Indirect comparison of the diagnostic performance of $^{18}$F-FDG PET/CT and MRI in differentiating benign and malignant ovarian or adnexal tumors: a systematic review and meta-analysis

Xianwen Hu 1,2, Dandan Li 3, Zhigang Liang 1, Yan Liao 1, Ling Yang 1, Rui Wang 1, Pan Wang 1 and Jiong Cai 1*

Abstract

Objective: To compare the value of fluorodeoxyglucose positron emission tomography (FDG-PET)/computed tomography (CT) and magnetic resonance imaging (MRI) in differentiating benign and malignant ovarian or adnexal tumors.

Materials and methods: English articles reporting on the diagnostic performance of MRI or $^{18}$F-FDG PET/CT in identifying benign and malignant ovarian or adnexal tumors published in PubMed and Embase between January 2000 and January 2021 were included in the meta-analysis. Two authors independently extracted the data. If the data presented in the study report could be used to construct a $2 \times 2$ contingency table comparing $^{18}$F-FDG PET/CT and MRI, the studies were selected for the analysis. The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) was used to evaluate the quality of the included studies. Forest plots were generated according to the sensitivity and specificity of $^{18}$F-FDG PET/CT and MRI.

Results: A total of 27 articles, including 11 $^{18}$F-FDG PET/CT studies and 17 MRI studies on the differentiation of benign and malignant ovarian or adnexal tumors, were included in this meta-analysis. The pooled sensitivity and specificity for $^{18}$F-FDG PET/CT in differentiating benign and malignant ovarian or adnexal tumors were 0.94 (95% CI, 0.87–0.97) and 0.86 (95% CI, 0.79–0.91), respectively, and the pooled sensitivity and specificity for MRI were 0.92 (95% CI: 0.89–0.95) and 0.85 (95% CI: 0.79–0.89), respectively.

Conclusion: While MRI and $^{18}$F-FDG PET/CT both showed to have high and similar diagnostic performance in the differential diagnosis of benign and malignant ovarian or adnexal tumors, MRI, a promising non-radiation imaging technology, may be a more suitable choice for patients with ovarian or accessory tumors. Nonetheless, prospective studies directly comparing MRI and $^{18}$F-FDG PET/CT diagnostic performance in the differentiation of benign and malignant ovarian or adnexal tumors are needed.

* Correspondence: jiong_cai@163.com
1Department of Nuclear Medicine, Affiliated Hospital of Zunyi Medical University, No.149 Dalian Road, Huichuan District, Zunyi, Guizhou Province, PR China 563003
Full list of author information is available at the end of the article

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
Keywords: Ovarian tumor, Positron emission tomography/computed tomography, Magnetic resonance imaging, Meta-analysis

Introduction

Ovarian cancer is the disease with the highest mortality rate among malignant tumors affecting the female reproductive tract. According to available statistics, more than 180 thousand women die of ovarian cancer every year worldwide [1]. Ovarian cancer is highly heterogeneous and can be classified into epithelial tumors, germ cell tumors, sex cord-stromal tumors, and other tumors [2]. Among these, epithelial ovarian cancer accounts for 90% of all cases [3]. The main risk factors for ovarian cancer are family genetic history, fertility factors, menstrual history, breastfeeding, height and body mass index, contraception, exercise, lifestyle, diet, gynecological diseases, psychological factors, and hormone replacement therapy [4–8]. Some studies suggested that smoking, a high-fat diet, ionizing radiation, talcum powder, and ABO blood type are also risk factors for ovarian cancer [9–11]. Moreover, Ness et al. have suggested that height (≥1.7 m) and body mass index (≥30 kg/m2) are also high-risk factors, while tubal ligation and short-term use of intrauterine devices can reduce the risk of ovarian cancer [10]. Moreover, Braem et al. [7] found that ovarian cancer was negatively correlated with increased parity, prolonged use of oral contraceptives, hysterectomy, younger age at natural menopause, exercise time, and annual shortening of menstrual cycles.

Ovarian cancer is an aggressive type of tumor. As no specific clinical symptoms are found in the early stage, diagnosis may be challenging. Mathieu et al. found that only 20–25% of ovarian cancer patients can be correctly diagnosed in the early stage [12]. Moreover, another study reported that about 60% of ovarian cancer patients are at an advanced stage at the time of diagnosis; these patients often have a poor prognosis and a low 5-year survival rate (below 30%) [13]. Therefore, developing a highly sensitive and specific diagnostic method may be crucial for early diagnosis, clinical staging, guiding treatment, and improving ovarian cancer prognosis.

Serological tumor markers and imaging methods, such as ultrasound, CT, MRI, and PET/CT, are the most common diagnostic approaches for ovarian cancer. Serum carbohydrate antigen (CA)125 is the most widely studied biomarker for ovarian cancer [14]. CA125 is elevated in 75% of patients with early-stage ovarian cancer, but its specificity is only 70.61% [5, 6]. CA125 is also positively expressed in other malignant tumors, including lung cancer, colorectal cancer, endometrial cancer, breast cancer, and lymphoma. Moreover, the expression level of CA125 is also increased in common pelvic benign diseases, such as adnexal cyst, endometriosis, uterine fibroids, and pelvic inflammatory disease [15]. In addition to CA125, recent studies have shown that HE4, a serum marker for ovarian cancer, has a specificity of more than 95% and a sensitivity of 70% [16, 17]. Moreover, carcinoembryonic antigen (CEA), gonadal hormone, CA72–4, CA15–3, and alkaline phosphatase have also been used as serum markers for ovarian cancer, but their sensitivity for detecting ovarian cancer is lower than 75% [17].

