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-00632730v1

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.

Journal of Clinical Ultrasound

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

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

SCHOLARONE™
Manuscripts

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~~

For Peer Review

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]

For Peer Review

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

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.

For Peer Review

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 patter recognition the observer had to provide a diagnosis of benignity or malignancy. Criteria for each type of lesion were defined in previous publications[~~13-16~~5,10]

For Peer Review

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[17,18]. [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.**

For Peer Review

RESULTS

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

Thirty-one patients underwent surgery with a histopathology result. Ten patients with a diagnosis of functional or hemorrhagic cyst were followed up with 2DUS at three and six months that showed a resolution of the lesion. **Table 2 summarizes the agreement between histopathologic result and both observers (A and B)**

~~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.

For Peer Review

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 cancer[~~7~~5] and for assigning specific diagnosis of some benign ovarian cysts[~~22~~. 10] Similar results have bee recently reported by Yazbek et al[~~8~~6]. This latter study also emphasized that the level

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

Regarding 3D US, only one study has assessed the reproducibility of the technique. Alcazar et al showed that evaluation of 3D volumes from adnexal masses is reproducible among two different observers with different level of expertise on 3D ultrasound. However, intra-observer agreement was higher for experienced examiner[12 9].

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

To date, no previous study has compared the reproducibility of 2D and 3D ultrasound in evaluating adnexal masses. In the present study we have analyzed the intra-observer and inter-observer agreement between two experienced observers using 2D US and 3D US in making a diagnosis of malignancy or benignity in adnexal masses. Analyzing intra-observer agreement we could conclude that 2D US and 3D US evaluation of adnexal masses by the same examiner provide similar diagnostic information in most of the cases when applying same criteria for malignancy suspicion in both techniques. This may be explained by the fact that both examiners are highly experienced in gynecologic ultrasound and adnexal masses evaluation. As a matter of fact, this may be a bias in our study, because intra-observer reproducibility could be lower in less experienced examiners, as shown by Yazbek et al[8 6]

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

For Peer Review

In conclusion, our data suggest that 3D US is slightly more reproducible**, but not significantly,** than 2D US. These findings, although need to be validated with further investigations, seem to be relevant in order to suggest the use of 3D US in routine clinical practice for the evaluation of adnexal masses.

For Peer Review

Acknowledgements

The authors are grateful to Mrs Natalia Garrofé for her secretarial support.

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, Sciot 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.~~

5Guerriero S, Alcazar JL, Pascual MA, Ajossa S, Gerada M, Bargellini R, Virgilio B, Melis GB. Intraobserver and interobserver agreement of grayscale typical

For Peer Review

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.

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 Alcazar JL, Rodriguez D, Royo P, Galvan R, Ajossa S, Guerriero S. Intraobserver and Interobserver reproducibility of 3-Dimensional power Doppler ultrasound indices in assessment of solid and cystic-solid adnexal masses. J Ultrasound Med 2008;27:16.~~

8 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.

9 Alcazar JL, Garcia-Manero M, Galvan R. Three-dimensional sonographic morphologic assessment of adnexal masses. J Ultrasound Med 2007;26:1007-1011.

~~10 Pascual MA, Carreras O, Hereter L, Gomez MT, Barri PN. Diagnóstico ecográfico del quiste dermoide de ovario. Prog Obst Gin 1993;36:508-512.~~

10Guerriero 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.

11 Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or parcial credit. Psychol Bull 1968;70:213-220.

For Peer Review

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 parcial credit. Psychol Bull 1968;70:213-220.~~

For Peer Review

~~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.~~

For Peer Review

For Peer Review

Abstract

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 ($p > 0.05$)

Conclusions

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

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

For Peer Review

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)[1-4]. 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[5,6].

