ABSTRACT  
Objective: To compare the diagnostic accuracy of different hormonal biomarkers and to find the most effective hormonal biomarker for the diagnosis of endometriosis.

Data Sources: We conducted a systematic search using PubMed, EMBASE, Cochrane Library, and China Biomedical Literature to identify relevant studies from the first day of databases to August 2018.

Methods of Study Selection: Two independent reviewers screened for study eligibility and extracted data. Random controlled trials, cross-sectional studies, case-control studies, and cohort studies evaluating the diagnostic accuracy of hormonal markers for endometriosis were included.

Tabulation, Integration, and Results: We included 17 studies that involved 1279 participants and evaluated 7 hormonal biomarkers. The pooled sensitivity and specificity in endometriosis were .79 (.71, .86) and .89 (.82, .94) for aromatase, .30 (.18, .46) and .80 (.65, .90) for human chorionic gonadotropin/luteinizing hormone receptor, .75 (.66, .83) and .47 (.34, .60) for estrogen receptor (ER)-α, .65 (.56, .74) and .68 (.55, .80) for ER-β, .45 (.38–.52) and .92 (.85–.97) for serum prolactin, .69 (.51, .83) and .30 (.16, .49) for estrogen sulfotransferase, and .73 (.60–.84) and .48 (.33–.63) for 17β-hydroxysteroid dehydrogenase type 2 (17βHSD2). Compared with human chorionic gonadotropin/luteinizing hormone receptor, ER-α, ER-β, estrogen sulfotransferase, and 17βHSD2, aromatase had a higher sensitivity, specificity, positive likelihood ratio, and diagnostic odds ratio. The specificities of aromatase and serum prolactin were comparable, but the sensitivity, positive likelihood ratio, and positive likelihood ratio of serum prolactin were much lower than that of aromatase.

Conclusion: Aromatase may be an excellent diagnostic test for endometriosis. However, because of the moderate quality of the included studies and the limited sample size, this result requires more research to validate. (PROSPERO registration number: PROSPERO 2018 CRD42018105126.) Journal of Minimally Invasive Gynecology (2019) 26, 1026–1035. © 2019 AAGL. All rights reserved.

Keywords: Endometriosis; Hormonal biomarkers; Diagnostic test accuracy; Meta-analysis; Adjusted indirect comparison
endometriosis [8,9]. This approach has a number of disadvantages, including but not limited to high cost, the need for general anesthesia, and the likelihood of postoperative adhesion formation [10]. In addition, only a third of women who undertake a laparoscopic procedure will receive a diagnosis of endometriosis; therefore, many disease-free women are unnecessarily exposed to surgical risk [10,11]. Fortunately, in the past few years noninvasive or minimally invasive methods such as imaging and biomarkers have been at the forefront of ongoing study [9].

Cytochrome P450 aromatase (aromatase), the product of the CYP19A1 gene, is a microsomal enzyme that converts C19 androgens, such as testosterone or androstenedione, into C18 estrogens, estradiol (E₂) or estrone, respectively [12,13]. It is expressed in several tissues including the ovaries, testes, placenta, adipose tissue, and bone osteoblasts [14,15]. A previous study had suggested that detection of aromatase in endometrial biopsies could be used as an outpatient screening test for endometriosis [16]. Luteinizing hormone (LH) and human chorionic gonadotropin (HCG) are glycoprotein hormones that bind to G protein–coupled membrane receptors, resulting in activation of adenylate cyclase [17,18]. HCG/LH receptor levels have been found to be elevated in many gynecologic diseases, such as endometrial carcinomas, choriocarcinomas, and adenomyosis [19–21]. Estrogen receptors alpha (ER-α) and beta (ER-β) are transcription factors involved in the regulation of many complex physiologic processes in humans [22]. The expression patterns of ER-α and ER-β in endometriotic lesions are different from those in the eutopic endometrium [23]. 17β-Hydroxysteroid dehydrogenase type 2 (17βHSD2) and estrogen sulfotransferase (EST) are considered to be important enzymes in estrogen metabolism of endometrium [24,25]. They are altered in both eutopic and ectopic endometrium (endometriosis) of patients with endometriosis [26]. Prolactin is an anterior pituitary hormone that functions in lactation and pregnancy [27]. Studies have shown that serum prolactin levels were significantly elevated in infertile patients with endometriosis [28,29].

Three Cochrane systematic reviews evaluated the diagnostic value of blood biomarkers, endometrial biomarkers, and combined noninvasive tests for the diagnosis of endometriosis [10,30,31]. Several biomarkers showed good diagnostic value, but no studies directly or indirectly compare the accuracy of different biomarkers. Therefore, it is not clear which individual biomarker or combined biomarker is most effective for detecting endometriosis. The objectives of this meta-analysis and indirect comparison are to assess the diagnostic accuracy of hormonal biomarkers for endometriosis and to compare the diagnostic accuracy of different index tests and determine which one is the optimal modality for the diagnosis of endometriosis.

