Measuring intraobserver and intermethod reliability of endometriotic cyst volumes: A comparison between MRI and 3D transvaginal ultrasound in endometriosis

Jasmijn Bergwerff\textsuperscript{1,\*}, Anneke M. F. Schreurs\textsuperscript{1}, Marit C. I. Lier\textsuperscript{1}, Jan Hein T. M. van Waesbergh\textsuperscript{2}, Lisette E. E. van der Houwen\textsuperscript{1} and Velja Mijatovic\textsuperscript{1}

Abstract

**Objectives:** To identify the intraobserver and intermethod reliability of three-dimensional transvaginal ultrasound (3D-TVUS) using the software VOCAL and XI VOCAL compared with magnetic resonance imaging (MRI) for volumetric measurement of ovarian endometrioma.

**Methods:** The intermethod and intraobserver reliability of endometrioma volumes were assessed in 16 women diagnosed with endometriosis through laparoscopy with histologic confirmation and presenting with uni- or bilateral endometriomas. In total, volumes of 23 endometriomas were assessed with two-dimensional and three-dimensional transvaginal ultrasound and 6 mm magnetic resonance imaging. Examinations took place at two moments in one menstrual cycle: day 2–4 (T0) and day 20–22 (T1).

**Results:** The intraclass correlation for intraobserver reliability is good to very good for all three techniques ranging from the lowest value of 0.953 to the highest of 1.000. MRI has the most narrow limits of agreement (−3.93 to 4.53), followed by XI VOCAL (−5.16 to 5.65) while VOCAL has the widest limits of agreement (−10.22 to 11.39). Intraclass correlations are poor in the comparison of XI VOCAL to MRI, moderate between VOCAL and XI VOCAL, and good for the comparison between VOCAL and MRI. Limits of agreement vary per technique. When comparing 3D imaging techniques with 2D TVUS, XI VOCAL versus 2D TVUS provide the smallest limits of agreement.

**Conclusions:** MRI and XI VOCAL provide the best intraobserver reliability. The different imaging techniques are not interchangeable. As TVUS is a more readily available and cost-efficient imaging technique the usage of XI VOCAL is advised.

**Keywords**

Endometrioma, MRI, transvaginal ultrasound, intraobserver reliability, intermethod reliability, three-dimensional ultrasound

Date received: 18 October 2021; accepted: 4 February 2022
Introduction

Endometriosis is a benign estrogen-dependent disorder, affecting approximately 2%–10% of the female population of reproductive age. Ovarian endometriosis is the most common presentation occurring in approximately 17%–44% of endometriosis patients. In clinical practice, visualization of endometriomas is easily achieved with two-dimensional transvaginal ultrasound (2D-TVUS). Major volumetric changes can be visualized in time by using three orthogonal planes, which can be useful for treatment choices in individual patients.

Determining (small) volumetric changes in endometriomas can give more insight into the development and circumstances under which they grow or decrease in size. Three-dimensional transvaginal ultrasound (3D TVUS) might be more accurate to measure the volume of an endometrioma by using software of Virtual Organ Computer-aided AnaLysis (VOCAL) and eXtended Imaging Virtual Organ Computer-aided AnaLysis (XI VOCAL). Both techniques provide volume calculations based on manual delineations of the contour of objects.

Previous research has demonstrated that 3D ultrasound has better accuracy in determining volumes of both regular and irregular shaped objects when compared to 2D-TVUS. In vitro research has identified that intraobserver reliability is high for both VOCAL and XI VOCAL. Splenic volume and parathyroid volume have previously been reliably and accurately measured using 3D ultrasound. In the uterus, endometrial and uterine niche volumes can be measured adequately by 3D ultrasound as well.

