 2016;24:205-213.
35. Stoelinga B, Hehenkamp WJ, Brolmann HA, Huirne JA. Real-time elastography for assessment of uterine disorders. *Ultrasound Obstet Gynecol*. 2014;43:218-226.
36. Tessarolo M, Bonino L, Camanni M, Deltetto F. Elastosonography: a possible new tool for diagnosis of adenomyosis? *Eur Radiol*. 2011;21:1546-1552.
37. Champaneria R, Abedin P, Daniels J, Balogun M, Khan KS. Ultrasound scan and magnetic resonance imaging for the diagnosis of adenomyosis: systematic review comparing test accuracy. *Acta Obstet Gynecol Scand*. 2010;89(11):1374-84.
38. Reinhold C, McCarthy S, Bret PM, Mehio A, Atri M, Zakarian R, et al. Diffuse adenomyosis: comparison of endovaginal US and MR imaging with histopathologic correlation. *Radiology*. 1996;199(1):151-8.
39. Rasmussen CK, Hansen ES, Dueholm M. Inter-rater agreement in the diagnosis of adenomyosis by 2- and 3-dimensional transvaginal ultrasonography. *J Ultrasound Med* 2019;38(3):657-66.
40. Rasmussen CK, Glavind J, Madsen LD, Uldbjerg N, Dueholm M. Repeatability of Junctional Zone Measurements Using 3-Dimensional Transvaginal Sonography in Healthy Fertile Women.*J Ultrasound Med* . 2016;35(7):1497-508.
41. Vannuccini S, Luisi S, Tosti C, Sorbi F, Petraglia F. Role of medical therapy in the management of uterine adenomyosis.Fertil Steril. 2018;109(3):398-405.

42. Exacoustos C, Zupi E. A new era in diagnosing adenomyosis is coming. Fertil Steril 2018;110(5):858.
43. Van den Bosch T, de Bruijn AM, de Leeuw RA, Dueholm M, Exacoustos C, Valentin L, et al. A sonographic classification and reporting system for diagnosing adenomyosis. 2018.
44. Van den Bosch T, Dueholm M, Leone FP, Valentin L, Rasmussen CK, Votino A, et al. Terms, definitions and measurements to describe sonographic features of myometrium and uterine masses: a consensus opinion from the Morphological Uterus Sonographic Assessment (MUSA) group.*Ultrasound Obstet Gynecol* . 2015;46(3):284-98.
45. Exacoustos C, Manganaro L, Zupi E. Imaging for the evaluation of endometriosis and adenomyosis. Best Pract Res Clin Obstet Gynaecol 2014;28(5):655-81.
46. Ates S, Ozcan P, Aydin S, Karaca N. Differences in clinical characteristics for the determination of adenomyosis coexisting with leiomyomas. *J Obstet Gynaecol Res* 2016;42(3):307-12.
47. Naftalin J, Hoo W, Nunes N, Holland T, Mavrelos D, Jurkovic D. The association between ultrasound features of adenomyosis and severity of menstrual pain.*Ultrasound Obstet Gynecol* . 2015.

Acknowledgment
We would like to thank Marie S. Isaachsen, at the Medical Library at the University of Oslo, for her professional help with designing and conducting the literature search.

