Diffusion weighted imaging for the differential diagnosis of benign vs. malignant ovarian neoplasms

Xiang-Fu Meng1*, Shi-Cai Zhu1*, Shao-Juan Sun1, Ji-Cai Guo2 and Xue Wang3

Departments of 1Radiology, 2Respiratory Medicine and 3Ultrasound, Linyi Traditional Chinese Medicine Hospital, Linyi, Shandong 276003, P.R. China

Received February 26, 2015; Accepted January 5, 2016

DOI: 10.3892/ol.2016.4445

Abstract. In order to assess the diagnostic accuracy of diffusion weighted imaging (DWI) in differentiating between benign and malignant ovarian neoplasms, a systemic meta-analysis was conducted. Relevant studies were retrieved from scientific literature databases, including the PubMed, Wiley, EBSCO, Ovid, Web of Science, Wanfang, China National Knowledge Infrastructure and VIP databases. Following a multi-step screening and study selection process, the relevant data was extracted for use in the present study. Statistical analyses were performed using Meta-disc software version 1.4 and STATA statistical software version 12.0. A total of 285 articles were retrieved from the database searches. Following a careful screening process, 10 case-control studies were selected for the present meta-analysis. The 10 studies investigated the efficacy of DWI in diagnosing ovarian neoplasms, and included a combined total of 1,159 subjects, of which 559 patients had malignant lesions and 600 had benign lesions. The results showed that the pooled sensitivity, pooled specificity, pooled positive likelihood ratio, pooled negative likelihood ratio, pooled diagnostic odds ratio (DOR) and area under the curve of the summary receiver operating characteristics curve of DWI for differentiating between benign and malignant ovarian neoplasms were 0.93, 0.89, 7.58, 0.10, 85.33 and 0.95, respectively. A subgroup analysis based on ethnicity revealed no significant difference between Asians and Caucasians. Another subgroup analysis by magnetic resonance imaging (MRI) type showed that the DORs for GE Healthcare Life Sciences and Siemens AG MRI. The DWI demonstrated an excellent diagnostic performance in discriminating between benign and malignant ovarian neoplasms, and predicted the surgical outcome in ovarian neoplasms.

Introduction

Ovarian neoplasms are the sixth most common type of cancer in terms of affected population and the seventh leading cause of cancer-associated mortality among all gynecological malignancies worldwide (1,2). Advanced ovarian neoplasms may exhibit extensive spread to the surrounding abdominal organs, including the stomach, intestine, colon, liver, lungs and pancreas, and may also promote ascites within the peritoneal cavity (3). Due to the rapid proliferation and spread of ovarian cancer within the abdominal cavity, affected patients usually undergo primary debulking surgery prior to chemotherapy with carboplatin and taxol, which may result in a disease-free survival time of 16-22 months; however, the 5-year survival rate is only 27% (4). Ovarian neoplasms have diverse morphological features and varied genetic and epigenetic alterations; consequently, distinguishing between benign lesions and malignant ovarian cancers is challenging (5). Certain evidence indicates that combination therapies are incapable of improving the clinical outcome of patients with advanced ovarian neoplasms; therefore, the early detection of ovarian malignancy is critical for patient survival (6).

Diagnostic imaging is employed to determine the suitability of a patient for surgery by assessing the primary tumor, disease volume and extent (7,8). Diffusion weighted imaging (DWI) is a non-invasive technique based on molecular water diffusion properties, and may be used to characterize tissue architecture based on microstructure, cellular density, microcirculation and cell organization (9-11). DWI has been demonstrated to be more accurate compared with computed tomography (CT) for characterizing the sites of implants of ovarian cancers, including metastases to the liver, lungs, kidneys, uterus and pancreas (12,13). In addition, the combination of DWI with gadolinium-enhanced magnetic resonance imaging (MRI), which is used for whole-body imaging of advanced ovarian cancers, was shown to increase the accuracy of determining the operability of the tumor and reduce the rate of false diagnosis, particularly for lesions that were obscured by impeded diffusion in the spleen (9).
Quantitative analysis of DWI is conducted by employing the apparent diffusion coefficients (ADCs) (14). In general, high ADC values indicate that water molecules may move freely, which suggests a low cellularity and better organization of the tissue structure, a typical feature of normal tissues. By contrast, low ADC values indicate the impeded mobility of water molecules, which suggests high cellularity, a characteristic feature of malignant lesions. Benign lesions, including simple cysts and hemangiomas, also have high ADC values due to the low cell density of the lesion (15-17). Although multiple studies have described the advantages of DWI over other techniques for the differential diagnosis of benign and malignant tumors, other reports have indicated that DWI has a comparable sensitivity and lower specificity compared with CT (18,19). In order to address this issue, a meta-analysis based on high quality published studies was conducted in the present study in order to evaluate the diagnostic value of DWI in discriminating between benign and malignant ovarian neoplasms.