In current practice, MRI has proven to be a diagnostic tool for various manifestations of endometriosis including endometriomas. MRI can also be used to perform volumetric measurements. To date, no research has been performed using 3D-TVUS and MRI imaging techniques to determine endometrioma volumes. Therefore, the objective of this study was to investigate whether MRI and TVUS are reproducible and reliable methods for volume measurements of endometriomas by evaluating the intraobserver- and intermethod reliability of three-dimensional imaging techniques. In addition, we aim to determine whether endometriomas show volumetric changes during one menstrual cycle since we hypothesize that endometriomas may undergo a volumetric increase due to estrogenic stimulation during the follicular phase.

Methods and materials

This study was part of a single center, prospective case-control study investigating Sampson retrograde menstruation theory by modern imaging technologies, which was performed between January 2010 and September 2016. The study was approved by the institutional review board of the Amsterdam UMC, location VUmc (METC VUmc 2009/329) and was registered in the Dutch Trial Register (Trial NL2106 (NTR2223)).

Recruitment took place at the endometriosis and fertility department of the Amsterdam UMC, location VUmc. All women who were diagnosed with an endometriosis cyst at transvaginal ultrasound during their appointment at the outpatient clinic and met the inclusion criteria were approached. Women diagnosed with endometriosis through laparoscopy with histological confirmation and presenting with uni- or bilateral endometriomas on TVUS were included in this study after obtaining informed consent. Other inclusion criteria for this prospective study were: (1) age ranging between 18 and 43 years and (2) a regular menstrual cycle (28 days ± 3 days). Patients were excluded if they (1) had a positive pregnancy test, (2) had any contra-indications for undergoing MRI, (3) used hormonal treatment, and/or (4) had a suspicion of uterine, peritoneal, enteral, and/or ovarian malignancy after transvaginal ultrasound and/or MRI.

Procedures

All participating women were seen for examinations at two time points during one menstrual cycle. The first examination took place at the beginning of the menstrual cycle (cycle day 2–4, T0) and the second examination was planned post ovulatory (cycle day 20–22, T1). At both time points a 2D and 3D TVUS using an Accuvix XQ (Medison, Hoofddorp, The Netherlands) and a MRI at 1.5 T (Avanto, Siemens, Erlangen, Germany) were performed. Women were asked to empty their bladder prior to transvaginal ultrasound. Transvaginal ultrasound examinations were performed according to the principles as also advised by the IDEA group. MRI included High Resolution T2-weighted Imaging in three planes and axial T1-weighted with fat saturation with a slice thickness of 6 mm. No intravenous contrast was used. All images were stored according to hospital policy. TVUS were performed by the same expert sonographer with experience in endometriosis. MRI images were described by the same radiologist specialized in endometriosis imaging. An overview of the procedures can be found in the supplementary data.

Volume measurements

The dataset collected according to the above stated procedures allowed for the determination endometrioma volumes at a later time point. Based on the three measurements recorded by 2D transvaginal ultrasound a volume could be derived using the following formula:

$$V = \frac{D_1 \times D_2 \times D_3 \times \pi}{6}$$

The obtained 3D image dataset was assessed by one observer (LvdH) using VOCAL and VOCAL XI. The VOCAL method made use of an 18° rotational step. The axial plane was used as a reference to mark the extremities of the cyst after which the cyst was rotated along its axis to obtain 10 rotational views. The area of the cyst was
manually traced on each plane. Subsequently a volume was automatically calculated by the program as well as a visual representation of the shape of the cyst (Figure 1). The XI VOCAL method uses 10 planes. Manual demarcation of the contours was performed per slice. The software automatically calculated the volume as well as the distance between the first and last slice (Figure 2). The use of the 18° rotational step and the 10 slices were chosen as both these techniques consist of 10 measurement steps. Endometriomas were assessed with MRI on axial T2-weighted images by one observer (JHvW). The surface area of the cyst was traced from which a volume was calculated using the following formula:

\[
\text{Surface area} \times \left( \frac{6}{1000} \right)
\]

In order to calculate intraobserver reliability all cyst volume measurements had to be performed twice by one observer. A period of at least 7 weeks separated the first and second cyst volume measurement.

