



HAL
open science

Intra- and inter-observer variability of 2D and 3D transvaginal sonography in the diagnosis of benign versus malignant adnexal masses.

Maria A Pascual, Betlem Graupera, Lourdes Hereter, Anna Rotili, Ignacio Rodriguez, Juan Luis Alcazar

► To cite this version:

Maria A Pascual, Betlem Graupera, Lourdes Hereter, Anna Rotili, Ignacio Rodriguez, et al.. Intra- and inter-observer variability of 2D and 3D transvaginal sonography in the diagnosis of benign versus malignant adnexal masses.. *Journal of Clinical Ultrasound*, 2011, 39 (6), pp.316. 10.1002/jcu.20808 . hal-00632730

HAL Id: hal-00632730

<https://hal.science/hal-00632730>

Submitted on 15 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Intra- and inter-observer variability of 2D and 3D transvaginal sonography in the diagnosis of benign versus malignant adnexal masses.

Journal:	<i>Journal of Clinical Ultrasound</i>
Manuscript ID:	JCU-10-121.R1
Wiley - Manuscript type:	Research Article
Keywords:	adnexal mass, three-dimensional ultrasound , agreement, ultrasonography

SCHOLARONE™
Manuscripts

Review

Abstract

Objective:

Purpose:

The purpose of this study was to compare the intra-observer and inter-observer agreement in assigning a diagnosis of malignancy or benignity to adnexal masses using 2-dimensional ultrasonography (2D US) and 3-dimensional ultrasonography (3D US).

Methods:

Two experienced observers (Observers A and B) performed a retrospective review of digitally stored 2D images and 3D data from a sample of 41 consecutive patients with a diagnosis of adnexal mass. Each observer independently, and blinded each other, evaluated the 2D static images of each adnexal mass and then the 3D volumes one-week later. The observers were required to classify the adnexal lesion as benign or malignant according to pattern recognition analysis. Intra-observer and inter-observer agreement were assessed by calculating the kappa index (κ).

Results:

Intra-observer agreement between 2DUS and 3DUS for the observer A was 1.00 and for the observer B was 0.69. Inter-observer agreement was 0.69 for 2D US and 1.00 for 3D US (p >0.05)

~~Intra-observer agreement between 2DUS and 3DUS for the observer A was 0.96 and for the observer B was 0.77. Inter-observer agreement was 0.73 for 2D US and 0.92 for 3D US~~

Conclusions:

Our results indicate that 3D US for assessing adnexal masses is reproducible, but not significantly, than 2D US

~~Our results indicate that 3D US for assessing adnexal masses is reproducible, even more than 2D US.~~

Running Head: Three and Two-dimensional ultrasound: intra-inter-observer agreement

Keywords: adnexal mass, three-dimensional ultrasound, agreement , and ultrasonography

INTRODUCTION

Adnexal masses encompass a wide variety of ovarian disorders including benign lesions, malignant tumors and masses affecting the para-ovarian region. The use of 2D US allows discriminating benign from malignant lesions by the subjective evaluation of the gray scale ultrasound features, the so called “pattern recognition” analysis)^{1-6-1.4}. However, conventional two-dimensional ultrasound (2D US) has several limitations. One of its major disadvantages is that it relies on operator’s ability. In fact, recent studies have shown that reproducibility and confidence in making a specific diagnosis of 2D US depends on examiner’s experience^{7,8 5,6}

1
2
3 Three-dimensional ultrasound (3D US) has recently become available for clinical practice.
4
5 This technique has the potential to overcome some of the limitations of 2D US because it enables a
6
7 cross-sectional imaging from data obtained by a single sweep of the US beam across the involved
8
9 organ. Furthermore, with 2D US a “static” anatomic section is displayed on a video monitor while
10
11 3D US allows displaying anatomical features in planes not possible with conventional 2D US^{9,10,7}.
12
13 Although the benefits of 3D US have been emphasized^{11,8}, its role in assessing adnexal masses has
14
15 not been elucidated and more studies are required to demonstrate if it improves differentiation
16
17 between benign and malignant ovarian lesions and if it adds any real benefit compared to
18
19 conventional 2D US. Notwithstanding, before to be introduced definitively in clinical practice any
20
21 diagnostic method should be demonstrated as reproducible. As a matter of fact, one study has
22
23 shown that 3D US seems to be reproducible for assessing adnexal masses^{12,9}. However, to the best
24
25 of our knowledge, no study has assessed whether 3D US is more reproducible or not than 2D US
26
27
28
29
30
31
32

33 The purpose of our study was to compare the intra-observer and inter-observer agreement
34
35 between two experienced observers using 2D US and 3D US for assigning a diagnosis of
36
37 malignancy or benignity in the same set of adnexal masses.
38
39
40

41 MATERIALS AND METHODS

42
43
44 Although this is a retrospective study, Institutional review board approval was asked for and
45
46 obtained before performing the study.
47
48
49

50
51 A retrospective review of transvaginal sonograms obtained between May 2006 and January
52
53 2007 and stored in the PACS at our institution from 41 consecutive patients diagnosed as having an
54
55 adnexal mass evaluated and treated at our institution was performed. Patients with ectopic
56
57 pregnancy, lack of 3D volume, absence of surgery or ultrasound follow-up and an unquestionable
58
59 ultrasound diagnosis of tubal disorder were excluded.
60

1
2
3 All images were obtained with the same type of scanner (Voluson 730 Expert, GE Healthcare,
4 Milwaukee, USA), equipped with a transvaginal multi-frequency (2.9-10 MHz) transducer. Each
5 patient was first examined using conventional 2D US. Representative images from the adnexal
6 mass were digitally stored. After 2D US was performed 3D volume box was activated and a 3D
7 volume was acquired. 3D box was adjusted trying to include the whole adnexal mass within it.
8 When the volume acquisition was completed the data file was sent using a DICOM (Digital
9 Imaging and Communication in Medicine) connection to a personal computer. Representative 2D
10 images and 3D volumes were selected by one investigator (BG) to be analyzed by two other
11 observers (MAP, LH) who were blinded to 2D-image and 3D-volume selection. All selected 2D
12 images were retrieved from the database and showed without the possibility of being modified,
13 whereas the 3D stored volumes were analyzed off-line in a dedicated workstation by using
14 multiplanar reformatting with the 4D-view™ software, version 5.0 (GE Healthcare, Milwaukee,
15 USA).

