BMJ Open

Diagnostic accuracy of endometrial sampling tests for detecting endometrial cancer: a systematic review and meta-analysis

Noha Abdelsattar Sakna, Marwa Elgendi, Mohamed Hamed Salama, Ahmed Zeinhom, Somaia Labib, Ashraf Fawzy Nabhan

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This systematic review included primary studies with appropriate designs to assess the diagnostic accuracy of endometrial sampling tests studies, thus minimising spectrum bias.

⇒ This systematic review used the quality assessment tool for diagnostic accuracy studies (QUADAS-2) tool for quality appraisal and GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) for assessing the certainty of evidence.

⇒ The meta-analysis provides a robust synthesis of the evidence of the diagnostic accuracy of endometrial sampling tests for women typically seen in clinical practice with diagnostic uncertainty.

⇒ The potential for publication bias was not assessed due to the lack of validated methods for diagnostic test accuracy reviews.

INTRODUCTION

Endometrial cancer (EC) is the most common gynaecological tumour in developed countries1 2 and the fifth most common malignant condition among women worldwide.3–7 Its incidence shows an increase in line with an increase in life expectancy. It is a major health concern being the 14th leading cause of death from cancer in women.3–7

Most cases affect postmenopausal women.8 Most women present with abnormal uterine bleeding.9 While 90% of women with EC initially present with postmenopausal bleeding (PMB), only 5%–10% of women with PMB are diagnosed with EC. Irregular premenopausal or perimenopausal bleeding is another important presenting symptom.9 EC is sometimes diagnosed in women without an abnormal uterine bleeding. This occurs during the investigation of a thick endometrial line found on an imaging technique performed for other reasons or incidentally in hysterectomy specimens performed for benign conditions.5 10

ABSTRACT

Objectives To determine the diagnostic accuracy of different endometrial sampling tests for detecting endometrial carcinoma.

Design Systematic review and meta-analysis of studies of diagnostic accuracy.

Data sources Cochrane Library, MEDLINE/PubMed, CINAHL, Web of Science and Scopus, from the date of inception of the databases to 18 January 2023. Additionally, the reference lists of included studies and other systematic reviews were thoroughly searched.

Eligibility criteria We included published cross-sectional studies that evaluated any endometrial sampling test (index tests) in women (participants) with clinical suspicion of endometrial carcinoma (target condition) in comparison with histopathology of hysterectomy specimens (reference standard). We excluded case–control and case series studies. No restrictions on language or date of publication were applied.

Data extraction and synthesis Two independent reviewers extracted study data and assessed study quality using the revised quality assessment tool for diagnostic accuracy studies (QUADAS-2). We used bivariate diagnostic random-effects meta-analysis and presented the results in a summary receiver operating characteristic curve. We assessed the certainty of evidence as recommended by the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) approach.

Results Twelve studies (1607 participants), published between 1986 and 2022, contributed data to the meta-analysis results. Seven studies were judged to be at a low risk of bias in all domains and all studies had low applicability concerns. The most studied index tests were Pipelle and conventional dilation and curettage (D&C). The sensitivity, specificity, positive likelihood ratio and negative likelihood ratio (95% CIs) for Pipelle were 0.774 (0.565 to 0.900), 0.985 (0.927 to 0.997), 97.000 (14.000 to 349.000) and 0.241 (0.101 to 0.442) and for conventional D&C were 0.880 (0.281 to 0.993), 0.984 (0.956 to 0.995), 97.000 (14.000 to 153.000) and 0.194 (0.007 to 0.732), respectively.

Conclusion High certainty evidence indicates that endometrial sampling using Pipelle or conventional D&C is accurate in diagnosing endometrial cancer. Studies assessing other endometrial sampling tests were sparse. 

Trial registration number https://osf.io/h8e9z.
Survival from EC depends on the stage at diagnosis, tumour grading, deep myometrial invasion, lymphovascular space invasion and nodal metastasis. Therefore, presenting with clinical suspicion of endometrial carcinoma need an accurate early detection remains the cornerstone for improving outcomes. Therefore, presenting with clinical suspicion of endometrial carcinoma need an accurate early diagnosis, followed by assessment of clinicopathological factors and molecular subtypes to stratify the risk of recurrence and to tailor adjuvant treatment, for example pre-operative radiomics analysis.