**Statistical analysis**

Statistical analysis was performed using SPSS 24.0 software (IBM, USA). Patient characteristics were calculated using median, interquartile ranges, and percentiles. Intraobserver reliability and intermethod reliability were calculated with intraclass correlation (ICC) (two way-mixed, absolute agreement, single measurement). The ICCs are reported together with 95% confidence intervals. To assess the value of the intraclass correlation, cut-off values advised by Martins and Nastri were used: a value of < 0.70 is considered a very poor correlation, a value of 0.70-0.90 is a poor correlation, 0.90-0.95 is a moderate correlation, 0.95-0.99 is a good correlation and a value of >0.99 is an excellent correlation. The limits of agreement for the intraobserver and intermethod reliability were calculated according to Bland and Altman. The limits of agreement are set to include 95% of the calculated values. (Supplemental Figures 4 to 12).

**Results**

In total, 23 cysts from 16 patients were included. Four patients were excluded due to missing or incomplete three-dimensional imaging data. Patient characteristics are presented in Table 1.

**Intraobserver reliability**

The intraclass correlation for intraobserver reliability is good to very good for all three techniques ranging from the
The lowest value of 0.953 to the highest of 1.000 (Table 2). The Bland-Altman plots along with the limits of agreement for the intraobserver reliability can be found in the Supplemental Data. MRI has the narrowest limits of agreement (−3.93 to 4.53), followed by XI VOCAL (−5.16 to 5.65) while VOCAL has the widest limits of agreement (−10.22 to 11.39). Measurement of larger cysts show the most variation between measurements when using VOCAL. The Bland Altman plot for XI VOCAL has one outlier removed due to poor visibility, as some parts of the cyst wall were not able to be differentiated from surrounding tissue due to shadows. The difference between measurement at first and second assessment was approximately 40 mL. The Bland Altman plot for MRI has two outliers (one cyst at both T0 and T1) removed due to a volume exceeding 600 mL (138 mm × 131 mm × 80 mm). Including this cyst volume results in more than double the limits of agreement (−11.73 to 10.97 opposed to −3.93 to 4.53). The difference between the measurement at first and second assessment was approximately 10 and 30 mL at T0 and T1 respectively.

### Intermethod reliability

Table 3 summarizes the results of the intraclass correlation for intermethod reliability. Intraclass correlations are poor in the comparison of XI VOCAL to MRI, moderate between VOCAL and XI VOCAL, and good for the comparison between VOCAL and MRI. The Bland-Altman plots along with the limits of agreement for the interobserver reliability can be found in the Supplemental Data. The limits of agreement are widest for XI VOCAL versus MRI (−36.96 to 10.54). Similar limits of agreement are found between VOCAL versus XI VOCAL (−7.80 to 23.40) as between VOCAL versus MRI (−18.27 to 8.98). Comparing the three-dimensional imaging techniques to

---

### Table I. Patient characteristics.

|                      | Median [IQR]       |
|----------------------|--------------------|
| Age                  | 32.50 [29.00–35.00]|
| BMI                  | 21.27 [20.48–24.58]|
| VAS score            | 7.10 [4.55–8.08]   |
| Dysmenorrhea         | 1.50 [0.63–2.38]   |
| Dyspareunia          | 1.00 [0.00–1.50]   |
| Pelvic pain score    | 1.5 [1.00–2.00]    |
| Number of cysts (%)  |                    |
| 1                    | 9 (56.25)          |
| 2                    | 7 (43.75)          |

*Menstruation pain score over the last 3 months.
Based on average scores from visits at T0 and T1.
0: no symptoms, 1: mild, 2: average, 3: severe.
2D transvaginal ultrasound, the limits of agreement between XI VOCAL versus 2D TVUS are smallest (−13.54 to 16.83), followed by VOCAL versus 2D TVUS (−16.92 to 36.85) and widest for MRI versus 2D TVUS (−18.45 to 50.25).

