APPENDIX A. Search strategy.

(Original search performed 2 September 2023; search updated on 21 March 2025.)

Ovid MEDLINE(R) ALL <1946 to March 20, 2025>

1 Hysteroscopy/5853

2 (hystero* or ((uterine or intra?uterine) adj3 endoscop*) or "intra?uterine surger*").mp. 18698

3 1 or 2 18698

4 exp Uterine Hemorrhage/ or Menstruation Disturbances/co or Menstruation/co 25655

5 (((abnormal or irregular or heavy or excessive or dysfunction* or absent or scanty or rare or frequent or infrequent or post?menopaus* or post?coit* or pre?menopaus* or peri?menopaus* or menopaus* or anovulat*) adj3 (h?emorrhag* or bleed* or menstrua* or period*)) or AUB or spotting or menometrorrhagia or menorrhagia or metrorrhagia or polymenorrh?ea).mp. 45737

6 4 or 5 61692

7 exp Diagnosis/ or diagnosis.fs. 10933212

8 (diagnos* or predict* or detect* or determin* or prognos* or confirm* or verif* or accura* or inaccura* or sensitiv* or specific*).mp. 17515198

9 7 or 8 20817432

10 3 and 6 and 9 2732

11 limit 10 to yr="1980 - Current" 2631

Embase <1974 to 2025 March 20>

1 hysteroscopy/ 16190

2 (hystero* or ((uterine or intra?uterine) adj3 endoscop*) or "intra?uterine surger*").mp. 30937

3 1 or 2 30937

4 uterus bleeding/ or exp menstruation disorder/co or menstruation/co 15012

5 (((abnormal or irregular or heavy or excessive or dysfunction* or absent or scanty or rare or frequent or infrequent or post?menopaus* or post?coit* or pre?menopaus* or peri?menopaus* or menopaus* or anovulat*) adj3 (h?emorrhag* or bleed* or menstrua* or period*)) or AUB or spotting or menometrorrhagia or menorrhagia or metrorrhagia or polymenorrh?ea).mp. 75990

6 4 or 5 83296

7 exp diagnosis/ or di.fs. 10251281

8 (diagnos* or predict* or detect* or determin* or prognos* or confirm* or verif* or accura* or inaccura* or sensitiv* or specific*).mp. 21962914

9 7 or 8 23765524

10 3 and 6 and 9 5946

11 limit 10 to yr="1980 -Current" 5901

Cochrane Library (Wiley)

Date Run: 21/03/2025 12:09:12

ID	Search	Hits
#1	MeSH descriptor: [Hysteroscopy] explode all trees	595
#2	hystero* OR ((uterine OR intra-uterine OR intrauterine) NEAR/3 endoscop*) OR ((intra-uterine OR intrauterine) NEXT surger*)	3649
#3	#1 OR #2	3649
#4	MeSH descriptor: [Uterine Hemorrhage] this term only	806
#5	MeSH descriptor: [Menstruation Disturbances] this term only and with qualifier(s): [complications - CO]	27
#6	[mh Menstruation/co]	0
#7	#4 OR #5 OR #6	832
#8	[mh Diagnosis] or [mh /di]	483116
#9	diagnos* OR predict* OR detect* OR determin* OR prognos* OR confirm* OR verif* OR accura* OR inaccura* OR sensitiv* OR specific*	946254
#10	#8 OR #9	1141753
#11	#3 AND #7 AND #10 with Cochrane Library publication date Between Jan 1980 and Dec 2025	89

APPENDIX B. Summary of primary included studies.

Leone et al.[23]

Leone et al.[23] prospectively followed 128 premenopausal [n=88; median age 41, interquartile range (IQR): 34–48] and postmenopausal (n=40; median age 57, IQR: 52–67) patients with abnormal bleeding who initially underwent guided aspiration tissue biopsy with an adapted flexible intrauterine catheter utilised during saline infused sonography [saline infused sonography (SIS); termed sonohysterographic sampling]. All patients had an indication for SIS of inhomogeneous endometrium >4 mm in premenopausal state or thickened endometrium >4 mm or inconclusive findings on two dimensional transvaginal ultrasound (TVS) in postmenopausal state. Patients who were identified at time of SIS as having diffuse endometrial thickening and who agreed to undergo an additional sequential series of diagnostic evaluations within one month (including operative hysteroscopy with directed biopsy under no anesthesia immediately followed by blind dilation and curettage under general anesthesia) were included. Histologic evaluation was performed without disclosing the type of biopsy that was performed to obtain the tissue. Median endometrial specimen areas (mm^2) were measured on all samples, and those obtained by aspiration biopsy at time of SIS and those obtained at time of visually directed hysteroscopic biopsy were found to not be significantly different than by dilation and curettage [dilation and curettage (D&C); p=0.18]. The reference standard was the histopathology results obtained at time of D&C. Comparisons were made regarding the diagnostic accuracy of the guided aspiration biopsy at time of SIS to visually directed hysteroscopic biopsy. The three outcomes of interest for this study were hyperplasia without atypia, endometrial intraepithelial neoplasia (EIN)/atypical hyperplasia, and malignancy.

