Authors' objectives
To determine the accuracy of endometrial thickness measurement by pelvic ultrasonography in the diagnosis of endometrial carcinoma and disease (hyperplasia and/or carcinoma) in women presenting with postmenopausal bleeding.

Searching
MEDLINE and EMBASE were searched for studies published between 1966 and 2000; the search terms were reported. In addition, the reference lists of all known primary and review articles were checked for relevant citations.

Study selection
Study designs of evaluations included in the review
The inclusion criteria pertaining to study design were not stated clearly, but it appears that diagnostic cohort studies were sought.

Specific interventions included in the review
To be included in the review, studies had to measure endometrial thickness using ultrasound imaging. One or both layers of the endometrium were measured; the cut-offs for an abnormal test ranged from less than or equal to 2 mm to less than or equal to 10 mm for single-layer measurement, and from 3 to 15 mm for the measurement of both layers. The transducer frequency ranged from 3.5 to 7.5 MHz.

Reference standard test against which the new test was compared
The included studies were required to confirm the diagnosis histologically. The reference standards used in the included studies were grouped as follows: examination of hysterectomy specimens or direct biopsy under hysteroscopic vision; examination of specimens from blind dilation and curettage (D-C) under general anaesthesia; or out-patient endometrial biopsy. The majority of studies included in the review used D-C, or D-C and endometrial biopsy.

Participants included in the review
To be included in the review, studies had to recruit women with postmenopausal bleeding. The length of amenorrhoea ranged from 6 to 24 months, although it was largely unreported in the studies. Where reported, the proportion of hormone replacement therapy (HRT) users ranged from 0 to 100%.

Outcomes assessed in the review
No clear inclusion criteria were specified in terms of the test outcome. For each study, a 2x2 table of the diagnostic test result and endometrial pathology was constructed. Likelihood ratios (LRs) were then calculated, with 95% confidence intervals (CIs). Data for endometrial carcinoma and endometrial disease (carcinoma or hyperplasia) were presented separately.

How were decisions on the relevance of primary studies made?
Two reviewers independently and blindly assessed the English language articles for inclusion. Any disagreements were resolved by consensus, and the kappa statistic was used to assess inter-observer agreement. It appears that non-English language articles were assessed by one reviewer only.

Assessment of study quality
Methodological quality was assessed on the basis of the following: how the population was recruited (consecutive or arbitrary); the length of amenorrhoea; population spectrum of HRT use; adequacy of description of the diagnostic test scanning method (transvaginal or transabdominal method of obtaining image and frequency of transducer); a priori determination of cut-off point; blinding; and completeness of follow-up. These criteria were used to sort the studies into
five quality groups. Two reviewers independently and blindly assessed the methodological quality of all the included English language articles. Any disagreements were resolved by consensus, and the kappa statistic was used to assess inter-observer agreement. It appears that non-English language articles were assessed by one reviewer only.

Data extraction

Two reviewers independently extracted the data and any disagreements were resolved by consensus. For each study, a 2x2 table of the diagnostic test result and endometrial pathology was constructed.

Methods of synthesis

How were the studies combined?
The LRs were pooled in a meta-analysis, weighting the log LR from each study in inverse proportion to its variance. The authors then calculated pre- and post-test probabilities of endometrial pathology (for a negative and a positive test result, using Bayes’ theorem), along with 95% CIs. These calculations were performed separately for the outcomes of endometrial carcinoma and endometrial disease. The meta-analysis was stratified by cut-off level and measurement technique (one or two endometrial layers).

How were differences between studies investigated?
Heterogeneity was assessed using the chi-squared test. Where heterogeneity was found, sensitivity analyses were conducted based on study characteristics and study methodological quality. An additional publication based on a subset of studies included in this review assessed the effects of delayed verification on estimates of accuracy (see Other Publications of Related Interest).

Results of the review

Fifty-seven studies were included in the review. It was unclear how many participants were recruited in these studies. Inter-observer agreement was high. For study eligibility it was 96% (kappa 0.91) and for the various components of study quality it ranged from 89 to 100% (kappa: 0.64 to 1.0).

The majority of the studies did not score highly on methodological quality. Most did not report whether participant recruitment was consecutive. Less than half of the studies included patient spectrum (with respect to HRT use). The diagnostic test was quite well described, but over half of the studies using a cut-off of less than or equal to 4 mm or less than or equal to 5 mm determined this cut-off point post hoc. Few studies were blinded. However, most did verify the diagnosis, and most described what the review authors considered to be an ideal reference standard.

The commonest cut-off levels used in the diagnostic tests were based on the measurement of both endometrial layers. These cut-offs were less than or equal to 4 mm (9 studies for endometrial carcinoma; 9 studies for endometrial disease) and less than or equal to 5 mm (21 studies for endometrial carcinoma; 19 studies for endometrial disease); hereafter referred to as cut-off A and cut-off B, respectively.