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35 Two observers **A and B** (MAP and LH with, 25 and 20 years experience of gynecologic US,
36 respectively) were required to analyze 2D images in a random order. One week after 2D images
37 review they reviewed 3D volumes in a different order of cases than 2D images were to avoid recall
38 bias (referencia). The observers were blinded to each other's findings, they were unaware of the
39 diagnostic report. They were also unaware of the physical examination's findings and
40 histopathology results. The only information they had was the age and hormonal status of the
41 patient. For "static" 2D images only one representative gray-scale sonogram was available. 3D
42 volumes were analyzed using multiplanar rendering **(Figures 1 and 2)**. According to patten
43 recognition the observer had to provide a diagnosis of benignity or malignancy. Criteria for each
44 type of lesion were defined in previous publications^{13-16,10}

1
2
3 Patients with suspected functional and hemorrhagic cysts were managed with follow-up scans
4
5 at three and six months until spontaneous resolution. Patients with persistent or suspicious masses
6
7 were submitted to surgery and mass removal with pathological confirmation (gold standard).
8
9

10
11 Statistical analysis was performed using the SAS 9.1 Institute Inc Software (Cary, NC,
12
13 USA.). The intra-observer and inter-observer agreement analysis for classifying the mass as benign
14
15 or malignant were evaluated with Cohen's kappa index (κ) at the 95% confidence interval (CI). A
16
17 kappa value of <0 indicates no agreement, ≤ 0.20 indicates poor agreement, 0.21-0.39 indicates fair
18
19 agreement, 0.40-0.59 indicates moderate agreement, 0.60-0.79 indicates substantial agreement, and
20
21 0.80-1.00 indicates almost perfect agreement^{17,18}. 11,12.
22
23
24
25
26

27 Intra-observer agreement refers to the agreement between 2D US and 3D US diagnosis of
28
29 each examiner. Inter-observer agreement refers to the agreement between examiners when
30
31 providing a diagnosis using 2D US or 3D US. It was considered that an statistical significant
32
33 existed (p value less 0.05) when 95% CI of kappa values did not overlap.
34
35
36
37
38
39
40
41

42 RESULTS

43
44
45 The mean age of patients was 38.8 ± 8.3 . Among the 41 adnexal masses included in the study,
46
47 there were 34 benign masses and 7 malignancies. Table 1 summarizes the histological diagnoses.
48
49

50
51 Thirty-one patients underwent surgery with a histopathology result. Ten patients with a
52
53 diagnosis of functional or hemorrhagic cyst were followed up with 2DUS at three and six months
54
55 that showed a resolution of the lesion. Table 2 summarizes the agreement between
56
57 histopathologic result and both observers (A and B)
58
59
60

~~The intra-observer agreement between 2D US and 3D US was of 0.96 (95% CI: 0.91-1.0) for observer A and of 0.77 (95% CI: 0.60-0.99) for observer B.~~

The intra-observer agreement between 2D US and 3D US was of 1.00 (95% CI: 1.00-1.00) for observer A and of 0.69 (95% CI: 0.35-1.00) for observer B, respectively. (Tables 3 and 4)

~~The inter-observer agreement for 2D US was 0.73 (95% CI: 0.57-0.89) and 0.92 (95% CI: 0.84-0.99) for 3D US.~~

The inter-observer agreement for 2D US was 0.69 (95% CI: 0.35-1.00) and 1.00 (95% CI: 1.00-1.00) for 3D US (Tables 5 and 6). This difference was not considered statistically significant.

Discussion

The diagnostic efficacy of conventional 2D US in the diagnosis of adnexal masses is well documented^{19,20 13}. However, this technique is highly operator-dependent and based mainly on the subjective assessment of mass characteristics. In spite of this fact, few studies have addressed the issue of reproducibility of this technique for classifying adnexal masses as benign or malignant.

Timmerman et al showed that 2D US is reproducible among observers for providing a diagnosis of malignancy or benignity in adnexal masses^{21 14}. However, in this study anamnestic data and color Doppler findings were also included in the evaluation and final diagnosis. Guerriero et al demonstrated that 2S US was reproducible among experienced examiners for diagnosing ovarian cancer⁷⁵ and for assigning specific diagnosis of some benign ovarian cysts^{22, 10}. Similar results have been recently reported by Yazbek et al⁸⁶. This latter study also emphasized that the level

1
2
3 of confidence for providing a diagnosis affects agreement among observers. All these studies used
4
5 static 2D images for analyzing intra- and inter-observer agreement.
6
7

8
9 Regarding 3D US, only one study has assessed the reproducibility of the technique. Alcazar et
10
11 al showed that evaluation of 3D volumes from adnexal masses is reproducible among two different
12
13 observers with different level of expertise on 3D ultrasound. However, intra-observer agreement
14
15 was higher for experienced examiner¹²⁹.
16
17

18
19 Three-dimensional ultrasound allows the storage of a 3D volume of the area of interest, in this
20
21 case an adnexal mass, and an increased number of post-processing imaging tools, not available for
22
23 2D static images, that could provide a more accurate anatomic evaluation^{23,24 8}
24
25

26
27 To date, no previous study has compared the reproducibility of 2D and 3D ultrasound in
28
29 evaluating adnexal masses. In the present study we have analyzed the intra-observer and inter-
30
31 observer agreement between two experienced observers using 2D US and 3D US in making a
32
33 diagnosis of malignancy or benignity in adnexal masses. Analyzing intra-observer agreement we
34
35 could conclude that 2D US and 3D US evaluation of adnexal masses by the same examiner provide
36
37 similar diagnostic information in most of the cases when applying same criteria for malignancy
38
39 suspicion in both techniques. This may be explained by the fact that both examiners are highly
40
41 experienced in gynecologic ultrasound and adnexal masses evaluation. As a matter of fact, this may
42
43 be a bias in our study, because intra-observer reproducibility could be lower in less experienced
44
45 examiners, as shown by Yazbek et al⁸⁶
46
47
48
49