Ultrasound is a commonly used imaging method for gynecological diseases due to its simplicity and non-radiation exposure. However, the small size of early ovarian cancer may be limited to the ovary and does not cause ovarian morphological changes, which can lead to false-negative results. Moreover, differentiating ovarian cancer from ovarian cystadenoma, immature teratoma, and other diseases using ultrasound may be challenging [18]. On the other hand, CT and MRI can provide the anatomical information of the ovarian and its surrounding tissues, which are of great clinical significance for determining the scope of invasion of ovarian cancer and the formulation of surgery plans. MRI is a biological magnetic spin imaging technology that uses the hydrogen atoms in the human body all over the body to be excited by radio frequency pulses in an externally strong magnetic field to produce nuclear magnetic resonance. After spatial coding technology, the detector detects and receives the nuclear magnetic resonance signal emitted in the form of electromagnetic, input it into the computer, after data processing and conversion, and finally the shape of the human body tissues is formed into an image for diagnosis [19]. MRI is superior to CT in soft tissue resolution but may not be enough when detecting tumors smaller than 5 mm [20]. PET/CT imaging integrates CT and PET to achieve integration, organically integrating anatomical imaging and functional imaging, which can clearly and intuitively reflect the changes in tumor cell metabolism, so as to accurately and early diagnose tumors [21]. Most malignant tumor cells have strong metabolism and corresponding increase in energy consumption. Glucose is one of the main energy sources of tissue cells, and 18F-FDG can reflect the glucose utilization status of normal tissues of the body. Therefore, compared to MRI and CT, 18F-FDG PET/CT imaging can show both structural and functional data of the tumor and is often used to examine tumor cells at the molecular stage, which leads to positive manifestations of high metabolic uptake and early detection of
lesions [20, 21]. However, not all tumors have high radionuclide uptake, such as bronchoalveolar carcinoma, neuroendocrine tumors, colon mucinous adenocarcinoma, prostate cancer, carcinoids and so on [22]. In addition, some inflammatory lesions such as abscess, granulomatous disease, atherosclerosis, or benign tumors such as colon adenoma, uterine fibroids also have poor tracer uptake [23, 24].

In this study, we conducted a systematic review and meta-analysis on the diagnostic value of MRI and 18F-FDG PET/CT in ovarian cancer and indirectly compared the differential diagnosis performance of MRI and 18F-FDG PET/CT in ovarian benign and malignant tumors.

Material and methods
Study search strategy
This systematic review and meta-analysis were performed in accordance with PRISMA 2009 guidelines [25]. The Pubmed and Embase databases were searched for articles reporting on MRI or 18F-FDG PET/CT in ovarian cancer that were then included in the study. The following search terms were used: (“PET/CT” OR “PET-CT” OR “positron emission tomography/computed tomography” OR “positron emission tomography-computed tomography” OR “MR” OR “Magnetic Resonance”) AND (“ovarian cancer” OR “ovarian tumor” OR “ovarian neoplasms” OR “adnexal mass” OR “adnexal lesions”. Articles published in English language between January 2000 and January 2021 were included.

Two independent reviewers examined all potentially suitable articles after reading the abstract. When the results of two independent reviewers were inconsistent, a group discussion was held until a consensus was reached.

Study selection
The studies needed to meet the following inclusion criteria: (i) published between January 2000 and January 2021; (ii) prospective or retrospective studies that evaluated the accuracy of 18F-FDG PET/CT or/and MRI in differentiating benign and malignant ovarian or adnexal tumors; (iii) reference standards that at least included histopathological examination results; (iv) research data that included or that allowed to derive true positive, false positive, false negative, and true negative values based on the sensitivity, specificity, accuracy, etc. provided in the article to construct a 2 × 2 contingency table. The exclusion criteria were: (i) the sample in the study was less than 10 patients; (ii) for MRI research, the magnetic field strength was < 1.5 T or the magnetic field strength information was not recorded; (iii) for PET/CT studies, other radiotracers were used; (iv) studies in which data or data subsets were published more than once.

Data extraction and quality assessment
This meta-analysis extracted the first author, publication time, country, sample size, average age, study design type, patient selection (consecutive or nonconsecutive), true-positive (TP), false-positive (FP), false-negative (FN), true-negative (TN) results from the included studies. Other extracted data included: CT technology for PET/CT, magnetic field strength for MRI, the interval between index tests and HP, positive reference standard, the cutoff value of SUVmax for PET/CT, and ADC value for MRI of differentiating benign and malignant ovarian tumors. The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) was used for quality assessment of enrolled studies [26]. Data extraction and critical evaluation were independently carried out by two authors, if consensus could not be reached, a third reviewer was included to resolve disputes.