Angioni et al.[24]

Angioni et al.[24] prospectively followed 319 postmenopausal patients with bleeding (median age 54, range 46–88) who underwent blind endometrial biopsy with Novak curette followed by subsequent outpatient diagnostic CO_2 hysteroscopy performed at least one week later. In this cohort, 8% reported use of menopausal hormone therapy (n=25). Patients with suspected benign intracavitary lesions, identified either by Novak curette histology or hysteroscopic visualisation, underwent subsequent operative hysteroscopy under general anesthesia. Histopathology from operative hysteroscopy specimens served as the reference standard for benign intracavitary lesions. For complex or atypical hyperplasia and endometrial carcinoma, hysterectomy specimens served as the reference standard. Patients with complex or atypical hyperplasia diagnosed by Novak curette (n=20) or operative hysteroscopy (n=7) underwent vaginal hysterectomy with bilateral adnexectomy. Patients with endometrial carcinoma (n=15) underwent abdominal hysterectomy with bilateral adnexectomy and pelvic lymphadenectomy. This study did not report sensitivity or specificity for malignancy (n=15), precluding diagnostic accuracy assessment for this outcome. The three outcomes of interest for this study were hyperplasia/EIN combined, endometrial polyp and submucosal myoma.

Gao et al.[25]

Gao et al.[25] retrospectively analyzed 177 postmenopausal patients who presented with bleeding for reasons other than cervical conditions, intrauterine device impaction, or hormone replacement therapy. Based on diagnostic evaluation received, these patients were categorised into a hysteroscopy group [n=96, mean age, standard deviation; (SD): 58.3 (6.9), median age 58.5, range 45–73] and a diagnostic curettage group [n=81, mean age (SD) 57.8 (7.1), median age=57, range 44–75]. Nine of the 96 patients in the hysteroscopy group had heavy vaginal bleeding and 87 had mild irregular vaginal bleeding. Eight of the 81 patients in the diagnostic curettage group had heavy vaginal bleeding and 73 had mild irregular vaginal bleeding. These characteristics did not differ significantly between the two groups. Diagnostic hysteroscopy was performed based on visual inspection using 5% glucose solution as distension medium. All 96 patients in the hysteroscopy group successfully completed diagnostic hysteroscopy. Visually directed biopsy of suspected areas was subsequently performed and used as the reference standard. Diagnostic curettage was performed using the fractional curettage approach. Nineteen of the 81 patients in the diagnostic curettage group had insufficient tissue and were classified as having negative results. Except for 5 patients who were diagnosed with endometrial malignancy and were immediately admitted for treatment, the remaining 76 patients in the diagnostic curettage group all underwent subsequent hysteroscopy directed biopsy as the reference standard. All tissue samples were sent for pathologic assessment. The five outcomes of interest for this study were hyperplasia and EIN combined, endometrial

malignancy, polyp, submucosal myoma, and endometritis.

Karageyim Karsidag et al.[26]

Karageyim Karsidag et al.[26] prospectively evaluated 36 postmenopausal patients (46–70 years) with postmenopausal bleeding and endometrial thickness >5 mm on TVS who had previously undergone blind D&C with normal histologic results but experienced recurrent or persistent AUB. All patients sequentially underwent TVS, saline infusion sonohysterography, and diagnostic hysteroscopy followed by hysteroscopy-directed biopsy. Each diagnostic modality was performed by a different experienced operator blinded to the results of the other tests to minimise diagnostic bias. Hysteroscopy-directed biopsy histopathology served as the reference standard. Endometrial polyps and submucosal myomas were considered pathologic outcomes, while atrophic, proliferative, secretory endometrium and hyperplasia without atypia were classified as normal. Malignancy and atypical hyperplasia were exclusion criteria and were not evaluated as outcomes. Sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratios, and diagnostic accuracy were calculated in the study for benign intracavitary lesions analyzed as a combined outcome relative to the reference standard, however the raw data was presented separately for polyps and myomas in the tables. In this study, ten patients had "insufficient material" on blind D&C and were classified as negative results for the purposes of diagnostic accuracy analysis. The two outcomes of interest and had separate data available were endometrial polyps and submucosal myomas.