**Objectives**
The aim of this systematic review and meta-analysis is to determine the diagnostic accuracy of different endometrial sampling tests for detecting endometrial carcinoma in women presenting with clinical suspicion of endometrial carcinoma.

**METHODS**
We conducted this systematic review using methodological approaches prespecified in a review protocol (online supplemental file 1), which was prospectively registered in the Centre for Open Science. We reported this review according to the Preferred Reporting Items for a Systematic review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines.

**Eligibility criteria**
We included published cohort selected cross-sectional studies that evaluated the diagnostic accuracy of different endometrial sampling methods, followed by verification with the most appropriate reference standard. This one-gate study design recruits suspicious participants to receive the index test and the reference test to determine their diagnosis. The cross-sectional relationship implies that the order of the reference and the index test can differ as long as the tests are performed at the same time or without delay that might introduce a risk of bias in the domain of flow and timing. Thus, we accepted studies where the reference and the index test were performed simultaneously, the reference test was performed after the index test or the index test was performed after the reference test.

Participants are women, of all ages, who had an endometrial sampling for a suspected endometrial hyperplasia or cancer. Index tests included different endometrial sampling techniques. Target conditions included endometrial hyperplasia or carcinoma. We accepted histopathology of hysterectomy specimens as the reference standard for the diagnosis of the target conditions. We excluded studies of a case–control design that included women with known endometrial carcinoma to matched controls.

**Information sources**
We developed a comprehensive search strategy to find published articles (online supplemental file 2). We searched the Cochrane Library, MEDLINE/PubMed, CINAHL, Web of Science and Scopus from the date of inception of the databases to 18 January 2023. We did not apply any restrictions on language or date of publication.

**Searching other resources**
We manually searched the reference lists in articles retrieved from electronic databases and relevant review articles.

**Study selection**
Two review authors (SL and NS) used a reference manager, to import and deduplicate search results. They independently screened titles and abstracts to identify potentially eligible studies for full-text retrieval. Two review authors (SL and ME) double checked the list of potentially eligible studies. Review authors working in pairs, independently assessed the full text using the predefined eligibility criteria. They resolved discrepancies by discussion and by consulting with the lead author (AN). We illustrated the study selection process using a PRISMA flowchart.

**Data collection process**
For each included study, at least two review authors (SL, NS and ME) independently extracted the following data: general information: title, journal, year and study design; sample size: total number of participants included and tested; baseline characteristics: age, index test, reference test; numbers of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) findings. We extracted these data for both target conditions (atypical endometrial hyperplasia (AEH) and EC). All numbers were double checked by two authors (MS and AZ). We summarised the data from each study in 2x2 tables (TP, FP, FN and TN), and we stored the data into a spreadsheet.

**Risk of bias and applicability**
Review authors, working in pairs, independently assessed the risk of bias of included studies and applicability of their results using quality assessment tool for diagnostic accuracy studies (QUADAS-2). They resolved discrepancies by discussion and by consulting with the lead author (AN). We addressed aspects of study quality involving the participant spectrum, index tests, target conditions, reference standards and flow and timing. We classified a study as having a high risk of bias if at least one of the domains of QUADAS-2 was judged as being high risk.

**Synthesis of results**
Each method of endometrial sampling was compared against the reference test. For each method of endometrial sampling, TP, TN, FP and FN were computed. We used bivariate diagnostic random-effects meta-analysis. We presented the results in a summary receiver operating characteristic curve (SROC). We used the Markov chain Monte Carlo procedure to generate summary positive and negative likelihood ratio (LR) and diagnostic OR for the bivariate model. When there are three or less studies, we used a univariate random-effects model. We explored heterogeneity using predefined subgroup analysis. For each of the target condition, we
presented the synthesis by the method of sample collection. We assessed the effect of risk of bias of included studies on diagnostic accuracy by performing a sensitivity analysis in which we excluded studies classified as having high risk of bias. We used R software V.4.2.222 and package ‘mada’.23