Statistical analyses
Stata software version 14.0 (Stata Corporation, College Station, TX, USA) was used for the statistical processing of this meta-analysis, and p < 0.05 was considered to be statistically significant. The sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and the area under the receiver operating characteristic (ROC) curves (AUC) with their 95% confidence intervals (CIs) for each individual study were calculated, according to the TP, FP, FN and TN values extracted from the enrolled study. The hierarchical logistic regression model was used to calculate general estimates of the sensitivity and specificity of the enrolled study, including the hierarchical summary receiver operating Characteristics (HSROC) model and concomitant variables. HSROC curves with 95% confidence and prediction regions were used to map the results for their sensitivity and specificity. PLR, NLR, and DOR were calculated by bivariate generalized linear mixed model and the random-effects model. Cochran's Q test and Higgins I² test were used to examine their heterogeneity [27]. In Cochran's Q test, p < 0.05 was the test standard, indicating the existence of heterogeneity. Higgins I² test was used to evaluate the degree of heterogeneity using the following criteria: inconsistency index (I²) < 50% was considered as irrelevant heterogeneity; I² = 50–80% was deemed as the possibility of moderate heterogeneity; I² > 80% suggested the possibility of significant heterogeneity. Two-sided p < 0.05 were considered as statistical significance across the included studies. The subgroup analysis of MRI and 18F-FDG PET/CT was carried out according to the sample size of the study, average age, study design type, patient selection, etc. The funnel plots and Deeks' asymmetry tests were used as the assessment of publication bias for MRI and 18F-FDG PET/CT [28].
Results

Literature search

The literature search of related subject terms initially produced 1894 articles, consisting of 1413 articles in PubMed and 481 articles in Embase. After gradually deleting overlapping, irrelevant comments, case reports/series, conferences, animal research, studies that no provided a full text, and articles not in the field of interest, 1791 articles were excluded, and the remaining 103 potentially eligible original texts were further evaluated. As not all of them were completely published in English (n = 5), it was not possible to extract sufficient data to construct a 2 x 2 contingency table (n = 14) and further exclude papers in areas of non-interest (n = 57). Finally, 27 papers on the differentiation of benign and malignant ovarian or adnexal tumors were included for meta-analysis [29–55]. The detailed process of document retrieval is shown in Fig. 1.

Study characteristics

The included 27 articles included 3730 patients with 3842 tumors, consisting of 10 PET/CT, 16 MRI, and 1 article that included PET/CT and MRI to differentiate benign and malignant ovarian or adnexal tumors. Among them, 17 studies were prospectively designed, 9 were retrospective studies, and 1 was unspecified. The sample size of the enrolled studies ranged from 30 to 1130 patients, and the average age of the patients ranged from 39.9 to 64.0 years old. For PET/CT, 7 out of the 11 studies recorded the cutoff value of the maximum standard uptake value (SUVmax) between benign and malignant tumors to study its diagnostic performance. As for MRI, 5 out of the 17 studies used the cutoff value of the apparent diffusion coefficient (ADC) value to distinguish benign and malignant ovarian tumors. In all the enrolled studies, histopathological examination was used as the reference standard, and 5 of them also included the follow-up time of at least half a year into the reference standard. The detailed characteristics of the enrolled studies and patients are summarized in Table 1, and the characteristics of PET/CT are shown in Table S1. The characteristics of MRI are shown in Table S2.