Changes in cyst volume

Table 4 provides an oversight of the average changes in cyst volumes measured with MRI at T0 and T1. The difference in mean cyst diameter is added to show relative growth of the cyst compared to volume change.

Discussion

In this study, we investigated the reliability of 18° VOCAL and XI VOCAL (10 slices) 3D-TVUS and MRI in the measurement of endometriomas. Intraobserver reliability was excellent for MRI. In comparison, XI VOCAL demonstrated very good intraobserver reliability. The imaging modalities are not interchangeable due to poor intermethod reliability especially with increasing cyst size. We were unable to observe any significant changes in cyst volume within one menstrual cycle.

To our knowledge, this is the first study to determine the intraobserver- and intermethod reliability of different 3D imaging techniques when measuring endometriomas. A strength of this study is that all patients included had a cystectomy after receiving both scans allowing for histological confirmation of all visualized cysts. In addition, we were able to obtain a relatively large dataset due to measurements at two time points during the cycle. A limitation in our study is the lack of an absolute volume to compare the imaging data to, as we were unable to measure the exact volume of the cysts during surgery. The lack of a comparison makes it difficult to determine an acceptable error margin for the 3D imaging techniques. However, a previous in vitro study described VOCAL and XI VOCAL to overestimate the size of irregular shaped objects by 7.8% and 4% respectively.8

For the intraobserver reliability, XI VOCAL and MRI can be reliably used to reproduce measurements. We found that MRI provides the highest level of reliability when measuring endometriomas. XI VOCAL did produce smaller limits of agreement than VOCAL which could be explained by its ability to more accurately depict irregularly shaped objects.8 Additionally, we did not repeat 2D transvaginal ultrasound measurements to perform a first and second assessment per timepoint (T0 and T1), and therefore cannot provide insight into intraobserver reliability for 2D transvaginal ultrasound. A comparison between XI VOCAL and VOCAL has previously only been

Table 2. Intraobserver reliability VOCAL, XI VOCAL, and MRI at T0 and T1.

|                      | ICC    | 95% CI           |
|----------------------|--------|------------------|
| Intraobserver reliability VOCAL T0 | 0.987  | 0.968–0.995      |
| Intraobserver reliability VOCAL T1 | 0.992  | 0.980–0.997      |
| Intraobserver reliability XI VOCAL T0 | 0.997  | 0.991–0.999      |
| Intraobserver reliability XI VOCAL T1 | 0.953  | 0.888–0.980      |
| Intraobserver reliability MRI 6 mm T0 | 1.000  | 1.000–1.000      |
| Intraobserver reliability MRI 6 mm T1 | 0.999  | 0.997–1.000      |

Table 3. Intermethod reliability VOCAL, XI VOCAL, and MRI. First assessment and second assessment at T0 and T1.

|                      | ICC    | 95% CI           |
|----------------------|--------|------------------|
| First assessment     |        |                  |
| Intermethod reliability VOCAL vs XI VOCAL T0 | 0.927  | 0.615–0.977      |
| Intermethod reliability VOCAL vs XI VOCAL T1 | 0.911  | 0.713–0.967      |
| Intermethod reliability VOCAL vs MRI T0 | 0.972  | 0.921–0.990      |
| Intermethod reliability VOCAL vs MRI T1 | 0.985  | 0.807–0.996      |
| Intermethod reliability XI VOCAL vs MRI T0 | 0.889  | 0.331–0.970      |
| Intermethod reliability XI VOCAL vs MRI T1 | 0.846  | 0.330–0.954      |
| Second assessment    |        |                  |
| Intermethod reliability VOCAL vs XI VOCAL T0 | 0.938  | 0.602–0.982      |
| Intermethod reliability VOCAL vs XI VOCAL T1 | 0.972  | 0.806–0.992      |
| Intermethod reliability VOCAL vs MRI T0 | 0.969  | 0.910–0.989      |
| Intermethod reliability VOCAL vs MRI T1 | 0.983  | 0.910–0.995      |
| Intermethod reliability XI VOCAL vs MRI T0 | 0.887  | 0.369–0.968      |
| Intermethod reliability XI VOCAL vs MRI T1 | 0.929  | 0.360–0.982      |
examined for placental volume by Cheong et al.\(^{20}\) who reported similar results of high intraobserver reliability for both methods. They did however report smaller limits of agreement with both VOCAL and XI VOCAL.