50
51 On the other hand, our data show that the agreement between observers with respect to the
52
53 characterization of an adnexal mass using 2D US images is slightly inferior to that attained with 3D
54
55 US data. We explained this result because 3D technique allows the possibility to elaborate the
56
57 acquired volumes that may display better the pathologic features of the lesion. We also believe that
58
59 the present study may be biased against 2D technique since real time scan is better than static image
60
evaluation for predicting malignancy in adnexal masses^{25 15}

1
2
3 In conclusion, our data suggest that 3D US is slightly more reproducible, but not
4 significantly, than 2D US. These findings, although need to be validated with further
5
6 investigations, seem to be relevant in order to suggest the use of 3D US in routine clinical practice
7
8 for the evaluation of adnexal masses.
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3
4
5
6
7 **Acknowledgements**
8

9 The authors are grateful to Mrs Natalia Garrofé for her secretarial support.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

References

- ~~1. Mais V, Guerriero S, Ajossa S, Angiolucci M, Paoletti AM, Melis GB. Transvaginal ultrasonography in the diagnosis of cystic teratoma. *Obstet Gynecol* 1995;85:48-52.~~
- 1 Buy JN, Ghossain MA, Hugol D, Hassen K, Scioc C, Truc JB, Poitout P, Vadrot D. Characterization of adnexal masses: Combination of color Doppler and conventional sonography compared with spectral Doppler analysis alone and conventional sonography alone. *AJR* 1996;166:385-393.
2. Alcazar JL; Laparte C, Jurado M, Lopez-Garcia G. The role of transvaginal ultrasonography combined with color velocity imaging and pulsed Doppler in the diagnosis of endometrioma. *Fertil Steril* 1997;67:487-491.
- ~~3. Pascual MA, Hereter L, Tresserra F, Carreras O, Ubeda A, Dexeus S. Transvaginal sonographic appearance of functional ovarian cyst. *Hum Reprod* 1997;12:1246-1249.~~
- 3 Brown DL, Doubilet PM, Miller FH, et al. Benign and malignant ovarian masses: Selection of the most discriminating gray-scale and Doppler sonographic features. *Radiology* 1998;208:103-110.
- 4 Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: The contribution of Doppler ultrasound. *Ultrasound Obstet Gynecol* 1999;14:338-347.
- ~~5 Guerriero S, Ajossa S, Lai MP, Mais V, Paoletti AM, Melis GB. Transvaginal ultrasonography associated with color Doppler energy in the diagnosis of hydrosalpinx. *Hum Reprod* 2000;15:1568-1572.~~
- 5 Guerriero S, Alcazar JL, Pascual MA, Ajossa S, Gerada M, Bargellini R, Virgilio B, Melis GB. Intraobserver and interobserver agreement of grayscale typical

- 1
2
3 [ultrasonographic patterns for the diagnosis of ovarian cancer. Ultrasound Med Biol.](#)
4
5
6 [2008;34:1711-6](#)
7
8
9 6 Yazbek J, Ameye L, Testa AC, Valentin L, Timmerman D, Holland TK, Van
10
11 Holsbeke C, Jurkovic D. Confidence of expert ultrasound operators in making a
12
13 diagnosis of adnexal tumor: effect on diagnostic accuracy and interobserver
14
15 agreement. *Ultrasound Obstet Gynecol.* 2010;35:89-93.
16
17
18 7 Raine-Fenning NJ, Campbell BK, Clewes JS, Kendall NR, Johnson IR. The
19
20 interobserver reliability of three-dimensional power Doppler data acquisition within
21
22 the female pelvis. *Ultrasound Obstet Gynecol* 2004;23:501-508.
23
24
25
26 ~~8 Alcazar JL, Rodriguez D, Royo P, Galvan R, Ajossa S, Guerriero S. Intraobserver~~
27
28 ~~and Interobserver reproducibility of 3-Dimensional power Doppler ultrasound indices~~
29
30 ~~in assessment of solid and cystic-solid adnexal masses. J Ultrasound Med 2008;27:16.~~
31
32
33 [8 Coyne L, Jayaprakasan K, Raine-Fenning N. 3D ultrasound in gynecology and](#)
34
35 [reproductive medicine. Women's Health \(Lond Engl\). 2008 Sep;4\(5\):501-16.](#)
36
37
38 9 Alcazar JL, Garcia-Manero M, Galvan R. Three-dimensional sonographic
39
40 morphologic assessment of adnexal masses. *J Ultrasound Med* 2007;26:1007-1011.
41
42
43
44 ~~10 Pascual MA, Carreras O, Hereter L, Gomez MT, Barri PN. Diagnóstico ecográfico~~
45
46 ~~del quiste dermoide de ovario. Prog Obst Gin 1993;36:508-512.~~
47
48
49 [10Guerriero S, Alcazar JL, Pascual MA, Ajossa S, Gerada M, Bargellini R, Virgilio B,](#)
50
51 [Melis GB. Diagnosis of the most frequent benign ovarian cysts: is ultrasonography](#)
52
53 [accurate and reproducible? J Womens Health. 2009 Apr;18\(4\):519-27.](#)
54
55
56
57 11 [Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled](#)
58
59 [disagreement or partial credit. Psychol Bull 1968;70:213-220.](#)
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- 12 [Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.](#)
- 13 [Valentin L. Use of morphology to characterize and manage common adnexal masses. *Best Pract Res Clin Obstet Gynaecol* 2004;18:71-79.](#)
- 14 [Timmerman D, Schwarzler , Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, Bourne TH. Subjective assessment of adnexal masses with the use of ultrasonography: An analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999;13:11-16.](#)
- 15 [Van Holsbeke C, Yazbek J, Holland TK, Daemen A, De Moor B, Testa AC, Valentin L, Jurkovic D, Timmerman D. Real-time ultrasound vs. evaluation of static images in the preoperative assessment of adnexal masses. *Ultrasound Obstet Gynecol.* 2008;32:828-31.](#)
- 16 ~~Pascual MA, Labastida R, Grases PJ, Carreras O, Hereter L, Dexeus S. Diagnóstico ecográfico de los quistes malignos de ovario. Estudio de las variables asociadas a malignidad mediante regresión logística no condicional. *Prog Obstet Ginecol* 1998;41:515-522.~~
- 17 ~~Pascual MA, Treserra F, Lopez L, Ubeda A, Grases PJ, Dexeus S. Role of Color Doppler Ultrasonography in the Diagnosis of Endometriotic Cyst. *J Ultrasound Med* 2000;19: 695-699.~~
- 18 ~~Pascual MA, Treserra F, Grases PJ, Labastida R, Dexeus S. Borderline cystic tumors of the ovary: Gray scale and color Doppler sonographic findings. *J Clin Ultrasound* 2002;30:76-82.~~
- 19 ~~Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213-220.~~