Quality assessment

The quality of all studies was considered satisfactory if it met at least 5 out of the 7 reference standards (7 reference standards include four items in the risk of bias, patient selection, index test, reference standard, flow and timing and three in application concerns, patient selection, index test, reference standard). Regarding the risk of bias for reference standards, all studies included at least histopathological examinations, which are
| Study /Year /Country | No. of patients | Mean age | Study design | Consecutive Scanner Cutoff value | Reference standard | Interval between index tests and HP | TP | FP | FN | FP |
|----------------------|----------------|----------|--------------|---------------------------------|--------------------|-----------------------------------|-----|-----|-----|-----|
| Castellucci P /2007 /Italy [29] | 50 | 64 | P | Yes | PET/CT (Non-CE) | 3.0 | HP + follow-up | ≤ 2 W | 28 | 0 | 4 | 18 |
| Kitajima K /2011 /Japan [30] | 108(111tumor) | 55.4 | NR | NR | PET/CT (Non-CE + CE) | 2.55 | HP | NR | 70 | 6 | 15 | 20 |
| Zytoon AA/2012 /Egypt [31] | 98 | 57.7 | P | yes | PET/CT (Non-CE + CE) | 4.3 | HP + follow-up | ≤ 4 W | 87 | 0 | 7 | 4 |
| Risum S/2007 /Denmark [32] | 97 | 60 | P | yes | PET/CT (Non-CE + CE) | NR | HP | ≤ 2 W | 57 | 3 | 0 | 37 |
| Tanizaki Y/2014 /Japan [33] | 160 | NR | R | NR | PET/CT (Non-CE + CE) | 2.9 | HP | NR | 54 | 5 | 13 | 88 |
| Dauwen H/2013 /Belgium [34] | 69 | 60 | P | Yes | PET/CT (Non-CE + CE) | NR | HP | ≤17d | 52 | 3 | 4 | 10 |
| Yamamoto Y/2008/Japan [35] | 30 | 47.7 | P | Yes | PET/CT (Non-CE) | 3.0 | HP | NR | 10 | 3 | 4 | 13 |
| Takagi H/2018 /Japan [36] | 76 | 59 | R | NR | PET/CT (Non-CE + CE) | 3.97 | HP | NR | 39 | 5 | 2 | 30 |
| Michielsen K/2013 /Belgium [37] | 32 | 61.9 | P | Yes | PET/CT (Non-CE + CE) | NR | HP | NR | 29 | 2 | 0 | 1 |
| Lee JW/2015/ Korea [38] | 39 | 51 | R | NR | PET/CT (Non-CE + CE) | 2.5 | HP | ≤ 7 W | 18 | 4 | 0 | 17 |
| Nam EJ/2009 /Korea [39] | 133 | 51 | P | NR | PET/CT (Non-CE + CE) | NR | HP | ≤ 4 W | 93 | 10 | 2 | 28 |
| Nam EJ/2009 /Korea [39] | 133 | 51 | P | NR | MRI (1.5 T) | NR | HP | ≤ 4 W | 69 | 7 | 3 | 6 |
| Kawahara K/2004/Japan [40] | 38 | 55.3 | P | yes | MRI (1.5 T) | NR | HP | ≤ 2 W | 21 | 2 | 2 | 13 |
| Kierans AS/2013 /USA [41] | 37 | 54 | R | Yes | MRI (1.5/3 T) | NR | HP | ≤ 137D | 6 | 3 | 3 | 25 |
| Türkoglu S/2020/Turkey [42] | 43 | 51.26 | R | Yes | MRI (1.5 T) | 0.93 | HP | ≤ 1 W | 15 | 5 | 8 | 15 |
| Michielsen K/2017/Belgium [43] | 161 | NR | P | Yes | MRI (3 T) | NR | HP | NR | 122 | 8 | 4 | 27 |
| Uehara T/2012 /Japan [44] | 50 | 51 | R | Yes | MRI (3 T) | NR | HP | NR | 18 | 3 | 1 | 28 |
| Booth SJ/2008/ UK [45] | 191 | 56 | R | NR | MRI (3 T) | NR | HP | NR | 90 | 22 | 8 | 71 |
| Shimada K/2017 /Japan [46] | 265 | NR | P | Yes | MRI (1.5 T) | NR | HP | ≤ 4 M | 52 | 18 | 2 | 193 |
| Zhang H/2019 /China [47] | 85 | 52.7 | R | Yes | MRI(1.5 T) | 1.162 | HP | NR | 51 | 3 | 11 | 20 |
| Zhang P/2012/ China [48] | 191(202tumor) | 56.5 | R | Yes | MRI (1.5 T) | 1.20 | HP | NR | 85 | 43 | 7 | 67 |
| Li W/2011/ China [49] | 127(131tumor) | B = 46.2,M = 59.9 | R | Yes | MRI (1.5 T) | 1.25 | HP | NR | 77 | 5 | 8 | 41 |
considered low-risk. Since most studies did not report the time interval between the index and reference standard tests, the risk of bias in flow and time was not assessed. Also, in two studies, the lifetime interval between the index test and the reference standard test (within 4 months and 137 days, respectively) was considered a higher risk [41, 46]. In terms of patient selection, all patients included in this study were suspected of having ovarian tumors detected by ultrasound or serum tumor markers, and the risks of publication bias and application concerns were considered low. The results of QUADAS-2 assessment are shown in Table S3.

### Diagnostic accuracy

The sensitivity of 11 studies containing $^{18}$F-FDG PET/CT methods to differentiate benign and malignant ovarian tumors ranged from 0.71 (95% CI, 0.42–0.92) to 1.0 (95% CI, 0.94–1.00), and the specificity was 0.33 (95% CI, 0.01–0.91) to 1.0 (95% CI, 0.81–1.00). Among them, a study that combined PET/CT with a Risk of Malignancy Index (RMI) based on serum CA-125, ultrasound examinations, and menopausal state showed high diagnostic value in differentiating benign and malignant ovarian tumors, with sensitivity and specificity of 1.0 (95% CI: 0.94–1.0) and 0.93 (95% CI: 0.80–0.98), respectively [32]. The pooled sensitivity and specificity of $^{18}$F-FDG PET/CT in differentiating benign and malignant ovarian tumors is 0.94 (95% CI, 0.87–0.97) and 0.86 (95% CI, 0.79–0.91), respectively, as shown in Fig. 2A. Both Cochrán’s Q test and Higgins I² test showed significant heterogeneity in sensitivity ($Q = 42.89$, $p < 0.01$; $I^2 = 74.35$) and specificity ($Q = 19.65$, $p < 0.01$; $I^2 = 70.00$).