Summary of findings table and assessment of the certainty of evidence
We prepared summary of findings tables, using GRADEpro GDT,24 to present the main results and key information regarding the certainty of evidence. We assessed the certainty of evidence as recommended by the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) approach.25

We rated the certainty of evidence as either high (when not downgraded), moderate (when downgraded by one level), low (when downgraded by two levels) or very low (when downgraded by more than two levels) based on five domains: risk of bias, indirectness, inconsistency, imprecision and publication bias. For each outcome, the certainty of evidence started as high when there were high-quality observational studies (cross-sectional or cohort studies) that enrolled participants with diagnostic uncertainty. If we found a reason for downgrading, we used our judgement to classify the reason as either serious (downgraded by one level) or very serious (downgraded by two levels).26 27

Patient and public involvement
None.

RESULTS
Study selection
Bibliographic database search identified 3386 records, and 23 additional records were identified by searching the reference lists of relevant records. After removing duplicates, 2679 titles and abstracts were screened. Full-text reports of 38 potentially eligible studies were assessed using the predefined eligibility criteria. Twelve studies were included (figure 1).

Study characteristics
The 12 studies recruited 1607 participants and were published between 1986 and 2022. All included studies had a one-gate cross-sectional design (table 1).

The majority of recruited women were postmenopausal. Participants had different presentations including postmenopausal bleeding, premenopausal abnormal uterine bleeding or women scheduled for hysterectomy for different indications.

Studies used different methods of endometrial sampling. This included Pipelle,28–30 GDP Tao device,31 Abradul cell sampler,32 Li brush,33 Uterine Explora Curette (UEC),34 Gynoscann,34 Novak,35 Masterson curette,36 hysteroscopy-directed37 or D&C.28 29 37–39 All studies used histopathology examination of hysterecetomy as the reference standard.

Risk of bias and concerns regarding applicability
Regarding bias, seven studies were classified as low risk in all domains. Two studies had a high risk of bias. In one study,28 only 25 (29.1%) participants had the reference standard and with unclear interval between index tests and reference standard. In the second study,37 25 (50%) hysteroscopic guided biopsies and 29 (58%) D&C biopsies had the reference standard. Four studies had an unclear risk of bias in the domains of patient selection, reference standard and flow and timing. All studies had low concerns regarding applicability.

Data synthesis
Diagnostic accuracy of endometrial sampling tests compared with histopathology of hysterectomy
Twelve studies reported data on EC.28–39 The SROC is depicted in figure 2 and online supplemental file 3.
Ten studies reported data on AEH using different methods of endometrial sampling. The SROC is depicted in figure 3. The GRADE summary of findings table presents the main results and key information regarding the certainty of evidence (table 2). For EC, out of 100 people with EC, 78 would be correctly diagnosed and 22 would be incorrectly diagnosed. Out of 100 people without EC, 99 would be correctly diagnosed and one would be incorrectly diagnosed (table 2) see full table version uploaded in supplementary file as online supplemental table S2.

Subgroup analysis by the specific test
Five studies examined the accuracy of D&C for diagnosing EC. Using D&C, compared with histopathology