Figure legends

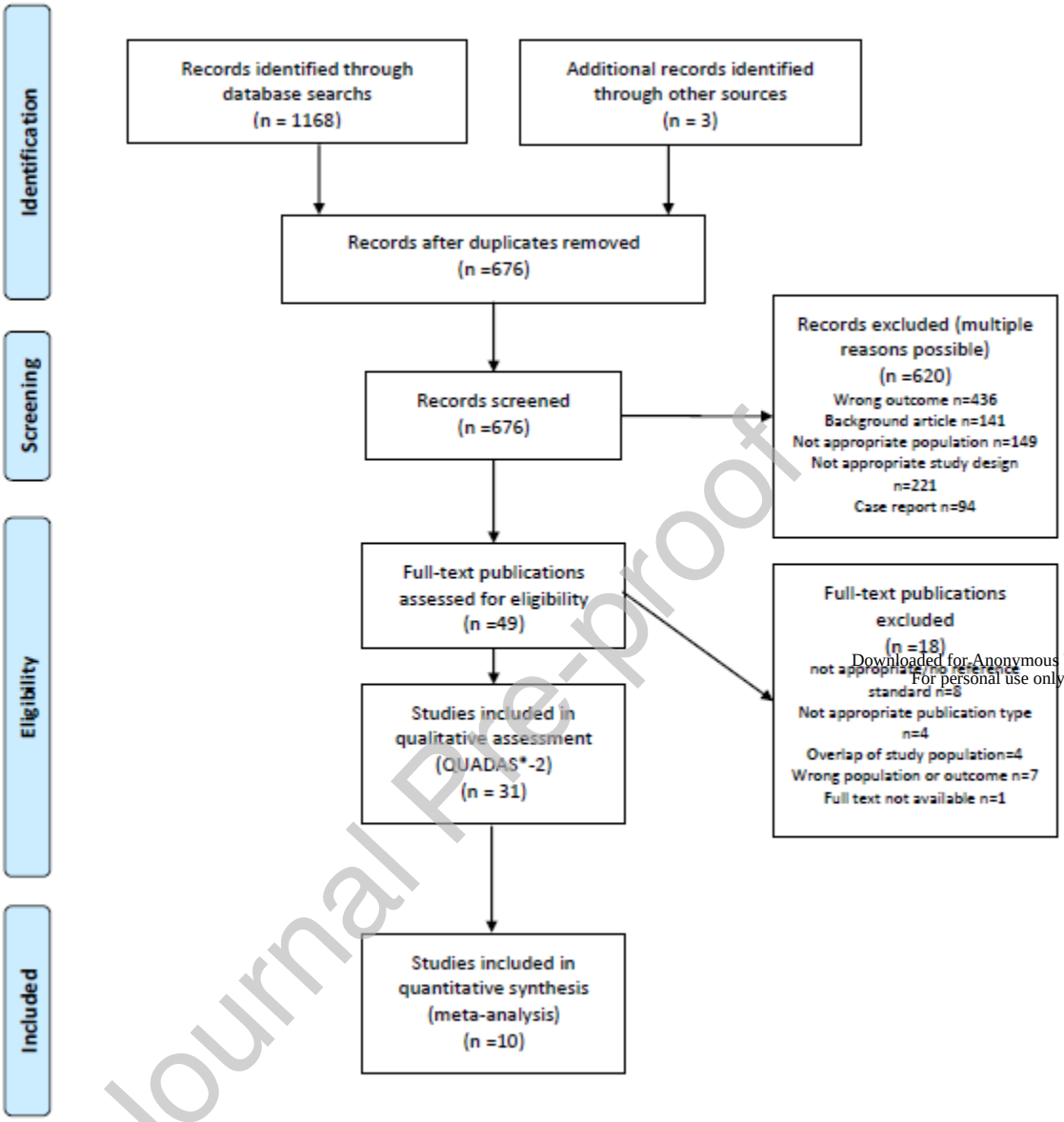

Figure 1. Flow Diagram with overview over publication selection. QUADAS; Quality Assessment of Diagnostic Accuracy Study-tool.