A total of 17 studies included the diagnostic performance of MRI in diagnosing ovarian cancer, with sensitivity and specificity ranging from 0.65 (95% CI: 0.43–0.84) to 0.97 (95% CI: 0.92–0.99), and 0.46 (95% CI: 0.19–0.75) to 0.97 (95% CI: 0.94–0.99); the combined sensitivity and specificity were 0.92 (95% CI: 0.89–0.95) and 0.85 (95% CI: 0.79–0.89), respectively, as shown in Fig. 2B. Among these studies, a study based on the Ovarian-Adnexal Reporting Data System Magnetic Resonance Imaging (O-RADSMRI) score confirmed that the method could be used for risk stratification of ultrasound uncertain ovarian-adnexal masses and demonstrated high diagnostic performance [55]. Moreover, a study conducted by Zhang et al. concluded that the radiomic features extracted from MRI are highly correlated with the diagnostic accuracy, classification, and patient prognosis of ovarian cancer [47]. Also, Cochrán’s Q test and Higgins I² test showed heterogeneity between studies in sensitivity ($Q = 151.02$, $p < 0.01$; $I^2 = 85.46$) and specificity ($Q = 54.27$, $p < 0.01$; $I^2 = 70.52$).

The pooled PLR and NLR of $^{18}$F-FDG PET/CT was 6.7 (95% CI: 4.3–10.4) and 0.07 (95% CI: 0.03–0.15), respectively. As for MRI, the combined effect estimates of PLR and NLR were 6.06 (95% CI: 4.24–8.66) and 0.09 (95% CI: 0.06–0.13), respectively. The combined DOR value of ovarian tumors diagnosed by $^{18}$F-FDG PET/CT was 95 (95% CI: 41–218), and the combined DOR value for MRI was 67 (95% CI: 38–118), respectively, as shown in Table 2. There was no statistical difference between the diagnostic odds ratio of MRI compared with that of PET/CT ($p = 0.81$). The area under the SROC curve of $^{18}$F-FDG PET/CT was 0.95, with a 0.93–0.96 of 95% CI. The difference between the 95% confidence contour and the prediction contour was significant, which also indicated the heterogeneity among studies. As for MRI, the area under the SROC curve was 0.95 (95%CI: 0.93–0.97), as shown in Fig. 3.

### Table 1 The principal characteristics of eligible studies (Continued)

| Study /Year /Country | No. of patients | Mean age | Study design | Consecutive | Scanner | Cutoff value | Reference standard | Interval between index tests and HP | TP | FP | FN | FP |
|----------------------|-----------------|----------|--------------|-------------|---------|-------------|---------------------|---------------------------------|----|----|----|----|
| **Fan X/2015/ China [50]** | 64 | 46.7 | R | NR | MRI (3 T) | 0.878 | HP | NR | 54 | 5 | 4 | 25 |
| **Schaib SA/2003/ UK [51]** | 104 (155 tumor) | 50 | P | Yes | MRI (1.5 T) | NR | HP | NR | 61 | 11 | 3 | 80 |
| **Gity M/2019/ Iran [52]** | 43 (49 tumor) | 39.9 | P | Yes | MRI (3 T) | NR | HP | NR | 21 | 9 | 1 | 18 |
| **Pereira PN/2018/ Brazil [53]** | 200 (237 tumor) | B = 47; M = 57.8 | P | Yes | MRI (1.5 T) | NR | HP + follow-up | NR | 75 | 4 | 4 | 154 |
| **Van TP/2007/ UK [54]** | 76 | B = 46; M = 57 | P | Yes | MRI (1.5 T) | NR | HP | NR | 23 | 7 | 2 | 44 |
| **Thomassin N/2020/France [55]** | 1130 | 49 | P | Yes | MRI (1.5/3 T) | NR | HP + follow-up | NR | 189 | 79 | 14 | 848 |

B Beligin, M Malignant, HP Histopathology, P Prospective, R Retrospective, NR Not report, W Week, M Month, D Day, MRI Magnetic resonance imaging, PET/CT Positron emission computer/Computed tomography, T Tesla

$^{18}$F-FDG PET/CT and MRI are two commonly used imaging modalities in the diagnosis and management of ovarian cancer. The sensitivity and specificity of $^{18}$F-FDG PET/CT in differentiating benign and malignant ovarian tumors ranged from 0.71 (95% CI, 0.42–0.92) to 1.0 (95% CI, 0.94–1.00), and the specificity was 0.33 (95% CI, 0.01–0.91) to 1.0 (95% CI, 0.81–1.00). Among them, a study that combined PET/CT with a Risk of Malignancy Index (RMI) based on serum CA-125, ultrasound examinations, and menopausal state showed high diagnostic value in differentiating benign and malignant ovarian tumors, with sensitivity and specificity of 1.0 (95% CI: 0.94–1.0) and 0.93 (95% CI: 0.80–0.98), respectively [32]. The pooled sensitivity and specificity of $^{18}$F-FDG PET/CT in differentiating benign and malignant ovarian tumors is 0.94 (95% CI, 0.87–0.97) and 0.86 (95% CI, 0.79–0.91), respectively, as shown in Fig. 2A. Both Cochrán’s Q test and Higgins I² test showed significant heterogeneity in sensitivity ($Q = 42.89$, $p < 0.01$; $I^2 = 74.35$) and specificity ($Q = 19.65$, $p < 0.01$; $I^2 = 70.00$).