### Table 1 Characteristics of included studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Design</th>
<th>Outcomes</th>
<th>Sample size</th>
<th>Index test</th>
<th>Reference test</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behnamfar and Arshad</td>
<td>2020</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>25</td>
<td>Pipelle, D&amp;C</td>
<td>Histopathology examination of hysterectomy</td>
<td>All women were menopausal, with postmenopausal bleeding (mean age 60)</td>
</tr>
<tr>
<td>de Leon et al</td>
<td>2016</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>42</td>
<td>GDP Tao</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women, mostly with abnormal uterine bleeding (median age 57)</td>
</tr>
<tr>
<td>Gerretsen and Velde</td>
<td>1987</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>61</td>
<td>Abradul</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women who had a total hysterectomy</td>
</tr>
<tr>
<td>Gungorduk et al</td>
<td>2014</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>189</td>
<td>Pipelle, D&amp;C</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women who had a total hysterectomy (mean age 49)</td>
</tr>
<tr>
<td>Han et al</td>
<td>2019</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>271</td>
<td>Li brush</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women who had a total hysterectomy</td>
</tr>
<tr>
<td>Ilavarasi et al</td>
<td>2019</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>106</td>
<td>Pipelle</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women with AUB</td>
</tr>
<tr>
<td>Kufahl et al</td>
<td>1997</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>181</td>
<td>Gynoscan, UEC</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women who had a total hysterectomy (median age 48)</td>
</tr>
<tr>
<td>Moradan et al</td>
<td>2017</td>
<td>A one-gate diagnostic test accuracy</td>
<td>EC</td>
<td>163</td>
<td>D&amp;C</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women with AUB resistant to medical therapy, or AUB in combination with an endometrial thickness of more than 12 mm on performing transvaginal ultrasound</td>
</tr>
<tr>
<td>Reddington et al</td>
<td>1995</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>25</td>
<td>Masterson curette</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women who had a total hysterectomy</td>
</tr>
<tr>
<td>Saadia et al</td>
<td>2011</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>50</td>
<td>D&amp;C</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women who had a total hysterectomy (median age 43)</td>
</tr>
<tr>
<td>Somasundar et al</td>
<td>2022</td>
<td>A one-gate diagnostic test accuracy</td>
<td>EC</td>
<td>54</td>
<td>D&amp;C, hysteroscopy</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women with AUB</td>
</tr>
<tr>
<td>Stever et al</td>
<td>1986</td>
<td>A one-gate diagnostic test accuracy</td>
<td>AEH, EC</td>
<td>440</td>
<td>Novak</td>
<td>Histopathology examination of hysterectomy</td>
<td>Premenopausal and postmenopausal women who had a total hysterectomy (mean age 40)</td>
</tr>
</tbody>
</table>

AEH, atypical endometrial hyperplasia; AUB, abnormal uterine bleeding; D&C, dilation and curettage; EC, endometrial cancer; UEC, Uterine Explora Curette.
examination of hysterectomy, out of 100 people with EC, 77 would be correctly diagnosed and out of 100 women without EC, 99 would be correctly diagnosed (online supplemental file 3). Three studies reported the accuracy of D&C for diagnosing AEH. Using D&C, compared with histopathology examination of hysterectomy, out of 100 people with AEH, 80 would be correctly and out of 100 women without AEH 97 would be correctly diagnosed.

Three studies examined the accuracy of Pipelle for diagnosing AEH or EC. Using Pipelle endometrial sampling, compared with histopathology examination of hysterectomy, out of 100 people with AEH, 74 would be correctly diagnosed and out of 100 women without AEH, 98 would be correctly diagnosed. Using Pipelle, compared with histopathology examination of hysterectomy, out of 100 people with EC, 77 would be correctly diagnosed and out of 100 women without EC, 99 would be correctly diagnosed (online supplemental file 3).

Diagnostic accuracy was examined in a single study for each of the following tests: Novak, GDP Tao device, Abradul cell sampler, Li brush, Uterine Explora Curette, Gynoscann, Z-sampler, Novak and Masterson curette. The measures of diagnostic accuracy, sensitivity, specificity, positive LR and negative LR, for each study are shown in online supplemental file 3.

### DISCUSSION

#### Summary of the evidence

This systematic review assessed the diagnostic accuracy of various endometrial sampling tests for detecting EC or AEH in women with suspected endometrial pathology. We included the one-gate cross-sectional study design where women with diagnostic uncertainty had the index test (endometrial sampling) with direct verification of index test results with an appropriate reference standard (histopathology of hysterectomy). Twelve DTA studies were identified, including 1607 women. All studies were published in peer-reviewed journals and most of these studies were of high methodological quality according to the QUADAS-2 criteria. Pooled estimates of diagnostic test accuracy were computed for various endometrial sampling tests. Robust evidence based on the results of the present review indicate that Pipelle or conventional curettage can correctly confirm or exclude EC or AEH. High-quality evidence is insufficient for hysteroscopy, GDP Tao device, Abradul cell sampler, Li brush, Uterine Explora Curette, Gynoscann, Z-sampler, Novak and Masterson curette.