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- ~~20 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.~~
- ~~21 Valentin L. Pattern recognition of pelvic masses by gray scale ultrasound imaging: The contribution of Doppler ultrasound. *Ultrasound Obstet Gynecol* 1999;14:338-347.~~
- ~~22 Valentin L. Use of morphology to characterize and manage common adnexal masses. *Best Pract Res Clin Obstet Gynaecol* 2004;18:71-79.~~
- ~~23 Timmerman D, Schwarzler , Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, Bourne TH. Subjective assessment of adnexal masses with the use of ultrasonography: An analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999;13:11-16.~~
- ~~24 Guerriero S, Alcazar JL, Pascual MA, Ajossa S, Gerada M, Bargellini R, Virgilio B, Melis GB. Diagnosis of the most frequent benign ovarian cysts: is ultrasonography accurate and reproducible? *J Womens Health*. 2009 Apr;18(4):519-27.~~
- ~~25 Maymon R, Herman A, Ariely S, Dreazen E, Buckovsky I, Weinraub Z. Three-dimensional vaginal sonography in obstetrics and gynaecology. *Human Reproduction Update* 2000;6:475-484.~~
- ~~26 Coyne L, Jayaprakasan K, Raine Fenning N. 3D ultrasound in gynecology and reproductive medicine. *Women's Health (Lond Engl)*. 2008 Sep;4(5):501-16.~~
- ~~27 Van Holsbeke C, Yazbek J, Holland TK, Daemen A, De Moor B, Testa AC, Valentin L, Jurkovic D, Timmerman D. Real-time ultrasound vs. evaluation of static images in the preoperative assessment of adnexal masses. *Ultrasound Obstet Gynecol*. 2008;32:828-31.~~

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1 2 3 4 1 **Abstract**

5 6 7 2 **Purpose**

8
9
10
11 3 The purpose of this study was to compare the intra-observer and inter-observer agreement in
12
13 4 assigning a diagnosis of malignancy or benignity to adnexal masses using 2-dimensional
14
15 5 ultrasonography (2D US) and 3-dimensional ultrasonography (3D US).
16
17

18 19 6 **Methods**

20
21
22
23 7 Two experienced observers (Observers A and B) performed a retrospective review of digitally
24
25 8 stored 2D images and 3D data from a sample of 41 consecutive patients with a diagnosis of adnexal
26
27 9 mass. Each observer independently, and blinded each other, evaluated the 2D static images of each
28
29 10 adnexal mass and then the 3D volumes one-week later. The observers were required to classify the
30
31 11 adnexal lesion as benign or malignant according to pattern recognition analysis. Intra-observer and
32
33 12 inter-observer agreement were assessed by calculating the kappa index (κ).
34
35
36
37

38 39 13 **Results**

40
41
42 14 Intra-observer agreement between 2DUS and 3DUS for the observer A was 1.00 and for the
43
44 15 observer B was 0.69. Inter-observer agreement was 0.69 for 2D US and 1.00 for 3D US ($p > 0.05$)
45
46
47

48 49 16 **Conclusions**

50
51
52 17 Our results indicate that 3D US for assessing adnexal masses is reproducible, but not significantly,
53
54 18 than 2D US.
55
56

57 19 **Running Head:** Three and Two-dimensional ultrasound: intra-inter-observer agreement
58
59
60
20

1
2
3 21 **Keywords:** adnexal mass, three-dimensional ultrasound, agreement , and ultrasonography
4
5
6
7 22

9 23 INTRODUCTION

10
11
12 24 Adnexal masses encompass a wide variety of ovarian disorders including benign lesions,
13 25 malignant tumors and masses affecting the para-ovarian region. The use of 2D US allows
14
15 26 discriminating benign from malignant lesions by the subjective evaluation of the gray scale
16
17 27 ultrasound features, the so called “pattern recognition” analysis)¹⁻⁴. However, conventional two-
18
19 28 dimensional ultrasound (2D US) has several limitations. One of its major disadvantages is that it
20
21 29 relies on operator’s ability. In fact, recent studies have shown that reproducibility and confidence in
22
23 30 making a specific diagnosis of 2D US depends on examiner’s experience^{5,6}.
24
25
26
27
28
29

30 31 Three-dimensional ultrasound (3D US) has recently become available for clinical practice.
32
33 32 This technique has the potential to overcome some of the limitations of 2D US because it enables a
34
35 33 cross-sectional imaging from data obtained by a single sweep of the US beam across the involved
36
37 34 organ. Furthermore, with 2D US a “static” anatomic section is displayed on a video monitor while
38
39 35 3D US allows displaying anatomical features in planes not possible with conventional 2D US⁷.
40
41 36 Although the benefits of 3D US have been emphasized⁸, its role in assessing adnexal masses has not
42
43 37 been elucidated and more studies are required to demonstrate if it improves differentiation between
44
45 38 benign and malignant ovarian lesions and if it adds any real benefit compared to conventional 2D
46
47 39 US. Notwithstanding, before to be introduced definitively in clinical practice any diagnostic method
48
49 40 should be demonstrated as reproducible. As a matter of fact, one study has shown that 3D US seems
50
51 41 to be reproducible for assessing adnexal masses⁹. However, to the best of our knowledge, no study
52
53 42 has assessed whether 3D US is more reproducible or not than 2D US
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66

The purpose of our study was to compare the intra-observer and inter-observer agreement between two experienced observers using 2D US and 3D US for assigning a diagnosis of malignancy or benignity in the same set of adnexal masses.