A total of 17 studies included the diagnostic performance of MRI in diagnosing ovarian cancer, with sensitivity and specificity ranging from 0.65 (95% CI: 0.43–0.84) to 0.97 (95% CI: 0.92–0.99), and 0.46 (95% CI: 0.19–0.75) to 0.97 (95% CI: 0.94–0.99); the combined sensitivity and specificity were 0.92 (95% CI: 0.89–0.95) and 0.85 (95% CI: 0.79–0.89), respectively, as shown in Fig. 2B. Among these studies, a study based on the Ovarian-Adnexal Reporting Data System Magnetic Resonance Imaging (O-RADSMRI) score confirmed that the method could be used for risk stratification of ultrasound uncertain ovarian-adnexal masses and demonstrated high diagnostic performance [55]. Moreover, a study conducted by Zhang et al. concluded that the radiomic features extracted from MRI are highly correlated with the diagnostic accuracy, classification, and patient prognosis of ovarian cancer [47]. Also, Cochrán’s Q test and Higgins I² test showed heterogeneity between studies in sensitivity ($Q = 151.02$, $p < 0.01$; $I^2 = 85.46$) and specificity ($Q = 54.27$, $p < 0.01$; $I^2 = 70.52$).

The pooled PLR and NLR of $^{18}$F-FDG PET/CT was 6.7 (95% CI: 4.3–10.4) and 0.07 (95% CI: 0.03–0.15), respectively. As for MRI, the combined effect estimates of PLR and NLR were 6.06 (95% CI: 4.24–8.66) and 0.09 (95% CI: 0.06–0.13), respectively. The combined DOR value of ovarian tumors diagnosed by $^{18}$F-FDG PET/CT was 95 (95% CI: 41–218), and the combined DOR value for MRI was 67 (95% CI: 38–118), respectively, as shown in Table 2. There was no statistical difference between the diagnostic odds ratio of MRI compared with that of PET/CT ($p = 0.81$). The area under the SROC curve of $^{18}$F-FDG PET/CT was 0.95, with a 0.93–0.96 of 95% CI. The difference between the 95% confidence contour and the prediction contour was significant, which also indicated the heterogeneity among studies. As for MRI, the area under the SROC curve was 0.95 (95%CI: 0.93–0.97), as shown in Fig. 3.
Publication bias

Deeks et al.’s funnel plot for publication bias for MRI and $^{18}$F-FDG PET/CT is shown in Fig. 4. The $p$ values of the slope coefficients were all greater than 0.05 (for PET/CT, $p = 0.52$, for MRI, $p = 0.08$), indicating that the possibility of publication bias between studies was low.

Exploration of heterogeneity

The results of the meta-regression analysis are summarized in Table 3. For both $^{18}$F-FDG PET/CT and MRI studies in the differentiation of benign and malignant ovarian tumors, the results of meta-regression analysis showed that the type of study design (prospective vs. retrospective) was a factor affecting heterogeneity ($p<0.01$). Specifically, for PET/CT, the sensitivity of prospective research design was higher than in retrospective research ($0.95$ [95%CI: 0.91–0.99] versus 0.94 [95%CI:0.85–1.00]), but the specificity was lower than retrospective research ($0.86$ [95%CI:0.77–0.94] versus 0.89 [95%CI:0.81–0.98]). As for MRI, the sensitivity and specificity of studies designed for prospective studies were higher than those in retrospective studies, which were $0.95$ (95%CI:...
In addition to the study design, the average age of patients between studies using PET/CT to differentiate benign and malignant ovarian tumors also showed heterogeneity. The sensitivity and specificity of the study with the average age of the enrolled patients older than 60 years old were higher than those in the group with the average age of enrolled patients younger than 60 years old, which were 0.97 (95%CI: [0.93–1.00]) versus 0.93 (95%CI: [0.87–0.99]) and 0.89(95%CI: [0.82–0.96]) versus 0.80(95%CI:0.73–0.87)), respectively, and the difference was statistically significant (p<0.01). Otherwise, the sample size, nor the use of CT enhancement technology affected the heterogeneity between PET/CT studies, with p values of 0.93 and 0.21, respectively.

In addition, in terms of the magnetic field strength of MRI, the sensitivity of using 3.0 T was higher than 1.5 T (0.95 [95%CI:0.91–0.98] vs 0.91 [95%CI:0.88–0.94]), but the specificity was lower (0.80 [95%CI:0.68–0.92] vs 0.86 [95%CI:0.81–0.92]). However, the difference was not statistically significant (p = 0.22). Also, the number of imaging planes (2 or 3) was not a factor affecting the accuracy of MRI diagnosis (p = 1.0).