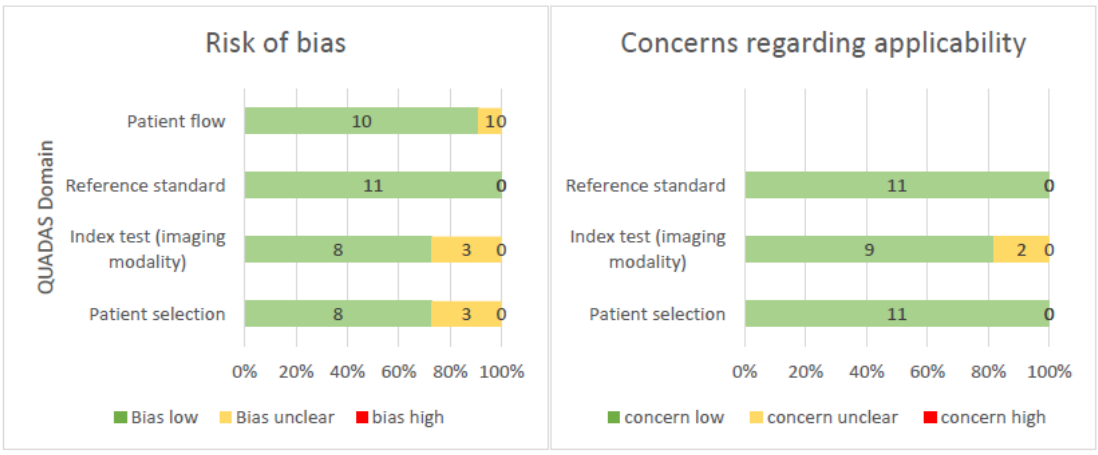

Figure 2. Proportion of included studies with high, low and unclear risk of bias or concerns regarding applicability [16-20,22,23,28,31,32]. QUADAS; Quality assessment of diagnostic accuracy studies.

