

Full Length Research Paper

Study on the effect of ultrasound operators' confidence on making diagnosis of ovarian tumor

Tianying Feng^{1*}, Si Ha¹, Gaojie Hu¹, Xiaolei Wang¹ and Huang Haiyan²

¹Department of Ultrasound, The Affiliated Hospital of Inner Mongolia Medical College, Hohhot, China.

²Department of Imaging and Interventional Radiology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, N.T. Hong Kong, China.

Accepted 16 August, 2011

This study was carried out to determine the accuracy of diagnosis of ovarian tumors by different levels of confidence and experience of ultrasound operators, and the association of histological diagnosis with confidence of ultrasound operators. Ultrasound images of representative cases of benign, borderline and invasive malignant ovarian tumors were independently evaluated by 4 experts and 4 trainees. The operators categorized the ovarian tumor into benign, borderline or invasive malignant, and they also gave a score which indicated the degree of confidence when they made diagnosis. The diagnostic accuracy in differentiating benign, borderline and invasive malignant ovarian tumors were calculated depending based on the level of confidence when diagnosis. 500 images of subjects histological diagnosed with ovarian tumor were examined. The confidence score of experts was significantly higher than that of trainees ($p < 0.05$). The diagnostic accuracy was significantly increased with the raising level of confidence (test for trend, $p < 0.05$). The borderline tumor was seemed difficult to diagnose, and had the lowest accuracy, sensitivity and specificity compared to benign and primary invasive tumors. The experts showed the higher accuracy (88.5) than that of trainees (72.5). The confidence score of experts was positive related to diagnostic accuracy, with the contingency coefficient for benign, borderline and primary invasive of 0.62, 0.33 and 0.56. Our study indicated the accuracy of expert ultrasound depends on the degree of confidence, and the borderline type of ovarian tumor showed the difficult to diagnose.

Key words: Ultrasound diagnosis, diagnostic confidence, ovarian tumors, ultrasound experience.

INTRODUCTION

The prognosis of ovarian tumor is poor, it is estimated that the overall mortality rate is about 75%. But when diagnosed at stage I, 90% of them can be cured. Meanwhile, strategies for early detection require high sensitivity and extremely high specificity to attain a positive predictive value of at least 10% (Ren et al., 2010). The color doppler ultrasonographic imaging is the common method used to diagnose the ovarian cancer. Several studies have suggested that subjective evaluation by expert sinologists is superior to the use of scoring systems and mathematical models when classifying ovarian tumor into benign or malignant (Valentin et al., 2009; Valentin, 1999; Valentin et al.,

2001). Previous study suggested the sensitivity and specificity for two sonologists were about 97% and 89% (Timmerman et al., 1999), and the accuracy of diagnosis may be influenced by the confidence and experience of ultrasound observers, physical characteristics of patients and patient's ability to cooperate with examination.

The ultrasound pattern-recognition technique enables accurate differentiation between benign and malignant tumor when used by expert ultrasound operators. However, we are not aware of the level of confidence in making a diagnosis of ovarian tumor between operators and whether the diagnostic accuracy depends on the level of confidence. The aim of our study was to assess the degree of confidence with which expert ultrasound operators differentiate between benign, borderline and invasive malignant ovarian tumors, and its effect on diagnostic accuracy and agreement.

*Corresponding author. E-mail: dengmin56@yahoo.com.cn.

Table 1. Diagnostic confidence of operators.

Operators	Confidence score		
	<4(%)	4-8(%)	>8(%)
4 Experts	28(1.4)	666(33.3)	1306(65.3)
4 Trainees	120(6.0)	1202(60.1)	678(33.9)

Table 2. Diagnostic performance according to the diagnostic confidence.

Confidence	N(%)	Accuracy(%)	Sensitivity (% , 95%CI)	Specificity (% , 95%CI)
<4	19(3.8)	31.3	14.3(0.4-58.0)	44.4(13.7-78.8)
4-8	202(40.4)	83.4	74.3(36.3-96.7)	83.1(63.4-87.8)
>8	279(55.9)	94.1	96.5(76.8-99.9)	97.4(89.2-99.9)
Total	500	87.3(72.5-93.6)	83.7(74.8-90.4)	88.4(84.4-91.7)

MATERIALS AND METHODS

Study design and sample

500 images of women who were histological diagnosed with ovarian tumor in the period between October 2008 and December 2010 were included from Department of Ultrasonography of Inner Mongolia Medical College and Guangdong Medical University. Cases were selected including a mix of representative examples of benign, borderline and invasive malignant ovarian tumors.