APPENDIX C. Summary of QUADAS-C risk of bias domains for the included studies.

In the patient selection domain, both Leone et al.[23] and Angioni et al.[24] prospectively enrolled cohorts of women with abnormal uterine bleeding and were determined to be at low risk of bias. In contrast, Gao et al.[25] was determined to have some concerns in this domain because of its retrospective design and the non-random allocation of patients to hysteroscopy or diagnostic curettage, introducing potential spectrum and selection bias. Karageyim Karsidag et al.[26] similarly showed some concerns due to incomplete reporting of enrollment procedures and eligibility criteria.

For the index test domain, two studies (Leone et al.,[23] Angioni et al.[24]) were determined to have some concerns. Neither study explicitly stated that interpretation of hysteroscopic findings or blind sampling results was performed blinded to the reference standard. In addition, diagnostic criteria for hysteroscopic interpretation were not always prespecified, increasing the risk of interpretive bias. This limitation is particularly relevant for visually based tests such as hysteroscopy.

For the reference standard domain, Leone et al.,[23] where dilation and curettage served as a uniform histopathologic reference, was of low risk of bias. Angioni et al.[24] had some concerns for differential verification bias because reference standards varied according to pathology (operative hysteroscopy for benign lesions and hysterectomy for premalignant or malignant disease). Gao et al.[25] used hysteroscopy-directed biopsy as the reference standard but stated it was performed for areas with abnormal findings from visualisation. It is unclear how pathology was obtained for patients without obvious lesions under visualisation, resulting in some concerns. Karageyim Karsidag et al.[26] similarly showed some concerns due to limited reporting on histologic verification.

The flow and timing domain was an area of bias for all studies. Leone et al.[23] had high-risk of bias because hysteroscopic biopsy was performed immediately before dilation and curettage, creating a risk that the reference standard was influenced by prior tissue removal (partial verification and incorporation bias). Angioni et al.[24] and Gao et al.[25] also demonstrated some concerns regarding whether all patients underwent the same reference standard and some had repeated testing or staged verification depending on initial results.

Across studies, comparability between index tests was limited. In Gao et al.,[25] hysteroscopy and diagnostic curettage were performed in different patient groups rather than in a paired design, increasing the risk of confounding. In Angioni et al.[24] and Leone et al.,[23] although patients underwent both tests, the sequence and conditional verification introduced differential verification bias favoring hysteroscopy.

Applicability concerns were low for patient populations and index tests, as all studies evaluated women with clinically relevant abnormal uterine bleeding in settings consistent with current practice. However, applicability of reference standards was determined to have some concerns in studies relying on hysteroscopy-directed biopsy rather than hysterectomy or full curettage, particularly for diffuse or multifocal disease.

APPENDIX D. Summary of relevant studies that were excluded.

Seven additional studies were identified that, although relevant to the clinical question, did not meet the predefined methodological criteria for inclusion.[27-33] Five of these studies used the same modality as both the index text and the reference standard, introducing incorporation bias and precluding independent evaluation of diagnostic performance.[28-30,32,33] Three studies reported aggregated diagnostic metrics across multiple pathologic entities without stratifying results by individual pathology, thereby limiting the ability to interpret condition-specific test performance.[29,31,33] One study exhibited both methodological limitations, reporting global outcomes while simultaneously employing the index test as the reference standard.[33]

In five of these studies, hysteroscopic visual inspection served as the index test, with clinicians providing real-time assessments of endometrial pathology followed by blind D&C for histologic confirmation.[28-30,32,33] Although these studies offer insight into clinical practice patterns, the reliance on blind curettage as the reference standard introduces important limitations, as blind sampling may miss focal lesions and therefore cannot reliably serve as a definitive comparator. As such, these studies contribute contextual perspective but cannot reliably inform pooled estimates of diagnostic accuracy.