Methods

Protocol

The protocol for this meta-analysis was registered on PROSPERO (International Prospective Register of Systematic Reviews; registration number CRD42018105126).

Search Strategy

A systematic search was performed using PubMed, Embase, Cochrane Library, and Chinese Biomedicine Literature to identify relevant studies from inception to August 2018. The searches were independently conducted by 2 authors (Y.G. and M.S.) on August 6, 2018. There were no limitations on the year of publication and publication languages. The search terms included 17β-hydroxysteroid dehydrogenase, CYP19, aromatase, estrogen receptor, estrogen sulfotransferase, leucine-rich G protein-coupled receptor 7, relaxin, anti-mullerian hormone, androgen receptor, progesterone receptor, prolactin, GnRH, gonadotropin-releasing hormone, HCG, hormonal marker, sensitivity, specificity, false-positive reactions, false-negative reactions, ROC curve, predictive value, endometriosis, Adenomyosis, and their synonyms. Full details of the literature search strategies are shown in Supplemental Methods. The references of relevant systematic reviews/meta-analyses were searched to identify additional potential studies.

Inclusion Criteria

Types of Studies

We included random controlled trials, cross-sectional studies, case-control studies, and cohort studies that evaluated the diagnostic accuracy of hormonal markers for endometriosis. These were either prospective or retrospective. There were no limitations in minimal quality, minimal sample size, or the number of patients.

Types of Patients

Study participants included women with suspected endometriosis based on clinical symptoms, pelvic examination, or both who undertook the index test as well as the reference standard. All participants received 1 or several index tests. There were no limitations in age, race, or nationality.

Type of Index Tests

Any type of hormonal biomarker aimed at evaluating the diagnostic value for endometriosis was included. The index test was 1 eutopic endometrial (including hormonal biomarkers in menstrual fluid) biomarker or 1 hormonal biomarker combined with other tests.

Reference Standards

The reference standard was the visualization of endometriosis at surgery (laparoscopy or laparotomy) with or
without histologic confirmation, because this was currently the best available test for endometriosis.

Outcomes
Primary outcomes were sensitivity (SEN), specificity (SPE), positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), the area under the curve (AUC), and their respective 95% confidence intervals (CIs). Secondary outcomes were relative diagnostic estimates of different hormonal biomarkers.

Exclusion Criteria
Studies that did not report or provide sufficient information to allow us to calculate the true-positive, false-positive, true-negative, and false-negative values; protocols, review articles, editorials, case reports, summaries, animal and cell studies, meta-analyses, and letters; and duplicate articles and methodologic articles were excluded.

Study Selection
We imported the literature search records into EndNote X7 (Thomson Reuters Corporation, Stamford, CT) literature management software. Two independent reviewers (Y.G. and X.M.) examined the title and abstract of studies found in the search to identify related studies. Then, the same 2 reviewers retrieved the full text of all possibly relevant studies and assessed the eligibility of each study according to the eligibility criteria. Conflicts were resolved by a third reviewer (J.T.).

Data Extraction
A draft data extraction sheet was developed using Microsoft Excel 2013 (www.microsoft.com; Microsoft Corp., Redmond, WA). One reviewer (Y.G., X.M., J.L., or J.W.) extracted the data and a second reviewer (B.W. or J.T.) checked the extracted data. We resolved discordant evaluations by discussion to reach consensus. Regarding the conflicts between true positives, false negatives, true negatives, and false positives, each reviewer calculated the data again and finally obtained a unified result. Regarding the conflict of quality evaluation, each reviewer reviewed the full text again to find the corresponding information to make a judgment and finally reached a consensus.

We extracted data from eligible studies including general information such as author name, year of publication, country of the first author, number of authors, journal name, country of the journal, funding, and types of studies; characteristics of study including age and number of participants, number and name of index test, and number and name of reference test; and the reported number of true positives, false negatives, true negatives, and false positives. If the studies did not report these values, we reconstructed 2 × 2 tables from the diagnostic estimates presented in the article for each index test.

Risk of Bias and Quality of Evidence
The Quality Assessment of Diagnostic Accuracy Studies 2 quality assessment tool was used to assess methodologic quality [32]. This tool consists of 4 key domains that discuss patient selection, index test, reference standard, and flow of patients through the study and timing of the index tests and reference standard. Each domain is assessed in terms of the risk of bias, and the first 3 domains are also assessed in terms of concerns about applicability. Each question is answered as “yes,” “no,” or “unclear.” If all questions of a domain are answered as “yes,” the level of risk of bias is judged as “low,” and if 1 or more questions of a domain are answered as “no,” the bias risk is judged as “high.” Concerns about applicability are rated as “low,” “high,” or “unclear” [32]. In keeping with Cochrane DTA Working Group recommendations, no summary score was calculated because this obscures the importance of individual quality items and can lead to inaccurate conclusions [33,34]. One reviewer (Y.G., X.M., J.L., or J.W.) assessed the methodologic quality of each study according to predefined criteria and a second reviewer (B.W. or J.T.) checked the assessment. We resolved disagreement by discussion.