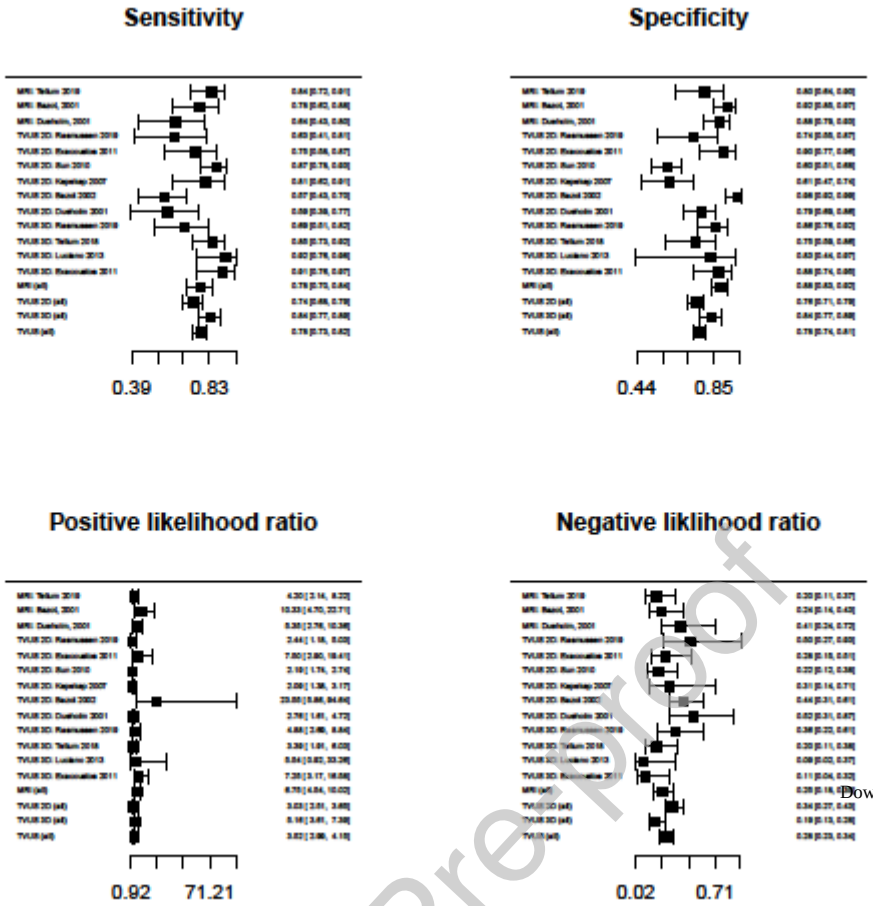

Figure 3a. Forest plots for sensitivity, specificity, positive likelihood ratio, negative likelihood ratio for each individual study and pooled [95%-confidence intervals]. Studies that used ≥1 diagnostic marker: Dueholm, Bazot, Kepkep. Studies that used at least 2 diagnostic markers: Exacoustos, Luciano, Rasmussen, Tellum.

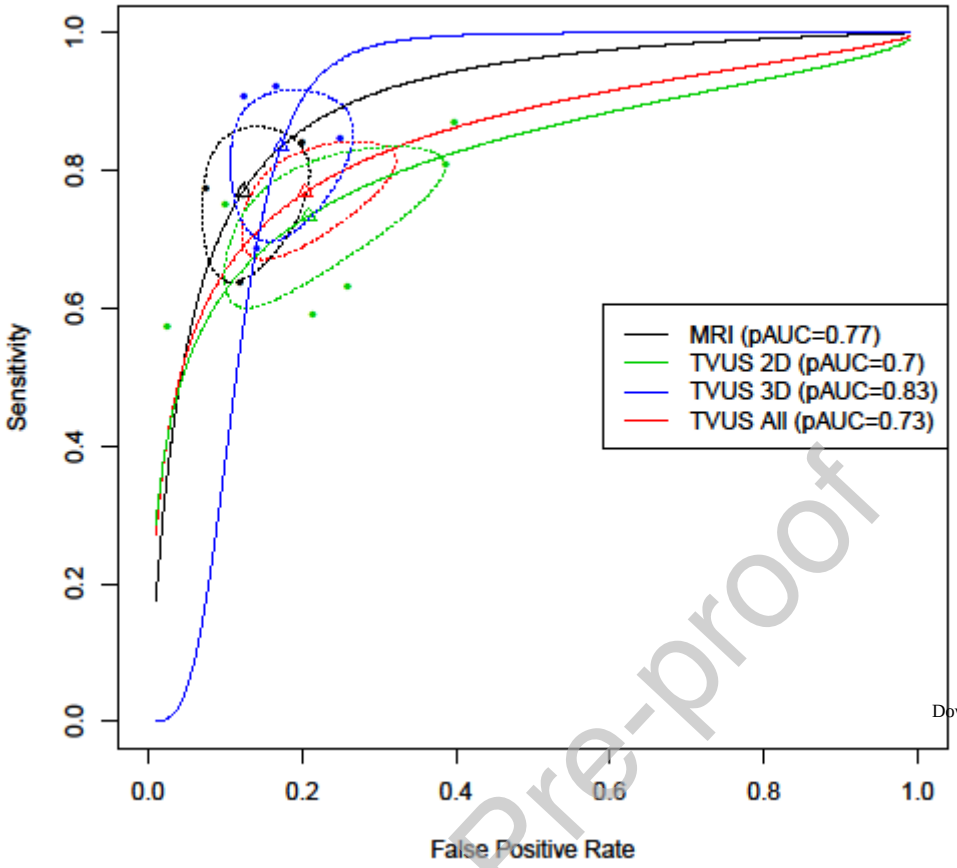

Figure 3b. Pooled area under the receiver operating characteristics curve (pAUC) for the different imaging modalities. The colored dots represent the individual studies, the areas marked with the dotted lines the 95% confidence interval. There was no difference of performance between the various modalities (comparison of pAUC MRI vs 2D-TVUS P=.45; MRI vs 3D-TVUS P=.47; MRI vs TVUS all P=.75) [16-20,22,23,28,31,32]. MRI; magnetic resonance imaging. TVUS; transvaginal ultrasound. 2D; 2-dimensional. 3D; 3-dimensional.

