Accuracy of hysteroscopy in the diagnosis of endometrial cancer and hyperplasia: a systematic quantitative review


Authors' objectives
To determine the accuracy of hysteroscopy in diagnosing endometrial cancer and hyperplasia in women with abnormal uterine bleeding, and to explore reasons for the heterogeneity of the outcomes.

Searching
MEDLINE and EMBASE were searched from January 1984 to December 2001, employing the MeSH and textwords for the term 'hysteroscopy' combined with the MeSH term 'diagnosis'. Studies reported in any language were considered. In addition, the Cochrane Library was searched, the reference lists of all known reviews and primary studies were examined, and the manufacturers of hysteroscopes were contacted.

Study selection
Study designs of evaluations included in the review
Observational studies in which the results of the diagnostic test of interest were compared with those of a reference standard (diagnostic accuracy studies) were eligible for inclusion in the review.

Specific interventions included in the review
Studies of hysteroscopy (direct endoscopic visualisation of the endometrial cavity) in the diagnosis of endometrial cancer and hyperplasia were eligible for inclusion in the review.

Reference standard test against which the new test was compared
The reference standard was endometrial histological findings. These were determined by either out-patient biopsy, dilation of the cervix and curettage of the endometrium, hysterectomy specimen, or by directed biopsy. The verification of diagnosis following hysteroscopy was performed either at the same time (simultaneous) or after a short delay (sequential).

Participants included in the review
Studies of women with abnormal premenopausal or postmenopausal uterine bleeding were eligible for inclusion in the review. The participants also included women taking hormone replacement therapy. Postmenopausal women represented 29% of the participants in the included studies.

Outcomes assessed in the review
The primary outcome measure was the accuracy with which endometrial cancer and hyperplasia were diagnosed. The secondary outcomes were failed procedures and major complications. A failed procedure was defined as the presence of technical problems (e.g. cervical stenosis, anatomic factors, structural abnormalities), inadequate visualisation (e.g. obscured by bleeding, debris) or patient factors (e.g. pain, intolerance), which prevented a final diagnosis.

How were decisions on the relevance of primary studies made?
The studies were identified by two reviewers independently, who were blinded to the author and title. The final inclusion and exclusion decisions were made with reference to a checklist, which consisted of items based upon the selection criteria. This checklist was piloted and the repeatability of its use was tested and confirmed. Any disagreements about study inclusion and exclusion were initially resolved by consensus, and when this was not possible, they were resolved by arbitration with a third reviewer. The agreement statistics among reviewers were computed.

Assessment of study quality
The methodological quality of the studies was assessed on the basis of study design (e.g. method of data collection) and relevant features relating to population (e.g. patient selection), description of the diagnostic test and histological
reference standard, the presence of verification bias (completeness and timing of verification by reference standard) and blinding. Five levels of study quality were also assigned. These ranked studies according to independence, blind comparison with reference standard, nature of population (e.g. whether consecutive, nonconsecutive, or ‘narrow’), and whether the reference standard was applied to all study participants. Studies of level 1 to 3 were considered to be of a high quality, while those ranked level 4 to 5 were of a low quality. English language articles were assessed by one reviewer, while foreign language articles were assessed by two independent reviewers following translation (when necessary). Any disagreements were resolved by consensus.

**Data extraction**
The authors do not state how the data were extracted for the review, or how many of the reviewers performed the data extraction.

Data were extracted on: study author and year, population characteristics (numbers, menopausal status, whether taking hormone replacement therapy), study quality level, reference standard, timing or completeness of verification, and percentage follow-up. In addition, the setting (in- or out-patient) and technical details pertaining to the hysteroscopic examination were also extracted. Endometrial cancer was the primary outcome; the data were extracted as 2x2 tables of the hysteroscopy results (benign or malignant) and the histological results. Similar contingency tables were produced for hysteroscopy result and endometrial disease (benign or disease). Failure rates were recorded, but were excluded from the tabulated data. The sensitivity, specificity and likelihood ratios (LRs) were calculated for each study, along with the 95% confidence intervals (CIs). When the tables contained 0 cells, 0.5 was added to each cell.

**Methods of synthesis**
How were the studies combined?
If the measures of sensitivity and specificity were found to be independent, as indicated by a lack of statistical correlation among them, the studies were combined statistically in a meta-analysis to produce summary pooled estimates of sensitivity and specificity. The summary LRs were calculated as the principal measure of diagnostic accuracy. If heterogeneity was found by a subgroup analysis, the results from the individual studies were initially pooled using both fixed-effect and random-effects models. The results obtained with a fixed-effect model are reported. If heterogeneity remained within the pre-specified clinical subgroups, it was decided to base inferences on the high-quality studies. Publication and related biases were explored through a funnel plot of diagnostic odd ratios (DORs) against the corresponding SEs. The correlation between estimated DORs and their SEs was tested using the adjusted rank method.