Statistical Analysis
Pairwise Meta-analysis
We calculated the SEN, SPE, PLR, NLR, DOR, and their corresponding 95% CIs for each index tests with STATA (version 12.0; Stata Corp., College Station, TX) and MetaDiSc 1.40 (Clinical Biostatistics Unit, Ramón y Cajal Hospital, Madrid, Spain). We also plotted SENs and SPEs in the summary receiver operating characteristic (SROC) space, using different symbols for different hormonal biomarkers. In addition, we used STATA (version 12.0; Stata Corp.) and Review Manager (version 5.3, 2014; The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen, Denmark) analysis software to build the hierarchical SROC graphics for each index test if applicable.

Indirect Comparisons between Competing Diagnostic Tests
We calculated relative diagnostic outcomes between index tests including relative SEN, relative SPE, relative DOR, and relative PLR. Then, we conducted indirect comparisons using the relative diagnostic outcomes. All analysis was performed using STATA (version 12.0; Stata Corp.) software.

Assessment of Publication Bias
Deek’s funnel plot was conducted to detect publication bias where there were more than 10 studies available for an index test [35].
Subgroup Analysis
We performed subgroup analyses between glandular and stromal and between proliferative phase and secretory phase. In addition, we also conducted indirect comparisons between different subgroups.

Results
Study Selection
The initial literature search identified 341 studies, and an additional 3 articles were identified from the reference lists of these articles. Of these, 325 were screened after the removal of duplicates. After screening titles and abstracts, there were conflicts in 17 studies, all of which were included to view the full text. Eventually, 286 records were excluded. The full text of the remaining 39 studies was retrieved for further scrutiny, and 22 studies were excluded because 3 studies did not evaluate the diagnostic accuracy of the hormonal biomarker for endometriosis, 10 studies assessed the diagnostic value of other biomarkers, 6 studies were abstracts or reviews, and there were conflicts in 3 studies because no detailed data were provided. Through calculations and discussions with a third reviewer, these 3 studies were unable to obtain complete data and were excluded. Finally, 17 studies met all the inclusion criteria and were included in our meta-analysis. Details of the flowchart of the literature search are shown in Fig. 1.

Study Characteristics
Thirteen studies assessed the diagnostic accuracy of aromatase for evaluating endometriosis, 2 studies assessed the diagnostic accuracy of serum prolactin, and 1 study evaluated EST, 17βHSD2, HCG/LH receptor, and ER-β for the diagnosis of endometriosis. The publication year of the 17 included studies ranged from 1999 to 2017. Of the 17 studies, 8 articles were published in English, 1 article was published in Spanish, and the remaining 8 studies were published in Chinese. The age of patients included in the study ranged from 21 to 55 years, but 4 articles did not report patient age. The reference tests of 3 studies were laparoscopy, 7 studies were laparoscopy or laparotomy, 4 studies were laparoscopy combined histology (biopsy-confirmed ectopic implants), and 3 studies were laparoscopy or laparotomy combined histology. Characteristics of the included studies are summarized in Table 1.

Assessment of Risk of Bias
The quality of included studies is summarized in Supplemental Fig. 1. In the domain of patient selection,
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Language</th>
<th>Design</th>
<th>No. of patients</th>
<th>Age (yr)</th>
<th>Study period</th>
<th>Index test</th>
<th>Reference test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dheenadayalu [37]</td>
<td>2002</td>
<td>UK</td>
<td>English</td>
<td>Prospective</td>
<td>60</td>
<td>21–46</td>
<td>NR</td>
<td>Endometrial aromatase</td>
<td>Laparoscopy</td>
</tr>
<tr>
<td>Johnson [39]</td>
<td>2004</td>
<td>Republic of Chile</td>
<td>Spanish</td>
<td>Prospective</td>
<td>43</td>
<td>25–43</td>
<td>NR</td>
<td>Endometrial aromatase</td>
<td>Laparoscopy</td>
</tr>
<tr>
<td>Wölfler [40]</td>
<td>2005</td>
<td>Austria</td>
<td>English</td>
<td>Prospective</td>
<td>64</td>
<td>21–48</td>
<td>NR</td>
<td>Endometrial aromatase</td>
<td>Laparoscopy, laparotomy + histology</td>
</tr>
<tr>
<td>Matsuzaki [43]</td>
<td>2006</td>
<td>France</td>
<td>English</td>
<td>Prospective</td>
<td>54</td>
<td>NR</td>
<td>2001.5–NR</td>
<td>Endometrial aromatase, 17βHSD2</td>
<td>Laparoscopy</td>
</tr>
<tr>
<td>Huang [44]</td>
<td>2008</td>
<td>China</td>
<td>Chinese</td>
<td>Prospective</td>
<td>100</td>
<td>32.9</td>
<td>2005.6–2007.5</td>
<td>Endometrial aromatase</td>
<td>Laparoscopy, laparotomy + histology</td>
</tr>
<tr>
<td>Hudelist [45]</td>
<td>2008</td>
<td>Austria</td>
<td>English</td>
<td>Prospective</td>
<td>45</td>
<td>22–45</td>
<td>NR</td>
<td>Endometrial HCG/LH receptor</td>
<td>Laparoscopy, laparotomy</td>
</tr>
<tr>
<td>Li [49]</td>
<td>2012</td>
<td>China</td>
<td>Chinese</td>
<td>Prospective</td>
<td>72</td>
<td>35–51</td>
<td>2009.6–2010.12</td>
<td>Endometrial aromatase</td>
<td>Laparoscopy, laparotomy + histology</td>
</tr>
<tr>
<td>Biibio [50]</td>
<td>2014</td>
<td>Brazil</td>
<td>English</td>
<td>Prospective</td>
<td>97</td>
<td>26–41</td>
<td>NR</td>
<td>Serum prolactin</td>
<td>Laparoscopy, laparotomy + histology</td>
</tr>
<tr>
<td>Zhou [51]</td>
<td>2017</td>
<td>China</td>
<td>Chinese</td>
<td>Prospective</td>
<td>90</td>
<td>NR</td>
<td>2014.5–2015.12</td>
<td>Endometrial aromatase</td>
<td>Laparoscopy, laparotomy + histology</td>
</tr>
</tbody>
</table>