Three-dimensional ultrasound (3D US) has recently become available for clinical practice. This technique has the potential to overcome some of the limitations of 2D US because it enables a cross-sectional imaging from data obtained by a single sweep of the US beam across the involved organ. Furthermore, with 2D US a "static" anatomic section is displayed on a video monitor while 3D US allows displaying anatomical features in planes not possible with conventional 2D US[7]. Although the benefits of 3D US have been emphasized[8], its role in assessing adnexal masses has not been elucidated and more studies are required to demonstrate if it improves differentiation between benign and malignant ovarian lesions and if it adds any real benefit compared to conventional 2D US. Notwithstanding, before to be introduced definitively in clinical practice any diagnostic method should be demonstrated as reproducible. As a matter of fact, one study has shown that 3D US seems to be reproducible for assessing adnexal masses[9]. However, to the best of our knowledge, no study has assessed whether 3D US is more reproducible or not than 2D US

For Peer Review

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).

For Peer Review

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

For Peer Review

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

Thirty-one patients underwent surgery with a histopathology result. Ten patients with a diagnosis of functional or hemorrhagic cyst were followed up with 2DUS at three and six months that showed a resolution of the lesion**.** Table 2 summarizes the agreement between histopathologic result and both observers (A and 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 0.69 (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[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[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 cancer[5] and for assigning specific diagnosis of some benign ovarian cysts[10]. Similar results have bee recently reported by Yazbek et al[6]. This latter study also emphasized that the level of confidence for providing a diagnosis affects agreement among observers. All these studies used static 2D images for analyzing intra- and inter-observer agreement.

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

For Peer Review

observers with different level of expertise on 3D ultrasound. However, intra-observer agreement was higher for experienced examiner[9].

Three-dimensional ultrasound allows the storage of a 3D volume of the area of interest, in this case an adnexal mass, and an increased number of post-processing imaging tools, not available for 2D static images, that could provide a more accurate anatomic evaluation[8].

To date, no previous study has compared the reproducibility of 2D and 3D ultrasound in evaluating adnexal masses. In the present study we have analyzed the intra-observer and inter-observer agreement between two experienced observers using 2D US and 3D US in making a diagnosis of malignancy or benignity in adnexal masses. Analyzing intra-observer agreement we could conclude that 2D US and 3D US evaluation of adnexal masses by the same examiner provide similar diagnostic information in most of the cases when applying same criteria for malignancy suspicion in both techniques. This may be explained by the fact that both examiners are highly experienced in gynecologic ultrasound and adnexal masses evaluation. As a matter of fact, this may be a bias in our study, because intra-observer reproducibility could be lower in less experienced examiners, as shown by Yazbek et al[6]

On the other hand, our data show that the agreement between observers with respect to the characterization of an adnexal mass using 2D US images is slightly inferior to that attained with 3D US data. We explained this result because 3D technique allows the possibility to elaborate the acquired volumes that may display better the pathologic features of the lesion. We also believe that the present study may be biased against 2D technique since real time scan is better that static image evaluation for predicting malignancy in adnexal masses[15]

In conclusion, our data suggest that 3D US is slightly more reproducible, but not significantly, than 2D US. These findings, although need to be validated with further investigations, seem to be relevant in order to suggest the use of 3D US in routine clinical practice for the evaluation of adnexal masses.

For Peer Review

References

1. Buy JN, Ghossain MA, Hugol, D, Hassen K, Sciot 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.

For Peer Review

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 parcial 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

For Peer Review

the preoperative assessment of adnexal masses. Ultrasound Obstet Gynecol. 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 fybroma	3	7.32
Hemorrhagic cyst	5	12.20
Mucinous cystoadenoma	4	9.76
Malignant tumor (Borderline tumors)	7	17.07
Paraovarian cyst	3	7.32
Functional cyst	5	12.20

For Peer Review

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

For Peer Review

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

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

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

For Peer Review

Replay to the Referees.

Reviewer 1

1. needs 3-5 representative images
   We have added Fig 1a and b and Fig 2a and b (page 4)

Reviewer 2.