How were differences between studies investigated?
Heterogeneity was investigated using sensitivity and specificity plots and the chi-squared test. To explore clinical sources of heterogeneity, the potential explanatory variables were defined a priori. Regression analysis using the log DOR was performed to investigate heterogeneity. Univariable analyses were initially performed followed by multivariable modelling. The variables included in the regression analyses were spectrum variability (menopausal status and setting) and study quality (patient selection, reference standard, verification bias and blinding); these were also adjusted for overall quality level. The models produced by multivariable analysis included only three variables: menopausal status, clinical setting and quality as a binary variable (levels 1 to 3 versus 4 to 5). Additional post-hoc analyses were performed when certain variables were considered informative or were recommended by peer reviewers, although the findings were considered in the context of hypothesis generation.

A sensitivity analysis was also performed; this considered inadequate histological specimens, precluding a definitive diagnosis following the reference test, as negative results. Intra-uterine polyps and fibroids were also excluded as part of a sensitivity analysis to examine whether the presence of these focal lesions affected estimates of diagnostic accuracy.

**Results of the review**
Sixty-five studies (n=26,346) were included in the review. Fifty-six (n=24,649) assessed the diagnosis of endometrial cancer and 41 looked at the diagnosis of endometrial hyperplasia.
Endometrial cancer (56 studies).

There was no significant association between sensitivity and specificity (correlation, r = 0.06; P = 0.65). Weighted by the number of cases, the overall sensitivity was 86.4% (95% CI: 84.0, 88.6) and the specificity was 99.2% (95% CI: 99.1, 99.3). The pooled LR for all studies was 60.9 (95% CI: 51.2, 72.5). The pre-test probability (prevalence) increased from 3.9% (95% CI: 3.7, 4.2) to 71.8% (95% CI: 67.0, 76.6) with a positive result, and decreased to 0.6% (95% CI: 0.5, 0.8) with a negative result. Heterogeneity of diagnostic performance among studies was significant. Heterogeneity was not explained by either study setting, menopausal status or study quality. The other potential explanatory variables defined post hoc did not significantly influence diagnostic accuracy. Further results of the different heterogeneity tests and sensitivity analyses carried out were reported and discussed in the review.

Endometrial disease.

There was no significant association between sensitivity and specificity (r = 0.05). The weighted overall sensitivity was 78.0% (95% CI: 76.3, 79.6) and the specificity was 95.8% (95% CI: 95.6, 96.1). The pre-test probability increased from 10.6% (95% CI: 10.2, 11.0) to 55.2% (95% CI: 52.4, 57.8) with a positive result, and decreased to 2.8% (95% CI: 2.4, 3.0) with a negative result. Heterogeneity was found in the overall and subgroup meta-analysis. Clinical setting and menopausal status were significant explanatory variables in the univariable analysis, as was the quality item of patient selection. Poor study quality, out-patient setting and postmenopausal status were associated with the significantly higher accuracy of hysteroscopy.

In 19 of the 65 studies reporting serious complications, 8 potentially serious complications were reported out of 9,413 successful procedures.

The rank correlation tests for publication and related biases found that funnel plot asymmetry was not statistically significant (P = 0.34 for endometrial cancer; P = 0.12 for endometrial disease).

Authors’ conclusions
Diagnostic hysteroscopy is safe, with a low incidence of serious complications and a small failure rate. The diagnostic accuracy of hysteroscopy is high for endometrial cancer, but only moderate for endometrial disease (cancer or hyperplasia).

CRD commentary
The review question and the study selection criteria were stated clearly. The literature review seemed reasonably comprehensive, with attempts made to identify additional studies to those found in the electronic searches. In addition, no language restrictions were employed. Sufficient information was provided on the literature selection and validation processes, but not the data extraction. The authors provide ample information on, and justification for the different statistical tests employed, which seem to have been appropriate for the range of analyses they have undertaken. The results and their implications were well presented and discussed, and were clear.

The authors’ conclusions seem appropriate in the light of the data presented and discussed

Implications of the review for practice and research
Practice: The authors state that hysteroscopy is highly accurate, and thereby clinically useful in diagnosing endometrial cancer in women with abnormal uterine bleeding. It is moderately useful in diagnosing endometrial disease.

Research: The authors did not state any implications for further research.

Funding
University of Birmingham Interdisciplinary Research Fund; the Birmingham Women’s Hospital Research and Development Programme.
Bibliographic details

PubMedID
12350192

Original Paper URL
http://jama.ama-assn.org/

Indexing Status
Subject indexing assigned by NLM

MeSH
Endometrial Hyperplasia /diagnosis /pathology; Endometrial Neoplasms /diagnosis /pathology; Female; Humans; Hysteroscopy; Predictive Value of Tests; Sensitivity and Specificity; Uterine Hemorrhage /etiolog

AccessionNumber
12002008556

Date bibliographic record published
31/03/2003

Date abstract record published
31/03/2003

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.