NR = not reported.
The total study population was 858 patients. The diagnostic value of aromatase for the diagnosis of endometriosis is presented in Supplemental Fig. 2. Thirteen studies [2,36–41,43,44,47–49,51] evaluated the diagnostic value of aromatase for the diagnosis of endometriosis. The total study population was 858 patients. The pooled estimates of these studies were as follows: SEN, .79 (95% CI, .71–.86); specificity, .89 (95% CI, .82–.94); and DOR, 31.27 (95% CI, 15.70–62.29). The AUC was .91 (95% CI, .88–.93).

One study [45] evaluated the value of HCG/LH receptor to detect endometriosis. The total study population was 45 patients. The pooled estimates of the study were as follows: SEN, .30 (95% CI, .18–.46); SPE, .80 (95% CI, .65–.90), and DOR, 1.56 (95% CI, .24–10.19). One study [46] evaluated the ability of ER-α and ER-β to detect endometriosis. The total study population was 90 patients. The SENs of ER-α and ER-β were .75 (95% CI, .66–.83) and .65 (95% CI, .56–.74). The SPEs of ER-α and ER-β were .47 (95% CI, .34–.60) and .68 (95% CI, .55–.80).

Two studies [42,50] reported the diagnostic efficacy of serum prolactin. There were a total of 187 patients. The pooled SEN, SPE, DOR, and AUC were .45 (95% CI, .38–.52), .92 (95% CI, .85–.97), 10.22 (95% CI, 4.35–24.05), and .83 (95% CI, .68–.97), respectively. One study [2] evaluated the value of EST to detect endometriosis. There were a total of 68 patients. The pooled SEN, SPE, and DOR were .69 (95% CI, .51–.83), .30 (95% CI, .16–.49), and .95 (95% CI, .34–2.66), respectively.

One study [43] reported the diagnostic efficacy of 17βHSD2 for endometriosis. There were 53 patients in total. The pooled SEN, SPE, and DOR were .73 (95% CI, .60–.84), .48 (95% CI, .33–.63), and 3.22 (95% CI, .18–.56.39), respectively.

Detailed diagnostic accuracy estimates are shown in Table 2. The SROC curves of aromatase and serum prolactin are presented in Supplemental Fig. 3A and the hierarchical SROC curve in Supplemental Fig. 3B.

### Results of Meta-analyses

#### Diagnostic Value of Hormonal Biomarkers for the Diagnosis of Endometriosis

Thirteen studies [2,36–41,43,44,47–49,51] evaluated the diagnostic value of aromatase for the diagnosis of endometriosis. The total study population was 858 patients. The pooled estimates of these studies were as follows: SEN, .79 (95% CI, .71–.86); specificity, .89 (95% CI, .82–.94); and DOR, 31.27 (95% CI, 15.70–62.29). The AUC was .91 (95% CI, .88–.93).

One study [45] evaluated the value of HCG/LH receptor to detect endometriosis. The total study population was 45 patients. The pooled estimates of the study were as follows: SEN, .30 (95% CI, .18–.46); SPE, .80 (95% CI, .65–.90), and DOR, 1.56 (95% CI, .24–10.19). One study [46] evaluated the ability of ER-α and ER-β to detect endometriosis. The total study population was 90 patients. The SENs of ER-α and ER-β were .75 (95% CI, .66–.83) and .65 (95% CI, .56–.74). The SPEs of ER-α and ER-β were .47 (95% CI, .34–.60) and .68 (95% CI, .55–.80).