Materials and methods


Selection and exclusion criteria. The published studies that were selected for the current meta-analysis fulfilled the following inclusion criteria: i) Studies are prospective or retrospective and evaluate the accuracy of DWI in discriminating between benign and malignant ovarian neoplasms; ii) studies must use histopathological results as the diagnostic gold standard; iii) studies must provide available data to calculate the sensitivity, specificity, negative likelihood ratio and positive likelihood ratio, and provide the number of ovarian neoplasm lesions; iv) studies must provide information on the type of MRI device used; and v) studies must be Chinese or English articles. The exclusion criteria were as follows: i) Studies do not conform to the inclusion criteria; ii) articles are abstracts, reviews, case reports, letters, meta-analyses or proceedings; iii) studies do not contain the number of benign or malignant neoplasms; iv) the data lacks integrity; v) publications are duplicates or studies contain overlapping data; and vi) studies are conducted in a non-human population. In addition, only the largest or most recently published study was included in instances where the author had published several studies based on one clinical dataset.

Data extraction and quality assessment. A standard data extraction form was used, and descriptive information on a range of factors was independently collected, including first author, year of submission, state, ethnicity, research design, number of lesions, gender, age and pathological types: sensitivity, specificity, accuracy, true positive, false positive, true-negative and false-negative values; and MRI apparatus types (manufacturer) and ADC threshold. Discrepancies during data extraction were resolved by reexamination of all items and discussion between reviewers. In order to determine whether the studies involved were of high quality, a quality assessment of diagnostic accuracy studies (QUADAS) analysis was used to evaluate the studies by two independent
parties (20). The 11 QUADAS items were: The spectrum of patients was representative of the patients that will receive the test in practice (QUADAS01); the selection criteria were clearly described (QUADAS02); the time period between the reference standard and index test was short enough to be reasonably sure that the target condition did not change between the two tests (QUADAS03); the whole sample or a random selection of the sample received verification using a reference standard of diagnosis (QUADAS04); patients received the same reference standard regardless of the index test result (QUADAS05); the reference standard was independent of the index test (QUADAS06); the execution of the index test was described in sufficient detail to permit replication of the test (QUADAS07); the execution of the reference
development.

Table I. Baseline characteristics of patients.

<table>
<thead>
<tr>
<th>Study author</th>
<th>Year</th>
<th>Ethnicity</th>
<th>Total</th>
<th>Benign</th>
<th>Malignant</th>
<th>Benign</th>
<th>Malignant</th>
<th>All patients</th>
<th>QUADAS score</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michielsen et al</td>
<td>2014</td>
<td>Caucasian</td>
<td>475</td>
<td>267</td>
<td>208</td>
<td>61.9±31.5</td>
<td>9</td>
<td>(9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kierans et al</td>
<td>2013</td>
<td>Caucasian</td>
<td>37</td>
<td>28</td>
<td>9</td>
<td>54.0±14.0</td>
<td>8</td>
<td>(25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Espada et al</td>
<td>2013</td>
<td>Caucasian</td>
<td>34</td>
<td>26</td>
<td>8</td>
<td>53.1±11.9</td>
<td>9</td>
<td>(12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang et al</td>
<td>2012</td>
<td>Asian</td>
<td>75</td>
<td>45</td>
<td>30</td>
<td>48.2±30.0</td>
<td>7</td>
<td>(28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cui et al</td>
<td>2012</td>
<td>Asian</td>
<td>84</td>
<td>34</td>
<td>50</td>
<td>50.8±13.1</td>
<td>7</td>
<td>(27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang et al</td>
<td>2012</td>
<td>Asian</td>
<td>202</td>
<td>74</td>
<td>128</td>
<td>56.5±15.3</td>
<td>8</td>
<td>(26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li et al</td>
<td>2012</td>
<td>Asian</td>
<td>131</td>
<td>46</td>
<td>85</td>
<td>46.2±15.5</td>
<td>7</td>
<td>(29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yuan et al</td>
<td>2011</td>
<td>Asian</td>
<td>45</td>
<td>22</td>
<td>23</td>
<td>47.8±26.5</td>
<td>7</td>
<td>(30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low et al</td>
<td>2009</td>
<td>Caucasian</td>
<td>34</td>
<td>7</td>
<td>27</td>
<td>58.5</td>
<td>9</td>
<td>(11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Han et al</td>
<td>2005</td>
<td>Asian</td>
<td>42</td>
<td>10</td>
<td>32</td>
<td>-</td>
<td>-</td>
<td>(31)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Age data are expressed as the mean ± standard deviation. Missing data are marked by ‘-‘. QUADAS, quality assessment of diagnostic accuracy studies; Ref., reference number.