This systematic review represents the synthesis of exclusively high-quality diagnostic accuracy studies. Every effort was made to minimise the risk of bias. In contrast to previously published reviews, we did not include case-control studies or studies that include only women

---

**Table 2** GRADE summary of findings table—diagnostic accuracy of all endometrial sampling tests compared with histopathology of hysterectomy

<table>
<thead>
<tr>
<th>Index test</th>
<th>Target condition</th>
<th>Number of studies (participants)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive LR (95% CI)</th>
<th>Negative LR (95% CI)</th>
<th>Certainty of the evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>All tests</td>
<td>Endometrial cancer</td>
<td>12 (1607)</td>
<td>0.781 (0.669 to 0.863)</td>
<td>0.989 (0.982 to 0.994)</td>
<td>74.800 (42.100 to 123.000)</td>
<td>0.226 (0.139 to 0.335)</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Atypical hyperplasia</td>
<td>10 (1390)</td>
<td>0.760 (0.639 to 0.849)</td>
<td>0.978 (0.953 to 0.989)</td>
<td>36.100 (16.400 to 69.700)</td>
<td>0.250 (0.155 to 0.368)</td>
<td>High</td>
</tr>
</tbody>
</table>

LR, likelihood ratio.
with endometrial carcinoma. Previously published reviews have included studies with case-control design (two-group design) which is generally not representative of a test’s accuracy in clinical practice.\(^\text{11} - \text{40}\) Spectrum bias through case-control design would lead to suboptimal patient selection. This falsely inflates sensitivity and specificity. Furthermore, both prevalence and predictive values depend on the ratio of people with and without the disease. In case-control studies, this ratio is constructed artificially, and thus prevalence and predictive values calculated from such a study are artefacts.\(^\text{41-44}\) Other reviews have included studies of index tests in women with known endometrial carcinoma. The ability of the test to rule out the disease can never be calculated if the study recruited only participants known to have the condition.\(^\text{44, 45}\)

Second, we only included studies that used histopathology examination of hysterectomy as a reference test. Previous reviews have used multiple verification tests and tests that might incorrectly classify the target condition.\(^\text{11} - \text{46-48}\) Errors in the reference test cause misclassification bias. Misclassification can significantly underestimate sensitivity and specificity. The magnitude of the bias depends on the disease prevalence, the accuracy of the index test and the degree of misclassification. The misclassification rate can vary from study to study depending on the methodology associated with the reference test for example, skill of the pathologist.\(^\text{49, 49}\)

The current review did not aim to assess the accuracy of endometrial sampling as a screening test in the general asymptomatic population. According to the most recent and rigorous recommendations, there is no study or major society guideline that recommend universal screening for EC in asymptomatic women.\(^\text{50, 51}\)

The current review focused on the first step in providing an accurate diagnosis in women presenting with a clinical suspicion of an endometrial carcinoma. Other syntheses of evidence have assessed the diagnostic and prognostic value of molecular and genomic profiling in endometrial carcinoma to tailor the most appropriate management strategies in early-stage cases.\(^\text{52}\)

**Limitations**

A comprehensive literature search was conducted and a meticulous screening process was performed. In order to reduce the risk of publication bias, we did not implement any language or publication status restriction. However, it was not possible to search grey literature. The possibility of unpublished studies always exists. The potential for publication bias was not assessed due to the lack of validated methods for diagnostic test accuracy reviews.

**Conclusions**

In conclusion, endometrial sampling, using Pipelle or conventional curettage, is accurate in diagnosing EC in women presenting with a clinical suspicion of an endometrial pathology. There is insufficient high-quality evidence regarding the diagnostic accuracy of other endometrial sampling tests.

**Twitter** Ashraf Fawzy Nabhan @ecebm

**Acknowledgements** We acknowledge the effort and the spirit of goodwill from Farida Elshafey, Menna Kamel and Aya Ramadan for their contribution to title and abstract screening. Thanks to Farida Elshafey for proofreading the manuscript.

**Contributors** AFN conceived the idea for this review and designed the review methods. NAS, ME and SL collaborated in searching, screening, selecting studies, data extraction and synthesis. AFN wrote the first draft of the manuscript and NAS, ME, AZ, MHS and AFN reviewed the manuscript. All authors read and approved the final version of the manuscript. AFN is the guarantor for this manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, conduct, or reporting or dissemination plans of this research.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.