Measurements

The color Doppler images was used for all the cases, and report which contained information on the color score by sonologist was put into an enclosed envelope. Histological diagnosis was used as the gold standard in determining the correctness of the ultrasound determination. After the images had been anonymized, they were evaluated independently by 8 operators (4 experts and 4 trainees). The operators were blinded to each other, the score of image and histological results were also blinded to them. But the clinical information of patients were available to them, including clinic symptoms, indication for the scan and staging of tumors. The operators were asked to provide the degree of confidence with which they made the diagnosis, confidence being classified into certain, probable and uncertain diagnosis. For each image, a visual scale from 1 to 10 was used to evaluate the operators' confidence in diagnosis, and the question was 'How confidence are you of your findings?'. The operators classified the mass into benign, borderline or invasive malignant ovarian tumors, and chose the predefined specific histological diagnosis.

Data analysis

StatsDirect™ was used for all statistical calculations. The overall accuracy, sensitivity, specificity, of mass images were calculated for experts and trainees, and stratification analysis was performed with regarded to different histological types. We assessed the degree of confidence of the 4 experts and 4 trainees in making diagnosis of benign, borderline or invasive malignant ovarian tumor. The diagnostic accuracy of experts and trainees in diagnosing benign, malignant borderline and invasive malignant tumors was calculated, and the relationship between the level of confidence and dialogistic

accuracy as well as histological type was also assessed.

RESULTS

The color Doppler images of 500 women diagnosed with ovarian tumors were collected. All mass images were assessed by the 8 examinations. Among these 500 cases, 255 masses (51%) were benign tumors, 130 masses(26%) were borderline ovarian tumors and 115(23%) were primary invasive ovarian tumors. The mean age of cases with ovarian tumor was 43±8 years, and was significant difference with that of borderline (45±11 years) and invasive ovarian tumors (38±9 years).

The diagnostic confidence differed significantly between experts and trainees (Table 1). The confidence score of experts was significantly higher than that of trainees. The diagnostic performance with regard to diagnostic confidence was showed in Table 2. The diagnostic accuracy showed to be low when the confidence score below 4 (accuracy =31.3%). But However, when the score was higher than 8, the total accuracy could be as high as 95%. The diagnostic accuracy significantly increased with the raising level of confidence (test for trend, $p < 0.05$). The borderline tumor showed the lowest diagnostic accuracy, sensitivity and specificity compared to benign and primary invasive tumors for operators (Table 3). The borderline tumors were more frequently as benign tumors than invasive tumors (Data not showed).

We further analyzed the correlation of the diagnostic confidence with diagnostic accuracy (Table 5). The diagnostic confidence score was positive related to diagnostic accuracy, and the contingency coefficient showed the confidence score was closely associated with benign and primary invasive tumor (contingency coefficient: 0.62, $p < 0.001$ and 0.56, $p < 0.001$, respectively), and moderately associated with borderline tumor (contingency coefficient: 0.33, $p < 0.001$) (Table 5).

Table 3. Diagnostic performance according to the histology type.

Histological type	N (%)	Accuracy	Sensitivity	Specificity
Benign	255(51)	88.5(79.7-96.9)	87.1(81.9-91.3)	90.0(85.1-93.7)
Borderline	130(26)	72.5(61.4-83.2)	64.9(55.2-73.7)	75.2(70.1-79.9)
Primary invasive	115(23)	87.3(72.5-93.6)	83.7(74.8-90.4)	88.4(84.4-91.7)

Table 4. Diagnostic performance according to the working year.

Operators	N (%)	Accuracy	Sensitivity	Specificity
4 Experts	4(50)	88.5(79.7-96.9)	87.1(81.9-91.3)	90.0(85.1-93.7)
4 Trainees	4(50)	72.5(61.4-83.2)	64.9(55.2-73.7)	75.2(70.1-79.9)

Table 5. Association of the confidence score with histology type and diagnostic accuracy.

Confidence	Accuracy		
	Benign (N,%)	Borderline (N,%)	Primary invasive (N,%)
Scores			
<4	31.2(1, 0.5%)	25 (9, 7.2%)	31.3 (8, 7.1%)
4-8	81.5(143, 56.2%)	67.5(97, 74.7%)	76.8(250, 50.0%)
>8	98.9(110, 43.3%)	90.2(23, 18.0%)	97.4(215, 42.9%)
Contingency coefficient(P value)	0.62, p<0.001	0.33, p<0.001	0.56, p<0.001

DISCUSSION

Our study has demonstrated a clear association between the level of diagnostic confidence and the diagnostic performance of the operators in determining the nature of ovarian tumors. The diagnostic accuracy of all the operators declined as the confidence in making the diagnosis decreased. The accuracy showed great difference in diagnosis of benign, borderline or primary invasive ovarian tumors, and the borderline tumor is the hardest to differentiate. Besides, the accuracy was modified by the working year and confidence score.