MATERIALS AND METHODS

Although this is a retrospective study, Institutional review board approval was asked for and obtained before performing the study.

A retrospective review of transvaginal sonograms obtained between May 2006 and January 2007 and stored in the PACS at our institution from 41 consecutive patients diagnosed as having an adnexal mass evaluated and treated at our institution was performed. Patients with ectopic pregnancy, lack of 3D volume, absence of surgery or ultrasound follow-up and an unquestionable ultrasound diagnosis of tubal disorder were excluded.

All images were obtained with the same type of scanner (Voluson 730 Expert, GE Healthcare, Milwaukee, USA), equipped with a transvaginal multi-frequency (2.9-10 MHz) transducer. Each patient was first examined using conventional 2D US. Representative images from the adnexal mass were digitally stored. After 2D US was performed 3D volume box was activated and a 3D volume was acquired. 3D box was adjusted trying to include the whole adnexal mass within it. When the volume acquisition was completed the data file was sent using a DICOM (Digital Imaging and Communication in Medicine) connection to a personal computer. Representative 2D images and 3D volumes were selected by one investigator (BG) to be analyzed by two other observers (MAP, LH) who were blinded to 2D-image and 3D-volume selection. All selected 2D images were retrieved from the database and showed without the possibility of being modified, whereas the 3D stored volumes were analyzed off-line in a dedicated workstation by using multiplanar reformatting with the 4D-view™ software, version 5.0 (GE Healthcare, Milwaukee, USA).

Two observers, A and B (MAP and LH with, 25 and 20 years experience of gynecologic US, respectively) were required to analyze 2D images in a random order. One week after 2D images review they reviewed 3D volumes in a different order of cases than 2D images were to avoid recall bias (referencia). The observers were blinded to each other's findings, they were unaware of the diagnostic report. They were also unaware of the physical examination's findings and histopathology results. The only information they had was the age and hormonal status of the patient. For "static" 2D images only one representative gray-scale sonogram was available. 3D volumes were analyzed using multiplanar rendering (Figures 1 and 2). According to pattern recognition the observer had to provide a diagnosis of benignity or malignancy. Criteria for each type of lesion were defined in previous publications^{5,10}

Patients with suspected functional and hemorrhagic cysts were managed with follow-up scans at three and six months until spontaneous resolution. Patients with persistent or suspicious masses were submitted to surgery and mass removal with pathological confirmation (gold standard).

Statistical analysis was performed using the SAS 9.1 Institute Inc Software (Cary, NC, USA.). The intra-observer and inter-observer agreement analysis for classifying the mass as benign or malignant were evaluated with Cohen's kappa index (κ) at the 95% confidence interval (CI). A kappa value of <0 indicates no agreement, ≤ 0.20 indicates poor agreement, 0.21-0.39 indicates fair agreement, 0.40-0.59 indicates moderate agreement, 0.60-0.79 indicates substantial agreement, and 0.80-1.00 indicates almost perfect agreement^{11,12}.

Intra-observer agreement refers to the agreement between 2D US and 3D US diagnosis of each examiner. Inter-observer agreement refers to the agreement between examiners when providing a diagnosis using 2D US or 3D US. It was considered that an statistical significant existed (p value less 0.05) when 95% CI of kappa values did not overlap.

RESULTS

1
2
3 91 The mean age of patients was 38.8 ± 8.3 . Among the 41 adnexal masses included in the study,
4
5
6 92 there were 34 benign masses and 7 malignancies. Table 1 summarizes the histological diagnoses.

7
8
9 93 Thirty-one patients underwent surgery with a histopathology result. Ten patients with a
10
11 94 diagnosis of functional or hemorrhagic cyst were followed up with 2DUS at three and six months
12
13 95 that showed a resolution of the lesion. Table 2 summarizes the agreement between histopathologic
14
15 96 result and both observers (A and B)

16
17
18 97 The intra-observer agreement between 2D US and 3D US was of 1.00 (95% CI: 1.00-1.00) for
19
20
21 98 observer A and of 0.69 (95% CI: 0.35-1.00) for observer B, respectively. (Tables 3 and 4)

22
23 99 The inter-observer agreement for 2D US was 0.69 (95% CI: 0.35-1.00) and 1.00 (95% CI:
24
25
26 100 1.00-1.00) for 3D US (Tables 5 and 6). This difference was not considered statistically significant.

27 28 29 101 **Discussion**

30
31 102 The diagnostic efficacy of conventional 2D US in the diagnosis of adnexal masses is well
32
33
34 103 documented¹³. However, this technique is highly operator-dependent and based mainly on the
35
36 104 subjective assessment of mass characteristics. In spite of this fact, few studies have addressed the
37
38
39 105 issue of reproducibility of this technique for classifying adnexal masses as benign or malignant.

40
41 106 Timmerman et al showed that 2D US is reproducible among observers for providing a
42
43
44 107 diagnosis of malignancy or benignity in adnexal masses¹⁴. However, in this study anamnestic data
45
46 108 and color Doppler findings were also included in the evaluation and final diagnosis. Guerriero et al
47
48
49 109 demonstrated that 2S US was reproducible among experienced examiners for diagnosing ovarian
50
51 110 cancer⁵ and for assigning specific diagnosis of some benign ovarian cysts¹⁰. Similar results have been
52
53 111 recently reported by Yazbek et al⁶. This latter study also emphasized that the level of confidence
54
55
56 112 for providing a diagnosis affects agreement among observers. All these studies used static 2D
57
58 113 images for analyzing intra- and inter-observer agreement.

59
60 114 Regarding 3D US, only one study has assessed the reproducibility of the technique. Alcazar et
115 al showed that evaluation of 3D volumes from adnexal masses is reproducible among two different

1
2
3 116 observers with different level of expertise on 3D ultrasound. However, intra-observer agreement
4
5
6 117 was higher for experienced examiner⁹.