Figure 3. (A) Sensitivity and (B) specificity analyses of the included studies on the diagnostic capabilities of diffusion weighted imaging for discriminating between benign and malignant ovarian neoplasms. CI, confidence interval.
standard was described in sufficient detail to permit replication (QUADAS08); the same clinical data was available at the time test results were interpreted and at the time the test was applied in practice (QUADAS09); uninterpretable test results were reported (QUADAS10); and withdrawals from the study were explained (QUADAS11).

**Statistical analysis.** The present meta-analysis was performed using STATA software (version 12.0; Stata Corp, College Station, TX, USA) and Meta-disc software (version 1.4; Meta-DiSc, Madrid, Spain). A random effects model or a fixed effects model was applied to calculate the diagnostic odds ratio (DOR). The multiple parameters of specificity, sensitivity, positive likelihood ratio, negative likelihood ratio and summary receiver operating characteristic (SROC) curve to count the area under the curve (AUC) were used to assess the diagnostic value of DWI in discriminating between benign and malignant ovarian neoplasms. A heterogeneity test, that included tests of threshold effect and non-threshold effect, was evaluated by applying the Spearman correlation coefficients (21) of the logarithms of sensitivity and 1-specificity. A non-threshold effect occurred with P>0.05 (P<0.05 was considered to indicate that a threshold effect existed). The heterogeneity test for a non-threshold effect was applied with an I² test (0%, no heterogeneity; 100%, maximal heterogeneity) (22) to reflect the potential for heterogeneity between studies. A random effects analysis was used for cases in which heterogeneity was observed between studies (P<0.05 or I²>50%); otherwise a fixed effects analysis was applied.

Data combination was conducted by applying a SROC to calculate the AUC (23), with scores that ranged from 0-1 in case of heterogeneity due to the threshold effect (an AUC of 1 was considered to be the maximal diagnostic value). If the heterogeneity was not caused by threshold effect, the targets, including specificity, sensitivity, positive likelihood ratio, negative likelihood ratio and DOR, were combined to produce

---

**Figure 4.** (A) The diagnostic OR forest plot of included studies on the diagnostic capabilities of diffusion weighted imaging in discriminating benign vs. malignant ovarian neoplasms. (B) The SROC curve of included studies on the diagnostic capabilities of diffusion weighted imaging in discriminating benign vs. malignant ovarian neoplasms. OR, odds ratio; CI, confidence interval; SROC, summary receiver operating characteristic; SENS, sensitivity; SPEC, specificity; AUC, area under the curve.
a forest plot. The studies were excluded in turn, in order to analyze the effect of single study on the overall results. Fagan’s nomogram was applied to estimate the pre-test and post-test probability of three diagnostic criteria, and a bivariate boxplot was used to judge the heterogeneity between studies. The included angle between the regression line in Deeks funnel plot (24) and DOR was applied to assess publication bias.

Results

Included studies. A total of 285 articles were retrieved from the 9 databases by keyword search. Following the exclusion of duplicates (n=89), reviews, letters or meta-analyses (n=42), non-human studies (n=53), and studies not relevant to our research topics (n=56), the remaining studies (n=45) were carefully reviewed. An additional 11 studies were excluded as they were not cohort or case-control studies (n=2), not relevant to DWI (n=4), or not relevant to ovarian neoplasms (n=5). Subsequent to additional reviewing of the remaining 34 trials, the studies that did not supply enough information were eliminated (n=4), and 10 high quality studies were included in the final meta-analysis. These 10 case-control studies, published between 2005 and 2014, provided the required information on the diagnostic value of DWI and ovarian neoplasms (9,11,12,25-31). The meta-analysis included a total of 1,159 ovarian tumor patients, comprising 559 benign tumor patients and 600 malignant tumor patients. Of the 10 studies, 6 were conducted on Asian patient populations and 4 on Caucasian patient populations. The QUADAS of included studies is shown in Fig. 1, and Table I presents the baseline characteristics of the selected study samples.

Although the majority of the studies were located in the middle region of the bivariate boxplot, two studies were located outside of the boxplot, which indicated heterogeneity between the studies (Fig. 2). The result of the Spearman correlation coefficient of the sensitivity logarithm and 1-specificity logarithm indicated that no threshold effect existed (0.345; P=0.328). The I² values of sensitivity (57.5%) and negative likelihood ratio (54.5%) were >50%, which indicated statistical heterogeneity between studies, and a random effects model was applied. However, the I² values for specificity (23.9%) and positive likelihood ratio (33.1%) were <50%, which suggested that no statistical heterogeneity existed; therefore, a fixed effects model was used.