The knowledge and experience of the operators is usually associated with the diagnostic accuracy and certainty. The confidence is extremely difficult to assess, as it may be influenced by complex factors which interact and influence the level of confidence, such as the amount of training or working years, time allocated to diagnosis, memory and emotions, etc. (Baranski and Petrusic, 1998; Rohrbaugh and Shanteau, 1999). The results of our study indicated the diagnostic accuracy was significantly affected by the operators' working years and confidence scores. This could be explained that the confidence score may represent the training and working experience, memory and emotions, etc.

The diagnostic accuracy in this study was the lowest in the case of borderline ovarian tumors, which were most

often misclassified as benign tumors. This may be because certain types of benign and borderline ovarian tumor usually share similar morphological features, which make it difficult in distinguish between the two types of tumors. In our study, the confidence score is highly associated with benign and primary invasive tumors and moderately associated with borderline tumor. This demonstrated the morphological features could influence the confidence score and accuracy. If the borderline ovarian tumor was misclassified into other type of tumor, it may induce overtreatment or miss diagnosis to treatment delayed.

The ultrasound examination is ideally to differentiate accurately between benign, borderline and invasive malignant ovarian tumors. But in practice, most ultrasound operators classify ovarian tumor as benign or malignant and only a few operators make diagnosis of borderline ovarian tumor. Our results shows the scores of diagnosing borderline tumor is the lowest, which suggests the morphological features of the borderline ovarian tumors is hard to understand, and the accuracy of ultrasound diagnosis could be increased if the lesions were classified as invasive or non-invasive instead. Therefore, many borderline ovarian tumors received similar treatment of benign lesions, and may take a risk that the tumor content would sometimes be spilled during minimally invasive surgery, and the misdiagnosis would

have an adverse effect of short-term outcomes and low survival rates (Fauvet et al., 2005).

There were several strengths in our study. Firstly, the subjects in our study were blinded to each other of experts and trainees; therefore, the measurement bias was avoided. Secondly, our study compared the ability to differentiate and make a correct histological diagnosis of the benign, borderline and invasive malignant ovarian tumors, and we also compared the diagnostic confidence in diagnosis. These data were not conducted in China before. Thirdly, the large sample of masses increased the power of statistic.

In conclusion, our study has shown that the level of confidence is positively associated with the diagnostic performance, and the accuracy is greatly influenced by the histological type of ovarian tumor. Borderline ovarian tumor is the main source of diagnostic uncertainty. If the diagnosis is probable, the morphological description of the tumor should be accompanied by a list of possible histological diagnoses to improve the diagnostic accuracy, because the ultrasound diagnosis is influenced by the level of confidence and working experience of operator. It is reasonable that the ultrasound report should be added with the report about the ultrasound examination in making diagnosis, and this is the responsibility of the ultrasound operator for the patients' health.

REFERENCES

- Baranski JV, Petrusic WM(1998). Probing the locus of confidence judgments: experiments on the time to determine confidence. *J. Exp. Psychol. Hum. Percept. Perform.*, 24: 929-945.
- Fauvet R, Boccara J, Dufournet C, Poncelet C, Darai E (2005). Laparoscopic management of borderline ovarian tumors: results of a French multicenter study. *Ann. Oncol.*, 16:403-410.
- Ren J, Cai H, Zhang X, Liu Z, Wang JS, Hwa YL, Zhang Y, Yang Y, Li Y, Jiang SW(2010). Tumor markers for early detection of ovarian cancer. *Exp. Rev. Mole. Diag.*, 10(6): 787-798.
- Rohrbaugh C, Shanteau J (1999). *Context, process and experience: research on applied judgment and decision making*. New York: John Wiley.
- Timmerman D, Schwärzler P, Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, Bourne TH (1999). Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. *Ultrasound Obstet. Gynecol.*, 13: 11-16.
- Valentin L (1999). Prospective cross-validation of Doppler ultrasound examination and grayscale ultrasound imaging for discrimination of benign and malignant pelvic masses. *Ultrasound Obstet. Gynecol.*, 14: 273-283.
- Valentin L, Hagen B, Tingulstad S, Eik-Nes S(2001). Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. *Ultrasound Obstet. Gynecol.*, 18: 357-365.
- Valentin L, Jurkovic D, Van Calster B, Testa AC, Van Holsbeke C, Bourne T, Vergote I, Van Huffel S, Timmerman D (2009). Adding a single CA-125 measurement to ultrasound performed by an experienced examiner does not improve preoperative discrimination between benign and malignant adnexal masses. A prospective international multicentre study of 809 patients. *Ultrasound Obstet. Gynecol.*, 34(3): 345-354.