1. excluding cases with no surgery/follow-up biases the study as far as sensitivity assessment is concerned. Clinicians are most concerned with borderline findings leading to a false-benign report. In this series cases missed due to underestimated borderline findings an no follow-up could not be included, while the reviewer judgment might have been positive. Could a linkage of the whole series with a Cancer registry be attempted ?
   Indeed it is a limitation of the study. but unfortunately we cannot relate our data set with a tumor registry
2. had the two reviewers seen the study cases before ? Strange and peculiar US features could have been remembered…
   It may be that some of the cases had been diagnosed by one of the reviewers, however the paper material is from the years 2006-2007 and were reviewed for the paper in 2009.
3. had 3DUS been used for reporting in the reality or was only considered for review ?
   Some pathologies are currently performed off-line 3DUS report. In the period in which these ovarian lesions were diagnosed (2006-2007) 3DUS report was not implemented in our Unit.
4. It would be interesting to know how the reviewer's report correlated with the original. It's the original we are interested in (with other clinical information, knowledge of Ca125….)
   Has not been correlated.
5. when reviewing 3DUS was 2DUS available ? This would occur in the reality
   The reviewers analyzed 3DUS cases without knowledge of the diagnosis of the lesion in 2DUS or access to the 2D images.
6. did reviewers report yes or no, or did they report on a scale (e.g. 1 to 5) with further grouping ? The latter would allow for studying different cut offs
   We evaluated two options: Benign/Malignant
7. A table with final outcome (cancer = 7 – no cancer = 34) listing the original and the reviewers' 2D and 3D report should be provided: the less reproducible the original report, the less reliable is the review scenario. With the limitation of the considered series sensitivity and PPV (orginal, review A and B, 2D and 3D shoudl be provided)
   Table 2 is added , and text in Results (page 5).
8. statistical analysis of the significance of interobserver reproducibility in2D and 3D shoudl be provided, given the small sample size
   The statistical significance of interobserver reproducibility has been evaluated with the CI of 95% k values (page 5).

For Peer Review

Reviewer 3

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

   Abstract (page 1) and Discussion (page 6) has been amended according to the observation

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

   Tables 3-6 has been added, and text in Results (pages 5-6).

Replay to the Editor:

We have made changes according to the suggestions

For Peer Review

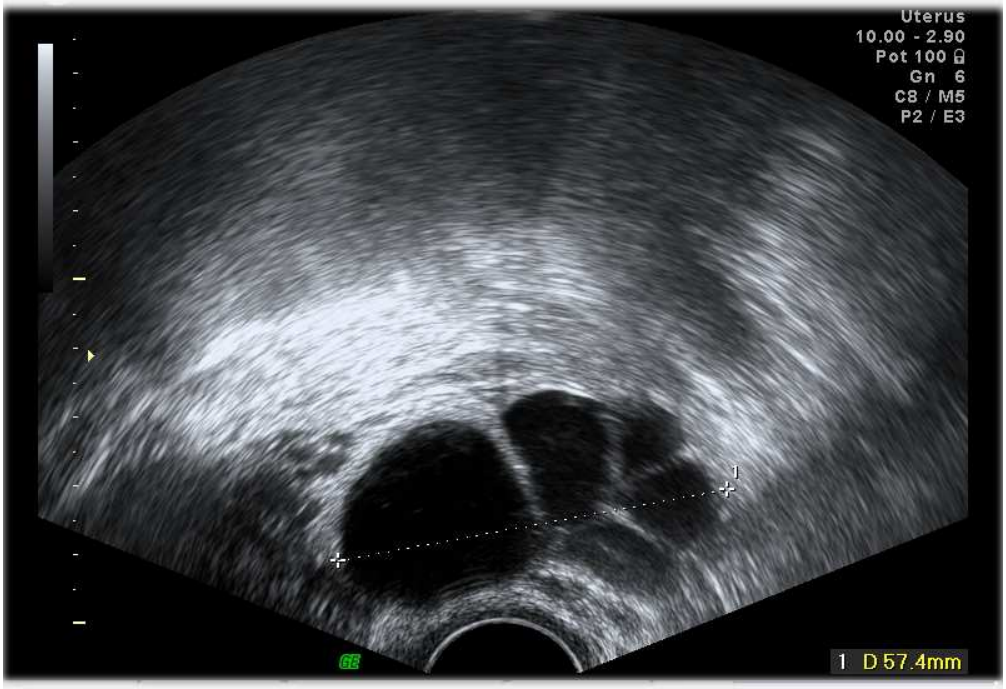

Figure 1a. Transvaginal 2DUS showing a septate complex cyst mass, in a 48-years old women.
169x116mm (120 x 120 DPI)