7
8
9 118 Three-dimensional ultrasound allows the storage of a 3D volume of the area of interest, in this
10
11 119 case an adnexal mass, and an increased number of post-processing imaging tools, not available for
12
13 120 2D static images, that could provide a more accurate anatomic evaluation⁸.

14
15
16 121 To date, no previous study has compared the reproducibility of 2D and 3D ultrasound in
17
18 122 evaluating adnexal masses. In the present study we have analyzed the intra-observer and inter-
19
20
21 123 observer agreement between two experienced observers using 2D US and 3D US in making a
22
23 124 diagnosis of malignancy or benignity in adnexal masses. Analyzing intra-observer agreement we
24
25 125 could conclude that 2D US and 3D US evaluation of adnexal masses by the same examiner provide
26
27
28 126 similar diagnostic information in most of the cases when applying same criteria for malignancy
29
30 127 suspicion in both techniques. This may be explained by the fact that both examiners are highly
31
32
33 128 experienced in gynecologic ultrasound and adnexal masses evaluation. As a matter of fact, this may
34
35 129 be a bias in our study, because intra-observer reproducibility could be lower in less experienced
36
37 130 examiners, as shown by Yazbek et al⁶

38
39
40 131 On the other hand, our data show that the agreement between observers with respect to the
41
42 132 characterization of an adnexal mass using 2D US images is slightly inferior to that attained with 3D
43
44
45 133 US data. We explained this result because 3D technique allows the possibility to elaborate the
46
47 134 acquired volumes that may display better the pathologic features of the lesion. We also believe that
48
49
50 135 the present study may be biased against 2D technique since real time scan is better than static image
51
52 136 evaluation for predicting malignancy in adnexal masses¹⁵

53
54
55 137 In conclusion, our data suggest that 3D US is slightly more reproducible, but not significantly,
56
57 138 than 2D US. These findings, although need to be validated with further investigations, seem to be
58
59 139 relevant in order to suggest the use of 3D US in routine clinical practice for the evaluation of
60
140 adnexal masses.

References

1. Buy JN, Ghossain MA, Hugol, D, Hassen K, Scioc C, Truc JB, Poitout P, Vadrot D. Characterization of adnexal masses: Combination of color Doppler and conventional sonography compared with spectral Doppler analysis alone and conventional sonography alone. *AJR* 1996;166:385-393.
2. Alcazar JL; Laparte C, Jurado M, Lopez-Garcia G. The role of transvaginal ultrasonography combined with color velocity imaging and pulsed Doppler in the diagnosis of endometrioma. *Fertil Steril* 1997;67:487-491.
3. Brown DL, Doubilet PM, Miller FH, et al. Benign and malignant ovarian masses: Selection of the most discriminating gray-scale and Doppler sonographic features. *Radiology* 1998;208:103-110.
4. Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: The contribution of Doppler ultrasound. *Ultrasound Obstet Gynecol* 1999;14:338-347.
5. Guerriero S, Alcazar JL, Pascual MA, Ajossa S, Gerada M, Bargellini R, Virgilio B, Melis GB. Intraobserver and interobserver agreement of grayscale typical ultrasonographic patterns for the diagnosis of ovarian cancer. *Ultrasound Med Biol.* 2008;34:1711-6
6. Yazbek J, Ameye L, Testa AC, Valentin L, Timmerman D, Holland TK, Van Holsbeke C, Jurkovic D. Confidence of expert ultrasound operators in making a diagnosis of adnexal tumor: effect on diagnostic accuracy and interobserver agreement. *Ultrasound Obstet Gynecol.* 2010;35:89-93.

- 1
2
3 164
4
5
6 165
7
8 166
9
10
11 167
12
13 168
14
15
16 169
17
18 170
19
20
21 171
22
23
24 172
25
26 173
27
28 174
29
30
31 175
32
33 176
34
35
36 177
37
38 178
39
40
41 179
42
43
44 180
45
46
47 181
48
49 182
50
51 183
52
53
54 184
55
56 185
57
58
59 186
60
7. Raine-Fenning NJ, Campbell BK, Clewes JS, Kendall NR, Johnson IR. The interobserver reliability of three-dimensional power Doppler data acquisition within the female pelvis. *Ultrasound Obstet Gynecol* 2004;23:501-508.
 8. Coyne L, Jayaprakasan K, Raine-Fenning N. 3D ultrasound in gynecology and reproductive medicine. *Women's Health* 2008;4:501-16.
 9. Alcázar JL, García-Manero M, Galván R. Three-dimensional sonographic morphologic assessment of adnexal masses: a reproducibility study. *J Ultrasound Med.* 2007;26:1007-11.
 10. Guerriero S, Alcazar JL, Pascual MA, Ajossa S, Gerada M, Bargellini R, Virgilio B, Melis GB. Diagnosis of the most frequent benign ovarian cysts: is ultrasonography accurate and reproducible? *J Womens Health.* 2009;18:519-27.
 11. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213-220.
 12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
 13. Valentin L. Use of morphology to characterize and manage common adnexal masses. *Best Pract Res Clin Obstet Gynaecol* 2004;18:71-79.
 14. Timmerman D, Schwarzler , Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, Bourne TH. Subjective assessment of adnexal masses with the use of ultrasonography: An analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999;13:11-16.
 15. Van Holsbeke C, Yazbek J, Holland TK, Daemen A, De Moor B, Testa AC, Valentin L, Jurkovic D, Timmerman D. Real-time ultrasound vs. evaluation of static images in

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

187 the preoperative assessment of adnexal masses. Ultrasound Obstet Gynecol.
188 2008;32:828-31.

For Peer Review

Table 1. Histologic diagnosis of adnexal masses.