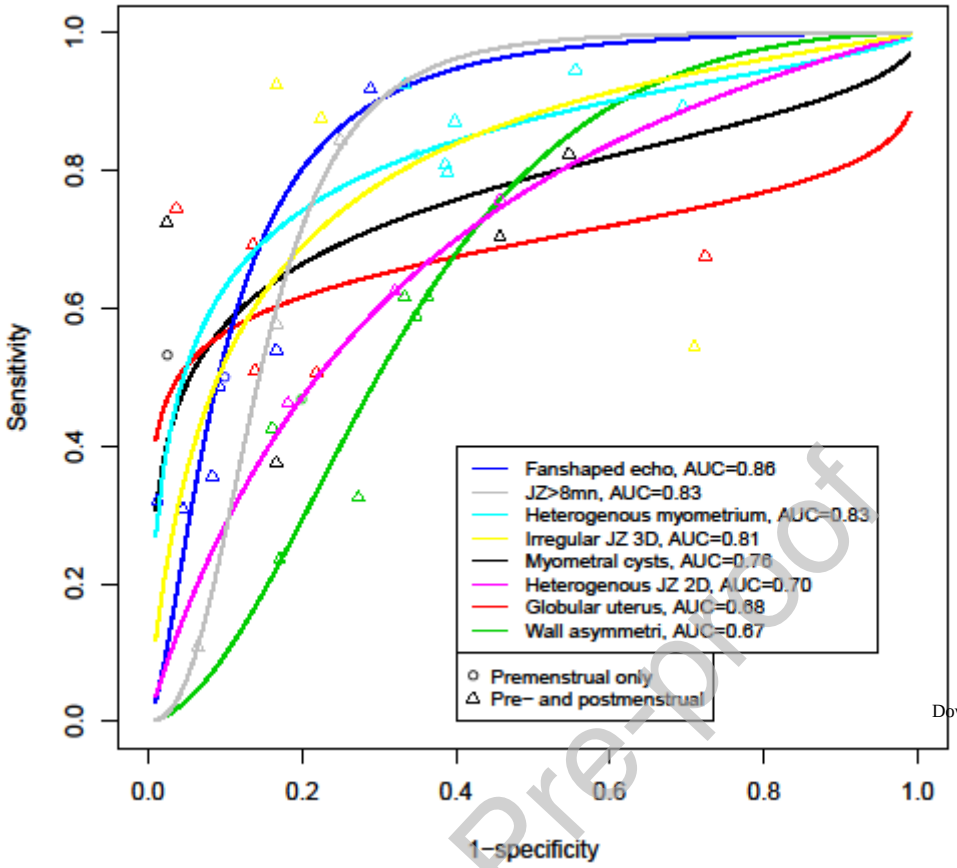

Figure 4. Pooled area under the receiver operating characteristics curves (pAUC) for individual diagnostic signs that were described for 2D- or 3D transvaginal ultrasound (TVUS) [16-20,22,23,28,31,32]. Note that the various signs have different diagnostic qualities and need to be combined for a reliable diagnosis.