Review

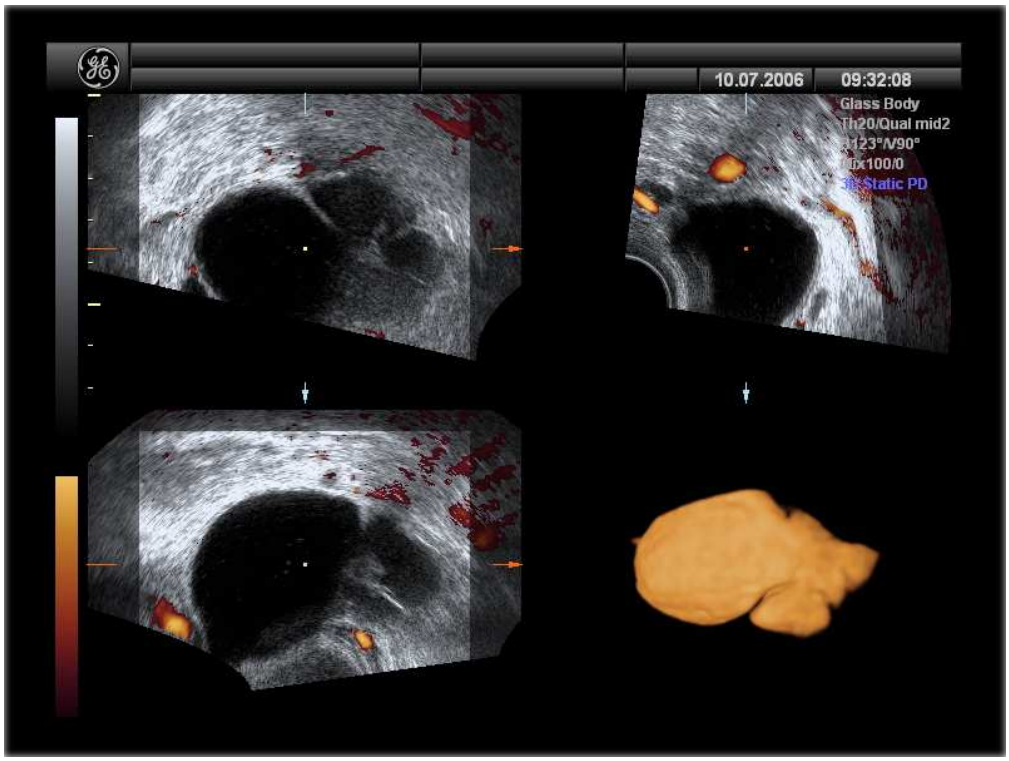

Figure 1b. Multiplanar display inverted mode of 3DUS clearly showing an hydrosalpinx
135x101mm (150 x 150 DPI)