Final diagnosis	N	%
Dermoid cyst	5	12.20
Endometrioma	9	21.95
Ovarian fibroma	3	7.32
Hemorrhagic cyst	5	12.20
Mucinous cystadenoma	4	9.76
Malignant tumor (Borderline tumors)	7	17.07
Paraovarian cyst	3	7.32
Functional cyst	5	12.20

Table 2 . Histopathologic result and 2D/3D-US diagnostic provided by both observers

(A and B)

	A-2D		A-3D		B-2D		B-3D	
	M	B	M	B	M	B	M	B
Malignant (M)	7	0	7	0	4	3	7	0
Benign (B)	0	34	0	34	0	34	0	34

Table 3 . Intra-observer agreement between 2D-US and 3D-US for examiner A

		3D-US	
		Benign	Malignant
2D-US	Benign	34	0
	Malignant	0	7

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 4 . Intra-observer agreement between 2D-US and 3D-US for examiner B

		3D-US	
		Benign	Malignant
2D-US	Benign	34	3
	Malignant	0	4

For Peer Review

Table 5 . Inter-observer agreement for 2D-US between examiners

		Examiner A	
		Benign	Malignant
Examiner B	Benign	34	3
	Malignant	0	4

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 6 . Inter-observer agreement for 3D-US between examiners

		Examiner A	
		Benign	Malignant
Examiner B	Benign	34	0
	Malignant	0	7

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3 Replay to the Referees.
4

5
6 Reviewer 1
7

- 8 1. needs 3-5 representative images

9 **We have added Fig 1a and b and Fig 2a and b (page 4)**
10
11

12
13 Reviewer 2.

- 14 1. excluding cases with no surgery/follow-up biases the study as far as sensitivity
15 assessment is concerned. Clinicians are most concerned with borderline findings
16 leading to a false-benign report. In this series cases missed due to underestimated
17 borderline findings and no follow-up could not be included, while the reviewer
18 judgment might have been positive. Could a linkage of the whole series with a
19 Cancer registry be attempted ?
20

21 **Indeed it is a limitation of the study. but unfortunately we cannot relate our data**
22 **set with a tumor registry**
23

- 24 2. had the two reviewers seen the study cases before ? Strange and peculiar US
25 features could have been remembered...

26 **It may be that some of the cases had been diagnosed by one of the**
27 **reviewers, however the paper material is from the years 2006-2007 and**
28 **were reviewed for the paper in 2009.**
29

- 30 3. had 3DUS been used for reporting in the reality or was only considered for review ?

31 **Some pathologies are currently performed off-line 3DUS report. In the**
32 **period in which these ovarian lesions were diagnosed (2006-2007) 3DUS**
33 **report was not implemented in our Unit.**
34

- 35 4. It would be interesting to know how the reviewer's report correlated with the
36 original. It's the original we are interested in (with other clinical information,
37 knowledge of Ca125....)

38 **Has not been correlated.**
39

- 40 5. when reviewing 3DUS was 2DUS available ? This would occur in the reality

41 **The reviewers analyzed 3DUS cases without knowledge of the diagnosis of**
42 **the lesion in 2DUS or access to the 2D images.**
43

- 44 6. did reviewers report yes or no, or did they report on a scale (e.g. 1 to 5) with further
45 grouping ? The latter would allow for studying different cut offs

46 **We evaluated two options: Benign/Malignant**
47

- 48 7. A table with final outcome (cancer = 7 – no cancer = 34) listing the original and the
49 reviewers' 2D and 3D report should be provided: the less reproducible the original
50 report, the less reliable is the review scenario. With the limitation of the considered
51 series sensitivity and PPV (original, review A and B, 2D and 3D should be provided)

52 **Table 2 is added , and text in Results (page 5).**
53

- 54 8. statistical analysis of the significance of interobserver reproducibility in 2D and 3D
55 should be provided, given the small sample size

56 **The statistical significance of interobserver reproducibility has been**
57 **evaluated with the CI of 95% k values (page 5).**
58
59
60

1
2
3 Reviewer 3

- 4 1. The conclusion in the abstract states "... 3D US ... is reproducible, even more than
5 2D US." This is repeated at the end of the paper. The authors do not show a
6 statistically significant difference in the kappas and the confidence intervals reported
7 overlap. Thus, we cannot conclude that 3D is more reproducible.

8
9 **Abstract (page 1) and Discussion (page 6) has been amended according**
10 **to the observation**

- 11 2. Please present the tables showing the data from which the kappas were computed.

12 **Tables 3-6 has been added, and text in Results (pages 5-6).**
13
14
15
16

17 Replay to the Editor:

18
19 **We have made changes according to the suggestions**
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Figure 1a. Transvaginal 2DUS showing a septate complex cyst mass, in a 48-years old women. 169x116mm (120 x 120 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