Intermethod reliability varies from poor to good between the 3D imaging techniques. In the comparison between 3D imaging techniques and 2D TVUS the smallest limits of agreement are found in the comparison between 2D TVUS and XI VOCAL. From the intermethod Bland-Altman plots it seems that the measurement differences increase with cyst size. The measurement differences relative to cyst size are however not statistically significant. For future research, we advise the use of one imaging modality to measure volumetric changes as the limits of agreement are smaller when using one measurement technique. XI VOCAL can be advised above VOCAL as XI VOCAL provides smaller limits of agreement and more accurate measurements with larger cyst sizes. The use of XI VOCAL can be a more accurate measurement technique in comparison to 2D-TVUS as it takes into account the irregular shape of the cyst yet has the same practical benefits as TVUS. The practical benefits being that TVUS is relatively cheap, less time-consuming, and readily available at most gynecological outpatient departments. The benefit of 3D-TVUS imaging techniques compared to 2D-TVUS for use in the clinical practice has yet to be determined.

With the current set up of the study, we developed the second objective to determine whether any cyclical variances occurred in cyst volume. During one menstrual cycle we found no difference in endometrioma volume, we were therefore unable to show endometriomal growth under estrogenic stimulation in a single natural cycle. It could also be hypothesized that the cyst grows due to a build-up of menstrual debris. However, as the measurements were limited to one menstrual cycle we were unable to show volumetric changes during menstruation. The lack of growth seen within one cycle can be explained by the slow progressive nature of endometriosis, making it unlikely that any significant anatomical changes can be measured within this timeframe. Future research could focus on tracking the development of endometrial cysts during multiple cycles, as it remains unclear under which circumstances endometriomas grow or decrease in size. In addition, the growth of endometriomas could be a source of pain in women with endometriosis.

In conclusion, for measuring cyst volumes MRI provides the most reliable measurements of all three techniques. However, accurate measurements can also be performed by using 3D TVUS (XI VOCAL) as a more readily available and cost-efficient imaging technique.

### Authorship

V.M., J.W., and L.H. conceived and designed the study. J.W. and L.H. contributed to the acquisition of data. V.M., L.H., J.W., A.S., M.L., and J.B. contributed to the analysis and interpretation of the data. All authors made contributions to drafting and revising of the article. All authors approved the final version of the article to be published.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Ethical approval
Ethical approval for this study was obtained from the institutional review board of the Amsterdam UMC, location VUmc (METC VUmc 2009/329).

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Informed consent to participate
Written informed consent was obtained from all subjects before the study.

Informed consent to publish
Written informed consent was obtained from the patients for their anonymized information to be published in this article.

Trial registration
Dutch Trial Register: NL2106 (NTR2223).

ORCID ID
Jasmijn Bergwerff https://orcid.org/0000-0003-4986-3248

Supplemental material
Supplemental material for this article is available online.