	N	Adenomyosis, n (%)	Postmenopausal, n (%)	Mean age, years (range)	MRI sequences	MRI Slice thickness/ intersection gap	MRI system	Diagnostic criteria; prospectively defined	Diagnostic criteria; determined posthoc	Comments
Tellum et al 2019 [18]	93	57 (61)	0 (0)	42.7 (30-50)	T2W TSE (cor, obl. ax, sag) T1W TSE T1W TSE FS	3mm/0.3mm (3D Vista: 1mm/-0.5mm)	3.0 Tesla 1.5 Tesla	• JZ_{max} ≥12mm • High-intensity spots on T2, • ill-defined low intensity areas	JZ-morphology (definition similar to MUSA description)	2 independent readers
Dueholm et al 2001 [16]	106	22 (21)	0 (0)	44.7 (28-58)	T2W TSE (cor, obl. ax, sag)	4mm/1mm	1.5 Tesla	• JZ_{max} ≥15 mm • JZ_{max} 12-15mm AND nonuniform, thickened JZ OR focal not well-demarcated high OR low intensity areas in the myometrium	• JZ_{diff} (=JZ_{max}-JZ_{min}) • JZ_{max}≥12mm	
Bazot et al 2001 [17]	120	40 (33)	37 (31)	51 (30-88)	T2W SE or TSE (cor, obl. ax. or sag) T1W SE	5mm/1mm	1.5 Tesla	• a large, regular, asymmetric uterus • JZ_{max} of at least 12 mm AND/OR an ill-defined, low-signal-intensity myometrial area distinguished from well-circumscribed masses related to myoma • $Ratio_{max}$* 40% • punctate high-intensity myometrial foci		• 27% of population had uterine carcinoma; • 2 independent readers

Table 1: Overview of characteristics of the included magnetic resonance imaging (MRI) studies. JZ; junctional zone. T2W or T1W; T2/T1-weighted. TSE; turbo spin echo. SE; spin echo. Cor; coronal. Sag; sagittal. Ax; axial. Obl; oblique. *$Ratio_{max}$ was defined as the ratio between the uterine wall and the JZ_{max} measured at the same place.

	Study	N	Adenomyosis, n (%)	Fibroids, n (%)	Mean age, year (range)	Postmenopausal women, n (%)	Histopathologic definition of adenomyosis*	Histopathologic sections taken from specimen, number	Type of transducer	Experience of sonographer, years/number of readers	Exclusion/inclusion c differing from other st
3D- TVUS	Rasmussen et al 2019 [32]	46	19 (41)	?	46 (44-47)	0 (0)	≥2mm or 1 microscopic powerfield	With 5-10mm interval	6-12Mhz	10/1	Excl.: if ≥5 fibroids or ut volume >300ml due to f
	Tellum et al 2018 [19]	95	59 (62)	51 (54)	42.7 (30-50)	0 (0)	≥2.5mm	8 ±2.6	5-9Mhz	6/1	Excl: Use of hormones/
	Luciano et al 2013 [22]	32* *only those without TCER or medical treatment	26 (80)	?	41.2 (34-54)	0 (0)	≥2.5mm	≥8	2.8-10Mhz	>5/	Excl: Fibroids >4cm siz than 3 fibroids >3cm
	Exacoustos et al 2011 [28]	72	32	54 (73%)	46.7 (38-52)	0 (0)	≥2.5mm	?	6-9Mhz	"expert"/1	Excl: GnRH therapy, fib >8cm or more than 3 >5 size
2D-TVUS	Sun et al 2010 [20]	213	85 (40)		44.7 (24-73)	28 (87)	≥2.5mm	4-8	7-9Mhz	5y/1	Included: n=15, (7%) w endometrial carcinoma hyperplasia
	Kepkep et al 2007 [23]	70	26 (37)	20(28)	49 (37-63)	18 (26%)	≥2.5mm	20	5Mhz	8, 3/2 readers	
	Bazot et al 2002 [31]	129	47 (36)	?	NA	39 (30%)	≥2.5mm	5-15	5-9Mhz	>5/1	Excl.: prior myomectom endometrial resection, r excluded without given
	Dueholm et al 2001 [16]	106	22 (21)	?	44.7 (28-58)	0 (0)	≥2mm or 1 powerfield	?	5-8Mhz	3, 6/1	

Table 2: Overview of characteristics of the study populations, technical details of the index test and reference standard of the studies included in the meta-analysis. *Depth of ectopic endometrial glands below the endometrial-myometrial junction. ? Indicates that the information is not provided in the manuscript. GnRH; gonadotropin-releasing hormone agonist. Excl; excluded. TVUS; Transvaginal ultrasound. TCER; Transcervical endometrium resection.