Discussion

The current study evaluated the diagnostic performance of MRI and 18F-FDG PET/CT in differentiating benign and malignant ovarian or adnexal tumors. Our results showed that MRI and PET/CT both had high and similar sensitivity and specificity in the diagnosis of ovarian cancer. All studies included patients at risk; some studies also included patients confirmed with ovarian or appendage masses through ultrasound examination or elevated serum marker CA125. It should be pointed out that the study of Lee et al [38] contains a large number of ovarian lesions from other tumor sources, so we excluded the data of patients with ovarian metastases when performing this meta-analysis.

18F-FDG PET/CT and MRI data showed significant heterogeneity in the pooled sensitivity and specificity results. According to the results of meta-regression analysis, the statistically significant factors that caused the heterogeneity between PET/CT studies may be attributed to the type of study design and the average age of the enrolled patients. Specifically, retrospective studies showed lower sensitivity and higher specificity than prospective studies, which may be related to the small number of retrospective studies [33, 36, 38]. Moreover, the sensitivity and specificity were higher when examining patients older than 60 years old compared to studies that included patients younger than 60 years old; yet, the reason for this remains unclear. As for MRI, similar conclusions were drawn, i.e., the sensitivity and specificity of prospective design research were higher than...
retrospective research. When performing retrospective studies, doctors cannot always obtain enough clinical data, which may result in lower sensitivity and specificity.

Our meta-regression analysis showed that the use of enhanced CT technology and low-dose CT was not a heterogeneous factor affecting diagnostic performance. Therefore, from the perspective of patients receiving radiation doses, the use of enhanced CT technology needs to be reconsidered in future research. Also, as for MRI, further studies may be needed to truly determine the added value of DWI sequences in identifying benign and malignant ovarian tumors because meta-regression analysis showed no statistical difference between using DWI sequences and not using DWI sequences. It is worth noting that in both PET/CT and MRI studies, only two studies included follow-up as the reference standard. Therefore, studies using only pathological biopsy and combined pathological biopsy and follow-up time as reference standards were not included in meta-regression analysis.

Table 3 The results of meta-regression analysis of PET/CT and MRI to differentiate benign and malignant ovarian tumors

| Covariates          | Subgroup          | No. of studies | Sensitivity(95%CI) | Specificity(95%CI) | P     |
|---------------------|-------------------|----------------|-------------------|-------------------|-------|
| **18F-FDG PET/CT**  | Study design      |                |                   |                   |       |
|                     | Prospective       | 7              | 0.95 [0.91–0.99]  | 0.86 [0.77–0.94]  | <0.01 |
|                     | Retrospective     | 3              | 0.94 [0.85–1.00]  | 0.89 [0.81–0.98]  |       |
|                     | CT technique      |                |                   |                   |       |
|                     | Without enhanced CT | 5          | 0.96 [0.92–1.00]  | 0.81 [0.73–0.90]  | 0.21  |
|                     | With enhanced CT  | 6              | 0.91 [0.83–0.99]  | 0.91 [0.85–0.96]  |       |
|                     | Sample size       |                |                   |                   |       |
|                     | >50               | 7              | 0.94 [0.89–0.99]  | 0.87 [0.80–0.94]  | 0.93  |
|                     | ≤50               | 4              | 0.94 [0.86–1.00]  | 0.84 [0.71–0.97]  |       |
|                     | Mean age          |                |                   |                   |       |
|                     | ≥60               | 4              | 0.97 [0.93–1.00]  | 0.89 [0.82–0.96]  | <0.01 |
|                     | <60               | 6              | 0.93 [0.87–0.99]  | 0.80 [0.73–0.87]  |       |
| **MRI**             | Study design      |                |                   |                   |       |
|                     | Prospective       | 9              | 0.95 [0.93–0.97]  | 0.86 [0.80–0.93]  | <0.01 |
|                     | Retrospective     | 8              | 0.88 [0.85–0.92]  | 0.83 [0.75–0.91]  |       |
|                     | Magnetic field strength |          | 0.91 [0.88–0.94]  | 0.86 [0.81–0.92]  | 0.22  |
|                     | With 1.5 T        | 12             | 0.95 [0.91–0.98]  | 0.80 [0.68–0.92]  |       |
|                     | Only 3.0 T        | 5              | 0.95 [0.91–0.98]  | 0.80 [0.68–0.92]  |       |
|                     | Scan sequence     |                |                   |                   |       |
|                     | With DWI          | 11             | 0.91 [0.88–0.95]  | 0.87 [0.82–0.93]  | 0.16  |
|                     | Without DWI       | 6              | 0.91 [0.91–0.98]  | 0.78 [0.67–0.90]  |       |
|                     | Sample size       |                |                   |                   |       |
|                     | >50               | 12             | 0.93 [0.91–0.96]  | 0.85 [0.79–0.91]  | 0.09  |
|                     | ≤50               | 5              | 0.85 [0.76–0.94]  | 0.83 [0.72–0.95]  |       |
|                     | No. of imaging planes |          | 0.92 [0.92–0.92]  | 0.83 [0.83–0.83]  | 1.00  |
|                     | 3                 | 7              | 0.93 [0.93–0.93]  | 0.85 [0.85–0.85]  |       |