Two studies [42,50] reported the diagnostic efficacy of serum prolactin. There were a total of 187 patients. The pooled SEN, SPE, DOR, and AUC were .45 (95% CI, .38–.52), .92 (95% CI, .85–.97), 10.22 (95% CI, 4.35–24.05), and .83 (95% CI, .68–.97), respectively. One study [2] evaluated the value of EST to detect endometriosis. There were a total of 68 patients. The pooled SEN, SPE, and DOR were .69 (95% CI, .51–.83), .30 (95% CI, .16–.49), and .95 (95% CI, .34–2.66), respectively.

One study [43] reported the diagnostic efficacy of 17βHSD2 for endometriosis. There were 53 patients in total. The pooled SEN, SPE, and DOR were .73 (95% CI, .60–.84), .48 (95% CI, .33–.63), and 3.22 (95% CI, .18–.56.39), respectively.

### Group Analysis

There was no statistical difference (p >.05) in SEN, SPE, and DOR of aromatase between the proliferative phase and secretory phase. The SEN (.61 [95% CI, .39–.80] vs .00 [95% CI, .00–.15]) of HCG/LH receptor in glandular tissues was significantly higher than that of stromal tissues (p <.05). Considering ER-α, there was a statistically significant difference in SEN (.93 [95% CI, .84–.98] vs .57 [95% CI, .43–.69]) and SPE (.27 [95% CI, .12–.46] vs .67 [95% CI, .47–.83]) between the proliferative phase and the secretory phase (p <.05). The SEN (.90 [95% CI, .80–.96] vs .40 [95% CI, .28–.54]) of ER-β in the proliferative phase was significantly higher than that in the secretory phase (p <.05), but the SPE (.90 [95% CI, .74–.98] vs .47 [95% CI, .28–.66]) of ER-β in the secretory phase was significantly higher than that in the proliferative phase (p <.05). Serum prolactin with a cut-off threshold of 14.8 ng/mL had a higher SEN than a cut-off threshold of 20.0 ng/mL (p <.05). The SEN (.93 [95% CI, .78–.99] vs .53 [95% CI, .34–.72]) of 17βHSD2 in glandular tissues was significantly higher than that in stromal tissues, whereas the specificity (.04 [95% CI, .00–.22] vs .91 [95% CI, .72–.99]) and PLR (.98 [95% CI, .86–1.11] vs 6.13 [95% CI, 1.57–24.04]) were lower than that in stromal tissues (p <.05) (Supplemental Table 1).

### Indirect Comparisons between Competing Diagnostic Tests

The SENs of aromatase, ER-α, ER-β, EST, and 17βHSD2 were significantly higher than that of HCG/LH receptor and serum prolactin (p <.05). The SPEs of aromatase, HCG/LH receptor, and serum prolactin were significantly higher than that of ER-α, EST, and 17βHSD2 (p <.05). The PLR and DOR of aromatase were significantly higher than that of HCG/LH receptor, ER-α, ER-β, and EST (p <.05). Compared with ER-β and serum prolactin, ER-α had a lower PLR and DOR (p <.05). The details of indirect comparisons are shown in Fig. 2.

---

**Table 2**

<table>
<thead>
<tr>
<th>Index test</th>
<th>No. of studies</th>
<th>Sample</th>
<th>SEN (95% CI)</th>
<th>SPE (95% CI)</th>
<th>PLR (95% CI)</th>
<th>NLR (95% CI)</th>
<th>DOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aromatase</td>
<td>13</td>
<td>858</td>
<td>.79 (.71–.86)</td>
<td>.89 (.82–.94)</td>
<td>7.22 (4.26–12.24)</td>
<td>.23 (.16–.33)</td>
<td>31.27 (15.70–62.29)</td>
</tr>
<tr>
<td>HCG/LH receptor</td>
<td>1</td>
<td>45</td>
<td>.30 (.18–.46)</td>
<td>.80 (.65–.90)</td>
<td>1.49 (0.64–3.47)</td>
<td>.82 (.33–2.09)</td>
<td>1.56 (2.24–10.19)</td>
</tr>
<tr>
<td>ER-α</td>
<td>1</td>
<td>90</td>
<td>.75 (.66–.83)</td>
<td>.47 (.34–.60)</td>
<td>1.41 (1.09–1.82)</td>
<td>.54 (.36–.81)</td>
<td>2.62 (1.42–4.84)</td>
</tr>
<tr>
<td>ER-β</td>
<td>1</td>
<td>90</td>
<td>.65 (.56–.74)</td>
<td>.68 (.55–.80)</td>
<td>2.05 (1.39–3.04)</td>
<td>.51 (.38–.69)</td>
<td>4.03 (2.14–7.58)</td>
</tr>
<tr>
<td>Serum prolactin</td>
<td>2</td>
<td>187</td>
<td>.45 (.38–.52)</td>
<td>.92 (.85–.97)</td>
<td>4.86 (2.40–9.82)</td>
<td>.59 (.39–.91)</td>
<td>10.22 (4.35–24.05)</td>
</tr>
<tr>
<td>EST</td>
<td>1</td>
<td>68</td>
<td>.69 (.51–.83)</td>
<td>.30 (.16–.49)</td>
<td>.98 (.72–1.35)</td>
<td>1.04 (.51–2.11)</td>
<td>.95 (0.34–2.66)</td>
</tr>
<tr>
<td>17βHSD2</td>
<td>1</td>
<td>53</td>
<td>.73 (.60–.84)</td>
<td>.48 (.33–.63)</td>
<td>2.38 (.04–138.12)</td>
<td>.53 (.36–.79)</td>
<td>3.22 (1.18–56.39)</td>
</tr>
</tbody>
</table>
Assessment of Publication Bias

We performed Deek’s test [35] to detect publication bias, and the results showed for aromatase p = .165, indicating a that there is less likelihood of publication bias. Because of the lack of necessary data, we did not explore the publication bias of the other 6 biomarkers. The result of publication bias is shown in Supplemental Fig. 4.