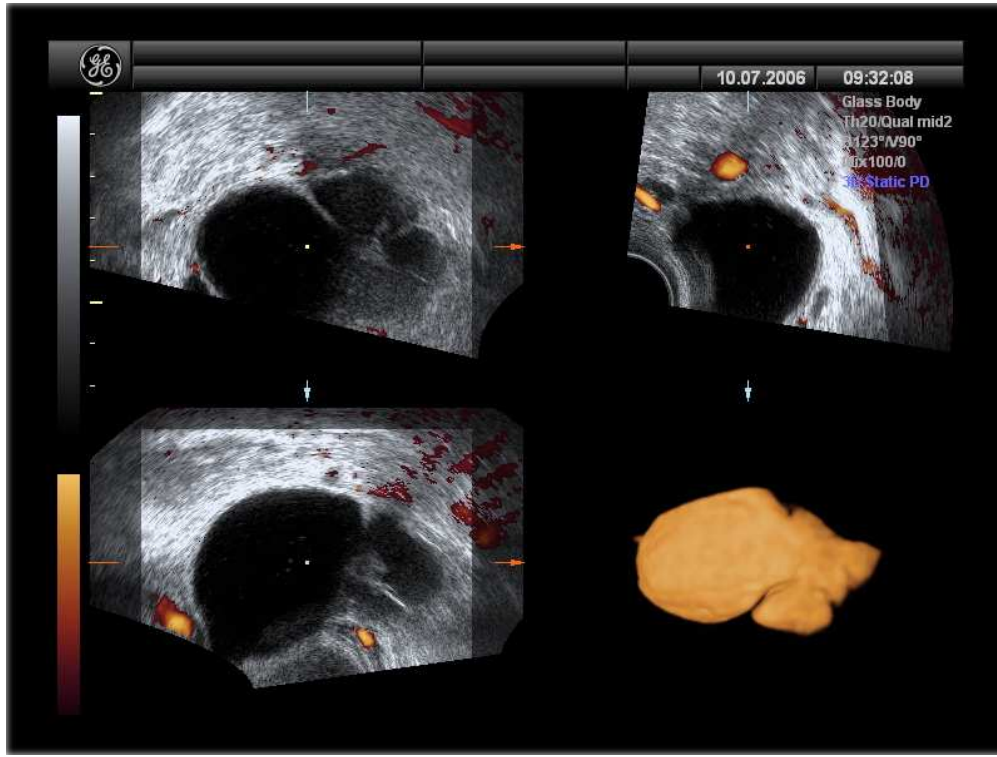


Figure 1b. Multiplanar display inverted mode of 3DUS clearly showing an hydrosalpinx 135x101mm (150 x 150 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

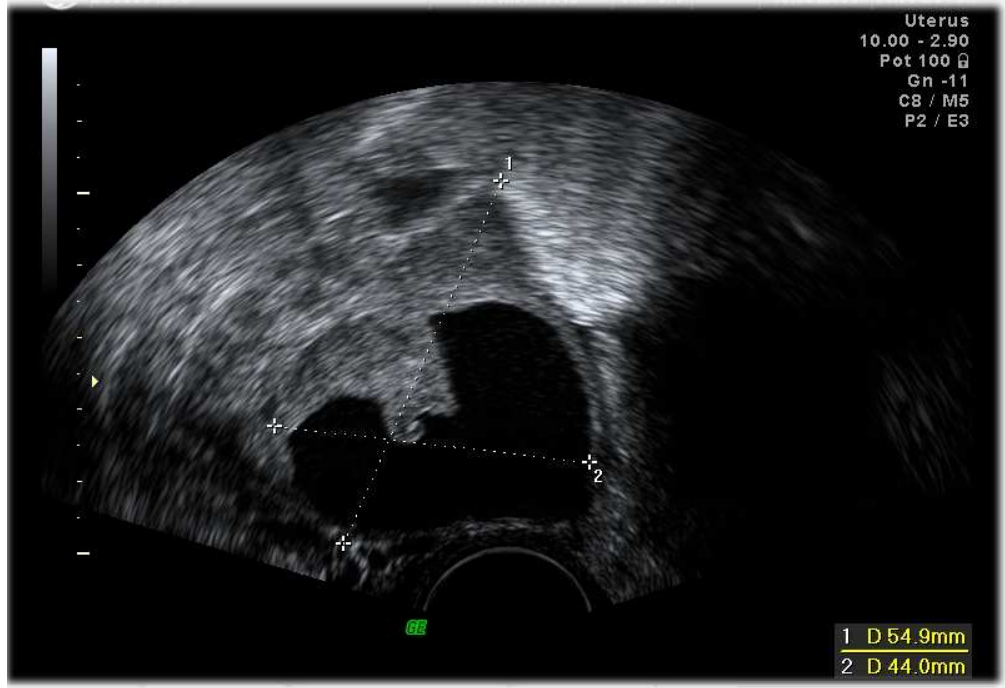


Figure 2a. 2DUS image shows a cyst with papillation in a 31-years old woman. The cyst disappeared at 3 months follow-up confirming that it was a functional- hemorrhagic cyst.
169x117mm (120 x 120 DPI)

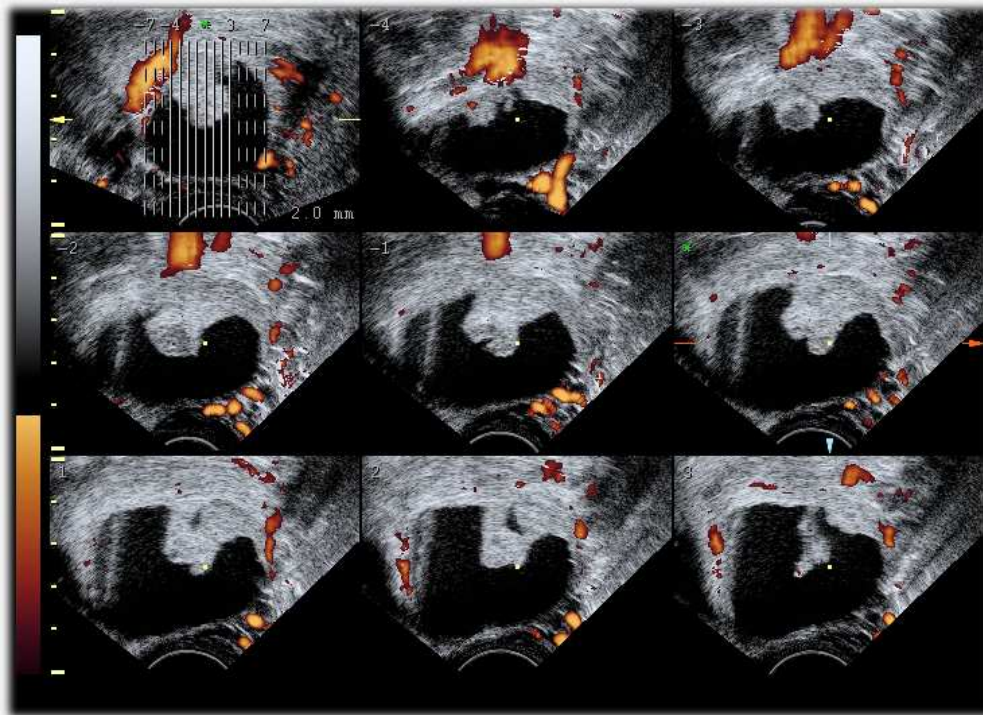


Figure 2b. 3D Power Doppler ultrasound by the Tomographic Ultrasound Imaging (TUI) allows the visualization of a complex cyst. Multiple slices above clearly show in each image completely absence of vascularization in the papillation.
125x90mm (150 x 150 DPI)