References
1. Dunselman GA, Vermeulen N, Becker C, et al. ESHRE guideline: management of women with endometriosis. *Hum Reprod* 2014; 29: 400–412.
2. Eskenazi B and Warner ML. Epidemiology of endometriosis. *Obstet Gynecol Clin North Am* 1997; 24: 235–258.
3. Busacca M and Vignali M. Ovarian endometriosis: from pathogenesis to surgical treatment. *Curr Opin Obstet Gynecol* 2003; 15: 321–326.
4. Giudice LC. Clinical practice. Endometriosis. *N Engl J Med* 2010; 362: 2389–2398.
5. Levine D, Brown DL, Andreotti RF, et al. Management of asymptomatic ovarian and other adnexal cysts imaged at US Society of Radiologists in Ultrasound consensus conference statement. *Ultrasound Q* 2010; 26: 121–131.
6. Guimaraes Filho HA, da Costa LL, Araujo Junior E, et al. XI VOCAL (eXtended Imaging VOCAL): a new modality for three-dimensional sonographic volume measurement. *Arch Gynecol Obstet* 2007; 276: 95–97.
7. Downey DB, Fenster A and Williams JC. Clinical utility of three-dimensional US. *Radiographics* 2000; 20: 559–571.
8. Barreto EQ, Milani HJ, Araujo Junior E, et al. Reliability and validity of in vitro volume calculations by 3-dimensional ultrasonography using the multiplanar, virtual organ computer-aided analysis (VOCAL), and extended imaging VOCAL methods. *J Ultrasound Med* 2010; 29: 767–774.
9. Hidaka H, Nakazawa T, Wang G, et al. Reliability and validity of splenic volume measurement by 3-D ultrasound. *Hepatol Res* 2010; 40: 979–988.
10. You SH, Suh SI, Ryoo IS, et al. Assessing the accuracy and reliability of ultrasonographic three-dimensional parathyroid volume measurement in a patient with secondary hyperparathyroidism: a comparison with the two-dimensional conventional method. *Ultrasoundograpy* 2017; 36: 17–24.
11. Bordes A, Bory AM, Benchaab M, et al. Reproducibility of transvaginal three-dimensional endometrial volume measurements with virtual organ computer-aided analysis (VOCAL) during ovarian stimulation. *Ultrasound Obstet Gynecol* 2002; 19: 76–80.
12. Marjolein Bij de Vaate AJ, Linskens IH, van der Voet LF, et al. Reproducibility of three-dimensional ultrasound for the measurement of a niche in a caesarean scar and assessment of its shape. *Eur J Obstet Gynecol Reprod Biol* 2015; 188: 39–44.
13. Martins WP, Ferriani RA, Barra DA, et al. Reliability and validity of tissue volume measurement by three-dimensional ultrasound: an experimental model. *Ultrasound Obstet Gynecol* 2007; 29: 210–214.
14. Raine-Fenning N, Campbell B, Collier J, et al. The reproducibility of endometrial volume acquisition and measurement with the VOCAL-imaging program. *Ultrasound Obstet Gynecol* 2002; 19: 69–75.
15. Nisenblat V, Bossuyt PM, Farquhar C, et al. Imaging modalities for the non-invasive diagnosis of endometriosis. *Cochrane Database Syst Rev* 2016; 2: CD009591.
16. Cranney R, Condous G and Reid S. An update on the diagnosis, surgical management, and fertility outcomes for women with endometrioma. *Acta Obstet Gynecol Scand* 2017; 96: 633–643.
17. Guerriero S, Abrao M, Hudelist G, et al. P29.03: ultrasonography in deep endometriosis: a consensus opinion from the International Deep Endometriosis Analysis (IDEA) group. A preliminary statement. *Ultrasound Obstet Gynecol* 2011; 38: 265–265.
18. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
19. Martins WP and Nastri CO. Interpreting reproducibility results for ultrasound measurements. *Ultrasound Obstet Gynecol* 2014; 43: 479–480.
20. Cheong KB, Leung KY, Li TK, et al. Comparison of interand intraobserver agreement and reliability between three different types of placental volume measurement technique (XI VOCAL, VOCAL and multiplanar) and validity in the in vitro setting. *Ultrasound Obstet Gynecol* 2010; 36: 210–217.