Discussion

Noninvasive diagnostic tests or minimally invasive tests will reduce the associated surgical risks, increase accessibility to a diagnostic test, and improve treatment outcomes [30]. Although multiple markers and imaging techniques have been explored as diagnostic tests for endometriosis, none has been implemented routinely in clinical practice and many have not been subject to systematic review [31]. This meta-analysis and indirect comparison summarized direct and indirect evidence to assess the diagnostic value of the hormonal biomarkers for endometriosis. We assessed methodologic quality using the Quality Assessment of Diagnostic Accuracy Studies 2 tool. Moreover, we performed indirect comparisons between different hormonal biomarkers to find a more effective biomarker for the diagnosis of endometriosis.

According to the pooled results of the 13 included studies, the SEN and SPE of aromatase in detecting endometriosis were .79 and .89. The PLR was higher than 7.00 and the
DOR higher than 31.00, which indicates the aromatase had a high diagnostic value. However, the SEN and SPE of our study were a little higher than those obtained in a previous meta-analysis conducted by Gupta et al [10], which included 8 studies with 444 patients; they obtained a mean SEN of .78 and a mean SPE of .74 for pelvic endometriosis. Compared with the study by Gupta et al [10], we included more studies, and the sample size of our study was approximately 1.93 times (858 patients in our study vs 444 patients) that of Gupta et al [10]. Large samples can improve the representativeness and statistical power [52], so our results tend to be more convincing. Subgroup analysis showed that the diagnostic value of aromatase in the proliferative phase was similar to that in the secretory phase. This indicates that obtaining a proliferative or secretory endometrial sample does not affect the diagnostic accuracy of aromatase for endometriosis.

The pooled SPE of HCG/LH receptor for the diagnosis of endometriosis was .80; however, the pooled SEN was .30. Although the SEN of glandular components was higher than .60, the SEN of stromal cells was almost 0, indicating that the HCG/LH receptor was mainly detected in glandular components and rarely detected in stromal cells. When detecting HCG/LH receptors, it was particularly important to obtain suitable tissue specimens based on the patient’s menstrual cycle. Many patients will not be diagnosed correctly if the specimens are not selected properly. Therefore, our findings did not support HCG/LH receptor as a diagnostic test for endometriosis.

Estrogen can directly stimulate the proliferation of cells in ectopic lesions and can also stimulate a variety of factors involved in the pathogenesis of ovarian endometriosis [53]. Previous studies suggested that the predominant expression of ER-α in both glandular epithelial and stromal cells may be essential for the development and growth of peritoneal and ovarian endometriosis [54]. The current study showed the SENs and SPEs of ER-α in the glandular cells and stromal cells were similar. However, its SEN and SPE showed periodical differences. The SEN of ER-α in the proliferative phase was higher than .90, but the SPE was lower than .30. In contrast, ER-α had a low SEN of .57 in the secretory phase. This indicates that ER-α cannot be a good diagnostic method for endometriosis. The physiologic functions of the 2 subtypes of ER were not fully understood, but it has been reported that 1 possible role of ER-β is to modulate ER-α transcriptional activity [53,55]. Just like ER-α, the SENs and SPEs of ER-β in the glandular cells and stromal cells were similar. The SEN and SPE of ER-β also showed periodical differences, and the SEN in the proliferative phase was higher than that in the secretory phase. ER-β had a SEN of .90 in the proliferative phase, which indicates that ER-β in the proliferative phase had a better diagnostic value. However, the SPE of ER-β was lower than .50, which indicated that more than 50.00% of patients may be misdiagnosed.

The pooled SEN and SPE of serum prolactin were .45 and .92, respectively. Despite the high SPE, the SEN for the diagnosis of endometriosis remains unacceptably low. Xavier et al [56] highlighted the importance of choosing the correct threshold for a putative biomarker. However, results of the subgroup analysis showed that using a serum prolactin with a cut-off of 14.8 ng/mL and 20.0 ng/mL produced SENs of .44 and .21, values obviously still below the ideal for a diagnostic test.