18F-FDG Fluorine-18 labeled deoxyglucose, PET Positron emission computer, CT Computed tomography, MRI Magnetic resonance imaging, DWI Diffusion weighted imaging, CI Confidence interval; *As for MRI, the mean age of all patients with ovarian tumors enrolled in the study was less than 60 years old, so they were not included in the subgroup analysis.
The results of previous meta-analysis studies have shown that PET/CT has good diagnostic performance in ovarian cancer distant metastasis and prognostic evaluation [56–58]. Specifically, the meta-analysis results of Han et al. showed that the pooled sensitivity and specificity of 18F-FDG PET/CT in identifying distant metastases of ovarian cancer was 0.72 (95% CI: 0.61–0.81) and 0.93 (95% CI: 0.85–0.97), respectively. Among them, the pooled sensitivity and specificity of PET/CT in the diagnosis of retroperitoneal lymph node metastasis was 0.77 (95% CI: 0.61–0.87) and 0.97 (95% CI: 0.93–0.99), respectively [56]. Meanwhile, another study by Han et al. showed that PET/CT has good diagnostic performance in ovarian cancer distant metastasis and prognostic evaluation [56–58]. Specifically, the meta-analysis results of 64 studies with 3722 patients, Xu et al. showed that the pooled sensitivity and specificity of PET/CT and PET for recurrent/metastatic ovarian cancer were 0.92 (95% CI: 0.90–0.93) and 0.91 (95% CI: 0.89–0.93), respectively [58]. Our meta-analysis included 11 PET/CT studies that differentiated benign and malignant ovarian or adnexal tumors with a good diagnostic performance of PET/CT in ovarian cancer [29–39]. However, the main purpose of our study was to compare the diagnostic performance of PET/CT and MRI in differentiating benign and malignant primary ovarian tumors. Our results showed that both methods had good diagnostic performance; thus, both methods should be recommended in clinical practice. Compared with PET/CT, MRI can shorten the examination time and lower medical costs [59]. E.g., if we consider GE or Siemens MRI with a magnetic field strength of 1.5 T; when doing lower abdomen and pelvic examinations, the scanning time of conventional T1WI, T2WI sequence plus DWI sequence is less than 10 min; the time for a PET/CT examination is often more than 15 min. In China, the cost of an MRI is significantly lower than that of PET/CT. Also, MRI does not produce ionizing radiation and is commonly recommended for female patients (protecting breasts and ovaries sensitive to radiation) [60]. On the other hand, PET/CT has good diagnostic performance in identifying benign and malignant ovarian tumors. PET/CT may also detect ovarian cancer metastases at the same time. PET/CT is recommended for ovarian cancer staging and treatment evaluation. Therefore, future prospective studies comparing whole-body MRI and PET/CT in the staging of ovarian cancer are warranted. Since the current study only discussed the diagnostic value of MRI and PET/CT in distinguishing benign and malignant ovarian tumors, based on the above advantages of MRI, we believe that MRI may be more suitable as an auxiliary examination method for the differentiation of benign and malignant ovarian tumors.

The main limitation of the present study was that it indirectly compared the diagnostic accuracy of PET/CT and MRI. Different methods and different characteristics of patients were included in those studies, resulting in great heterogeneity in the estimation of diagnostic accuracy, limiting the quality of this meta-analysis. Secondly, in both PET/CT and MRI studies, inconsistent interpretation of the results was also a major drawback. For instance, some studies classified borderline ovarian tumors as benign tumors [31, 32], but some studies classified them as malignant tumors [30, 33, 35, 39, 46, 52, 54]. Moreover, most of the studies included in the analysis did not describe the quantitative data of ovarian tumors, such as the average size of benign and malignant tumors, the SUVmax value for the PET/CT study, the ADC value for the MRI study, which limits further subgroup analysis research. Nonetheless, the study with a large sample size indirectly compared MRI with 18F-FDG PET/CT still provides a reference for the differential diagnosis of benign and malignant ovarian or adnexal tumors.

Conclusion
MRI and 18F-FDG PET/CT showed to have high and similar diagnostic performance in the differential diagnosis of benign and malignant ovarian or adnexal tumors. MRI is a promising non-radiation imaging technology, which may be a more favorable choice for patients with ovarian or accessory tumors. Prospective studies directly comparing MRI and 18F-FDG PET/CT diagnostic performance in the differentiation of benign and malignant ovarian or adnexal tumors are needed in the future.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12885-021-08815-3.

Additional file 1 Table S1. FDG PET/CT characteristics. Table S2. MRI characteristics. Table S3. Risk of bias and application concerns for included studies were assessed by the QUADAS-2 tool.

Acknowledgements
None.

Authors’ contributions
Xianwen Hu performed investigation, methodology, visualization, and writing-first draft; Jiong Cai did investigation and research, methodology, writing-review and editing; Dandan Li did formal analysis and software operations; Yan Liao and Zhigang Liang did research methods and essay writing; Ling Yang and Rui Wang did data curation; Pan Wang participated in the review and revision of the paper; All authors reviewed the manuscript. The author(s) read and approved the final manuscript.