The pre-test probability of endometrial carcinoma was 14% (95% CI: 13.3, 14.7). A negative test result reduced the post-test probability of carcinoma to 1.2% (95% CI: 0.4, 2.9) at cut-off A and 2.3% (95% CI: 1.2, 4.8) at cut-off B. Conversely, a positive test result increased the post-test probabilities of carcinoma to 24.2% (95% CI: 19.7, 29.2) and 26.1% (95% CI: 21.1, 31.6), respectively. The LR estimates from the cut-off A studies did not show evidence of significant heterogeneity, although none of these studies were of a good quality. The pooled estimates of LRs for cut-off B studies were heterogeneous, and sensitivity analyses showed no explanation. When the analysis was restricted to just the four best-quality studies, the negative test result reduced the post-test probability of carcinoma to 2.5% (95% CI: 0.9, 6.4).

The pre-test probability of endometrial disease was 26% (95% CI: 25, 27). A negative test result reduced the post-test probabilities of disease to 2.4% (95% CI: 1.3, 3.9) at cut-off A and 5% (95% CI: 2.9, 9.1) at cut-off B. A positive test result increased the post-test probabilities of disease to 43.3% (95% CI: 36.6, 46.7) and 47.9% (95% CI: 40.4, 55.6), respectively. The LR estimates from the cut-off A studies did not show evidence of significant heterogeneity, but none of these studies were of a good quality. Again, heterogeneity in the result for cut-off B studies could not be explained.
by sensitivity analyses. Using the pooled estimate from the four best-quality studies only, a negative test result reduced
the post-test probability of disease to 2.7% (95% CI: 0.9, 6.9).

Further analyses were reported in the paper. The additional analysis published as a separate report assessed the effects
of delayed verification. This analysis was restricted to the 15 included studies that included a reference standard
examination obtained by an independent endometrial sampling technique and that provided explicit information on the
time between the index test and reference standard. The pooled diagnostic odds ratio for studies that reported
immediate verification (<24 hours between tests) was 30.6 (95% CI: 9.1, 102.6), compared with 15.6 (95% CI: 7.1,
34.1) for studies that reported delayed verification (>24 hours between tests). Sensitivity ranged from 88 to 100% in
studies that reported immediate verification and from 67 to 100% in those that reported delayed verification.
Specificity ranged from 31 to 83% in studies that reported immediate verification and from 39 to 77% among those
that reported delayed verification.

Authors’ conclusions
Ultrasonic measurement of endometrial thickness is of limited diagnostic use in the prediction of endometrial
hyperplasia or carcinoma, but it is a good test for the exclusion of endometrial pathology.

CRD commentary
The review objectives and literature search were adequate, and the inclusion criteria were generally clear. Efforts were
made to reduce the possibility of bias by performing the eligibility assessment and methodological quality assessment in
duplicate and blinded for most papers. In addition, the data extraction was also conducted in duplicate. Overall, the
reported inter-observer agreement was high.

The methodological quality of the included studies was assessed quite comprehensively, but some relevant aspects were
not investigated. For example, there was no mention of how the different reference standard tests were picked and
when they were performed (in relation to the diagnostic test under study and the start of treatment). Furthermore, it was
not explicit how the studies were grouped into the five quality categories; it seems that the review authors did not
consider the use of post hoc cut-offs when grouping for quality. The review authors chose to pool the results of each
study, despite evidence of heterogeneity for some of the results. It was unclear whether the sensitivity analyses
conducted to investigate this heterogeneity were pre-planned. These analyses did not explain the heterogeneity. The
question remains as to how meaningful the combined result is and how reliable an estimate of post-test probability is,
based upon such a result. The review authors clearly described the limitations of the included studies, but when drawing
their conclusions they slightly overstated the results. The conclusions appeared to be based on the results of the four
studies judged to be of the best methodological quality. However, this did not appear to have been a pre-planned
subgroup analysis, and the authors did not specify how they determined the quality cut-off. Despite these studies being
described as ‘best quality’, only one reported consecutive recruitment, the transducer frequencies were different in each
study, only three studies were blinded, and one used a different reference standard to the other three. It was also unclear
if the four ‘best quality’ studies used a post hoc diagnostic cut-off. With such differences and limitations in the
individual studies, it is difficult to judge how reliable the combined result is.

Implications of the review for practice and research
Practice: The authors stated that the ultrasonic measurement of endometrial thickness alone could not rule in
endometrial disease. However, they suggested that by examining both layers, a cut-off of less than or equal to 5 mm
could be used to rule out endometrial disease with good certainty. They then went on to caution that ruling out
endometrial cancer is very important, and they would be wary of reliance on pooled estimates from only four studies.

Research: The authors stated that there was an urgent need for better quality primary accuracy studies using ideal
reference standards to guide decision-making. They also stated that further meta-analyses should focus on the inclusion
of more recent publications and approaches using individual patient data, thus allowing more robust subgroup analyses.

Bibliographic details

PubMedID
12225294

Other publications of related interest

Indexing Status
Subject indexing assigned by NLM

MeSH
Endometrial Hyperplasia /pathology /ultrasonography; Endometrial Neoplasms /pathology /ultrasonography; Endometrium /pathology /ultrasonography; Female; Humans; Likelihood Functions; Postmenopause; Research Design /standards; Uterine Hemorrhage /etiology

AccessionNumber
12002002148

Date bibliographic record published
31/08/2005

Date abstract record published
31/08/2005

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.