Quantitative data synthesis. The results of the present meta-analysis demonstrated that DWI has high diagnostic capabilities in discriminating between benign and malignant ovarian neoplasms, based on the outcomes of pooled sensitivity [0.93; 95% confidence interval (CI), 0.91-0.95; Fig. 3A], pooled...
specificity (0.89; 95% CI, 0.86-0.91; Fig. 3B), pooled positive likelihood ratio (7.58; 95% CI, 6.00-9.56), pooled negative likelihood ratio (0.10; 95% CI, 0.06-0.16), pooled DOR (85.33; 95% CI, 57.15-127.40) (Fig. 4A) and the AUC of the SROC curve (0.95; Fig. 4B). The results of the pooled likelihood ratio suggested a limited clinical value of DWI for discriminating between benign and malignant ovarian neoplasms (Fig. 5A). The results of Fagan's nomogram indicated that the pre-test
probability ratio (20%) x positive likelihood ratio (8.00) yielded a post-test probability ratio of 66%, while the pre-test probability ratio x negative likelihood ratio (0.08) yielded a post-test probability ratio of 2% (Fig. 5B). Subgroup analyses of ethnicity and MRI types were conducted. The results indicated that the DORs for patients of Asian and Caucasian descent were 86.37 (95% CI, 48.35-154.27) and 84.33 (95% CI, 48.44-146.81), respectively (Fig. 6A), suggesting no statistically significant differences between Asians and Caucasians. Notably, the DORs for GE Healthcare Life Sciences and Siemens AG MRIs were 100.76 (95% CI, 65.28-155.53) and 30.85 (95% CI, 10.40-91.53), respectively; thus, the diagnostic value of GE Healthcare Life Sciences machine appears to be superior to that of the Siemens AG equipment (Fig. 6B).

Sensitivity analysis and publication bias. A sensitivity analysis was applied in order to estimate the stability of the present meta-analysis. The outcomes of the sensitivity analysis revealed that none of the included studies had an influence on the pooled DOR of the DWI for discriminating between benign and malignant ovarian neoplasms (Fig. 7A). The included angle between the regression line and the DOR axis was close to 90° in Deeks, which indicated that no publication bias was present in the current meta-analysis (Fig. 7B).

Discussion

In order to evaluate the diagnostic value of DWI in discriminating between benign and malignant ovarian neoplasms, a systematic meta-analysis was conducted. The outcomes of the present study indicated that DWI has a high diagnostic value with regard to discriminating between benign and malignant ovarian neoplasms. Ovarian neoplasms are the sixth most prevalent cancer and the seventh leading cause of cancer-associated mortality among all types of gynecological malignancy worldwide (1,2). Due to the high frequency of peritoneal and distant metastases and the complicated nature of ovarian neoplasms, the patient survival rate is poor (27,32). Diagnostic imaging is used to determine the suitability of the patient for surgery by assessing the primary tumor or by describing the disease volume and extent (7,8). DWI, as a functional MRI technique, provides non-invasive and high-resolution information on tissue characteristics based on molecular water diffusion properties, and the imaging output is useful to determine tissue microstructure, density, microcirculation and cellular organization (9,11,33). Several studies describe a close association between cell density and ADC values (16,33-35). High ADC values are present in benign lesions, including primary neoplasms that contain low cell densities, while low ADC values are typically observed in malignant neoplasms, due to high cell densities and altered membrane permeability (15,16). Based on the results of the current meta-analysis, a major conclusion is that DWI demonstrates an excellent diagnostic performance for discriminating between benign and malignant ovarian neoplasms.

Subgroup analyses based on ethnicity and MRI devices were performed. The result of the subgroup analysis on ethnicity indicated no statistically significant differences between Asians and Caucasians. However, the subgroup analysis on the MRI machine type demonstrated that a better diagnostic performance of DWI was achieved with GE Healthcare Life Sciences equipment compared with Siemens AG machines.

The present study was subject to certain limitations. First, the number of studies included was small, and several relevant unpublished studies and abstracts were not included, which may have influenced the conclusion of the current study. A second limitation of the present study was that the sample size was relatively small, leading to a lower confidence in assessing the value of DWI for discriminating between benign and malignant ovarian neoplasms. Third, only studies published in English and Chinese were included, with studies published in other languages excluded, which potentially eliminated important studies that may have influenced the conclusion of the present meta-analysis. Additional studies are required that involve a larger sample size and diverse ethnicities for a more comprehensive evaluation of the diagnostic performance of DWI in ovarian neoplasms.

In conclusion, DWI demonstrates an excellent diagnostic value for discriminating between benign and malignant ovarian neoplasms, and may provide a reliable predictive tool for the surgical outcome in ovarian neoplasms.

References