Review

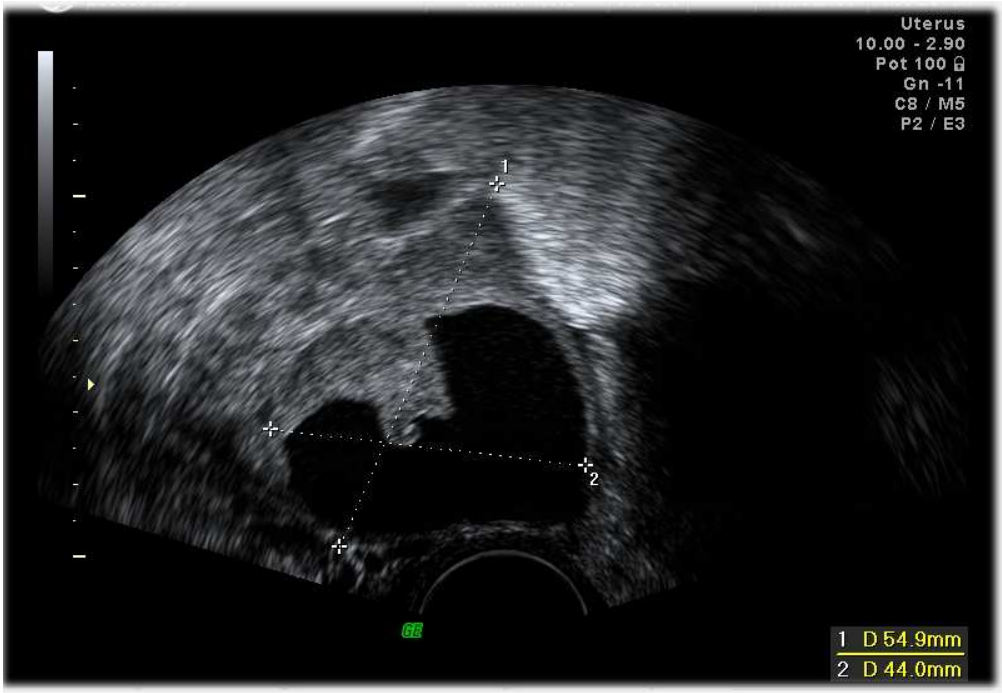

Figure 2a. 2DUS image shows a cyst with papillation in a 31-years old woman. The cyst disappeared at 3 months follow-up confirmating that it was a functional- hemorragic cyst.
169x117mm (120 x 120 DPI)

Review

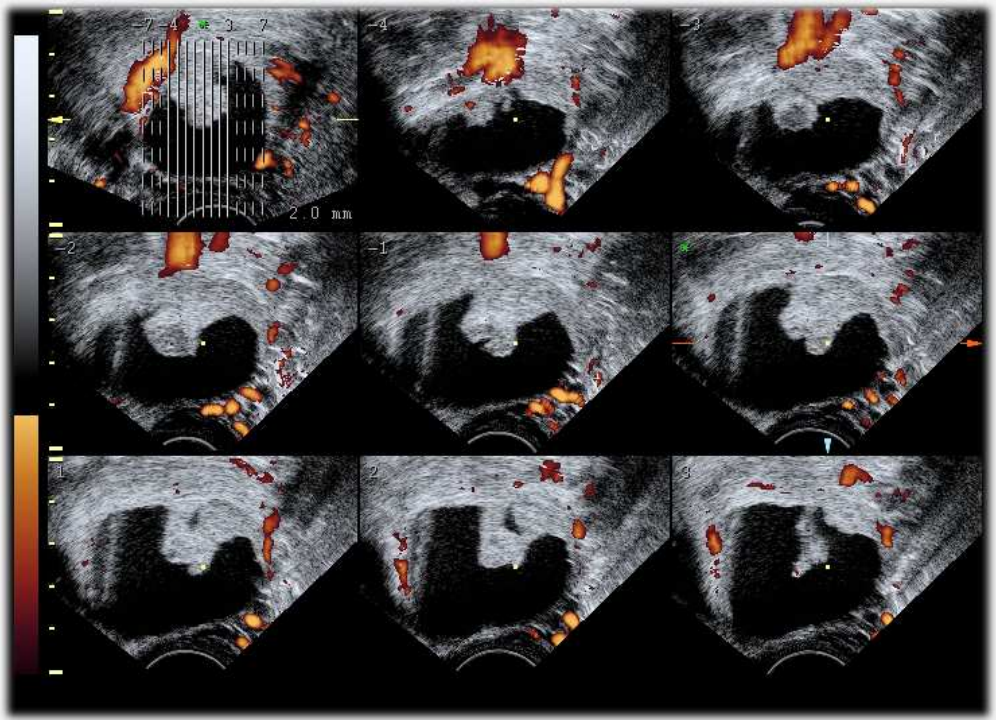

Figure 2b. 3D Power Doppler ultrasound by the Tomographic Ultrasound Imaging (TUI) allows the visualization of a complex cyst. Multiple slices above clearly show in each image completely absence of vascularization in the papillation.
125x90mm (150 x 150 DPI)

Review