For EST, the SEN was .69, and the corresponding value for 17βHSD2 was .73. These approaches can be risky, however; presently, these 2 techniques cannot predict endometriosis with a high specificity. Kitawaki et al [57] demonstrated that progesterone induces 17βHSD2 expression, not in normal endometrium from disease-free women but in eutopic endometrium from endometriosis patients [43]. The current study indicated a differential expression of 17βHSD2 in epithelial cells and stromal cells of the eutopic endometrium. The epithelial cells had a high SEN of .93 but a very low SPE of .04. In contrast, the stromal cells had a high SPE of .91 but a low SEN of .53. These data suggested that 17βHSD2 was either not sensitive or specific enough to be clinically useful in diagnosing endometriosis.

To compare the diagnostic accuracy of different hormonal biomarkers, we conducted indirect comparisons. Compared with HCG/LH receptor, ER-α, ER-β, EST, and 17βHSD2, aromatase had a higher SEN, SPE, PLR, and DOR, although some estimates were not statistically significant between them. The SPEs of aromatase and serum prolactin were comparable, but the SEN, PLR, and DOR of serum prolactin were much lower than that of aromatase. These results indicate that aromatase had better diagnostic value than the other 6 hormonal biomarkers in the diagnosis of endometriosis. In addition, based on the results of the SROC curve, the AUC of aromatase was close to 1, which also suggested that aromatase may be an excellent diagnostic test for endometriosis.

This review systematically evaluated the diagnostic value of 7 hormonal biomarkers to provide the most accurate test for diagnosing endometriosis. The main strengths of the review were that, for the first time, this study used indirect comparative meta-analysis to analyze the diagnostic accuracy of hormonal biomarkers, which clearly shows the difference between the diagnostic values of different biomarkers. Second, in addition to conducting indirect comparisons between different competitive diagnostic tests, we also performed indirect comparisons of different subgroups of each hormonal biomarker.

However, this study also had some limitations. First, although we included 17 studies with 7 hormonal biomarkers, there were a small number of studies for most index tests. This may undermine the reliability of the summary estimates from the meta-analyses. Second, almost all studies did not provide enough information to determine the interpretation of the index tests without knowing the results of the reference standard. Third, most studies did not report the time interval between diagnostic test and...
reference standard test. Fourth, the same index test may have different diagnostic values for different grades of endometriosis. However, we were unable to conduct a subgroup analysis to explore this issue because of the lack of necessary data. Finally, the methodologic quality of the included studies was generally moderate, which may affect the dissemination of evidence summarized in this study.

In conclusion, our study indicated that aromatase had a relatively high SEN and SPE for the diagnosis of endometriosis. Combined with higher DOR and AUC, we believe that aromatase may be an excellent diagnostic test for endometriosis. However, because of the moderate quality of the included studies and the limited sample size, this result requires more research to validate.

Acknowledgment

We thank Jiarui Wu and Junhai Jia for supporting the methodology.

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.jmig.2019.04.004.

References


Supplemental Methods

Search Strategies

PubMed

#1 “17β HSD” [Title/Abstract] OR “17-β hydroxysteroid dehydrogenase” [Title/Abstract]
#2 “17β HSD2” [Title/Abstract] OR “17-β hydroxysteroid dehydrogenase type 2” [Title/Abstract]
#3 “Cytochrome P450 Family 19” [Mesh] OR “CYP19” [Title/Abstract] OR “aromataze cytochrome P450” [Title/Abstract]
#4 “Receptors, Estrogen” [Mesh] OR ER[Title/Abstract] OR “estrogen receptor” [Title/Abstract] OR “oestrogen receptor” [Title/Abstract] OR “oestrogen receptors” [Title/Abstract] OR “estrogen receptors” [Title/Abstract]
#5 “Estrogen Receptor alpha” [Mesh] OR ER-α[Title/Abstract] OR “oestrogen receptor-alpha” [Title/Abstract] OR “oestrogen receptor alpha” [Title/Abstract] OR “estrogen receptor alpha” [Title/Abstract]
#7 EST [Title/Abstract] OR “oestrogen sulphotransferase” [Title/Abstract] OR “oestrogen sulphotransferase” [Title/Abstract] OR “oestrogen sulfotransferase” [Title/Abstract] OR “oestrogen sulfotransferase” [Title/Abstract]
#8 LGR7 [Title/Abstract] OR “leucine-rich G protein-coupled receptor 7” [Title/Abstract]
#9 “Relaxin” [Mesh] OR Relaxin [Title/Abstract]
#10 “Anti-Mullerian Hormone” [Mesh] OR AMH [Title/Abstract] OR “anti-mullerian hormone” [Title/Abstract]
#11 “Receptors, Androgen” [Mesh] OR “AR” [Title/Abstract] OR “androgen receptor” [Title/Abstract] OR “androgen receptors” [Title/Abstract]
#12 “Receptors, Progesterone” [Mesh] OR PR [Title/Abstract] OR “progesterone receptor” [Title/Abstract] OR “progesterone receptors” [Title/Abstract] OR “progestogen receptor” [Title/Abstract]
#13 “Prolactin” [Mesh] OR PRL [Title/Abstract] OR prolactin [Title/Abstract] OR lactogen [Title/Abstract]
#16 “Hormonal marker” [Title/Abstract] OR “Hormonal markers” [Title/Abstract]
#17 OR/1−16
#19 “Endometriosis” [Mesh] OR Endometrio* [Title/Abstract]
#20 “Adenomyosis” [Mesh] OR “adenomyosis” [Title/Abstract]
#21 #19 OR #20
#22 #17 AND #18 AND #21
Supplemental Fig. 1
Risk of bias and applicability concerns of included studies.

Supplemental Fig. 2
Detailed information of risk of bias and applicability concerns for each included study.
**Supplemental Fig. 3**

(A) SROC curves for aromatase and serum prolactin. (B) Hierarchical SROC curve for aromatase.

**Supplemental Fig. 4**

Publication bias of aromatase using Deek's funnel plot. ESS = effective sample size
### Supplemental Table 1

**Subgroup Analysis of Aromatase, HCG/LH Receptor, ER-α, ER-β, Serum Prolactin, and 17βHSD2**

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Subgroup</th>
<th>No. of studies</th>
<th>Sample</th>
<th>SEN (95% CI)</th>
<th>SPE (95% CI)</th>
<th>PLR (95% CI)</th>
<th>NLR (95% CI)</th>
<th>DOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aromatase</strong></td>
<td>Cycle phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proliferative phase</td>
<td>5</td>
<td>168</td>
<td>.84 (.75–.91)</td>
<td>.92 (.84–.97)</td>
<td>7.10 (1.46–34.57)</td>
<td>.20 (.11–.37)</td>
<td>45.01 (9.89–204.72)</td>
</tr>
<tr>
<td></td>
<td>Secretory phase</td>
<td>5</td>
<td>180</td>
<td>.86 (.76–.93)</td>
<td>.91 (.84–.96)</td>
<td>6.66 (3.24–13.67)</td>
<td>.19 (.11–.32)</td>
<td>42.22 (14.90–119.63)</td>
</tr>
<tr>
<td><strong>HCG/LH receptor</strong></td>
<td>Cycle phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proliferative phase</td>
<td>1</td>
<td>29</td>
<td>.31 (.14–.52)</td>
<td>.91 (.75–.98)</td>
<td>3.28 (.97–11.14)</td>
<td>.76 (.58–1.01)</td>
<td>4.30 (1.01–18.34)</td>
</tr>
<tr>
<td></td>
<td>Secretory phase</td>
<td>1</td>
<td>16</td>
<td>.30 (.12–.54)</td>
<td>.83 (.52–.98)</td>
<td>1.80 (.43–7.53)</td>
<td>.84 (.57–1.23)</td>
<td>2.14 (.36–12.89)</td>
</tr>
<tr>
<td><strong>ER-α</strong></td>
<td>Specimen</td>
<td>Glandular</td>
<td>1</td>
<td>45</td>
<td>.61 (.39–.80)</td>
<td>.82 (.60–.95)</td>
<td>3.35 (1.35–8.61)</td>
<td>.48 (.28–.83)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stromal</td>
<td>1</td>
<td>45</td>
<td>.00 (.00–.15)</td>
<td>.96 (.77–1.00)</td>
<td>.32 (.01–.74)</td>
<td>1.05 (.93–1.18)</td>
</tr>
<tr>
<td><strong>ER-β</strong></td>
<td>Specimen</td>
<td>Glandular</td>
<td>1</td>
<td>90</td>
<td>.73 (.60–.84)</td>
<td>.43 (.26–.63)</td>
<td>1.29 (.91–1.83)</td>
<td>.62 (.34–1.11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stromal</td>
<td>1</td>
<td>90</td>
<td>.77 (.64–.87)</td>
<td>.50 (.31–.69)</td>
<td>1.53 (1.04–2.25)</td>
<td>.47 (.26–.84)</td>
</tr>
<tr>
<td><strong>Serum prolactin</strong></td>
<td>Cut-off</td>
<td>14.8 ng/mL</td>
<td>1</td>
<td>97</td>
<td>.44 (.32–.58)</td>
<td>.94 (.80–.99)</td>
<td>7.56 (1.92–29.81)</td>
<td>.59 (.47–.75)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20.0 ng/mL</td>
<td>1</td>
<td>97</td>
<td>.21 (.12–.33)</td>
<td>.97 (.85–1.00)</td>
<td>7.02 (1.96–51.36)</td>
<td>.82 (.71–.94)</td>
</tr>
<tr>
<td><strong>17βHSD2</strong></td>
<td>Specimen</td>
<td>Glandular</td>
<td>1</td>
<td>53</td>
<td>.93 (.78–.99)</td>
<td>.04 (.00–.22)</td>
<td>.98 (.86–1.11)</td>
<td>1.53 (.15–15.89)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stromal</td>
<td>1</td>
<td>53</td>
<td>.53 (.34–.72)</td>
<td>.91 (.72–.99)</td>
<td>6.13 (1.57–24.04)</td>
<td>.51 (.34–.77)</td>
</tr>
</tbody>
</table>