The diagnostic work-up of women with postmenopausal bleeding
van Hanegem, N.

Citation for published version (APA):
van Hanegem, N. (2015). The diagnostic work-up of women with postmenopausal bleeding

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 7

The accuracy of endometrial sampling in women with postmenopausal bleeding: a systematic review and meta-analysis

N. van Hanegem, M.M.C. Prins, M.Y. Bongers, B.C. Opmeer, D.S. Sahota, B.W.J. Mol & A. Timmermans

European Journal of Obstetrics and Gynecology and Reproductive Biology, accepted with minor revisions
Chapter 7

Abstract

Postmenopausal bleeding (PMB) can be the first sign of endometrial cancer. In case of thickened endometrium, endometrial sampling is often used in these women. In this systematic review, we studied the accuracy of endometrial sampling for the diagnoses of endometrial cancer, atypical hyperplasia and endometrial disease (endometrial pathology, including benign polyps).

We systematically searched the literature for studies comparing the results of endometrial sampling in women with postmenopausal bleeding with two different reference standards: blind dilatation and curettage (D&C) and hysteroscopy with histology. We assessed the quality of the detected studies by the QUADAS-2 tool. For each included study, we calculated the fraction of women in whom endometrial sampling failed. Furthermore, we extracted numbers of cases of endometrial cancer, atypical hyperplasia and endometrial disease that were identified or missed by endometrial sampling.

We detected 12 studies reporting on 1029 women with postmenopausal bleeding: five studies with dilatation and curettage (D&C) and seven studies with hysteroscopy as a reference test. The weighted sensitivity of endometrial sampling with D&C as a reference for the diagnosis of endometrial cancer was 100% (range 100-100%) and 92% (71-100%) for the diagnosis of atypical hyperplasia. Only one study reported sensitivity for endometrial disease, which was 76%. When hysteroscopy was used as a reference, weighted sensitivities of endometrial sampling were 90% (range 50-100%), 82% (range 56-94%) and 39% (21-69%) for the diagnosis of endometrial cancer, atypical hyperplasia and endometrial disease, respectively. For all diagnoses studied and the reference test used, specificity was 98-100%. The weighted failure rate of endometrial sampling was 11% (range 1-53%), while insufficient samples were found in 31% (range 7-76%). In these women with insufficient or failed samples, an endometrial (pre) cancer was found in 7% (range 0-18%).

In women with postmenopausal bleeding, the sensitivity of endometrial sampling to detect endometrial cancer and especially atypical hyperplasia and endometrial disease, including endometrial polyps, is lower than previously thought.
The accuracy of endometrial sampling

Introduction

Postmenopausal bleeding (PMB) is one of the most frequent complaints with which women present in the outpatient gynaecology clinic. As PMB might be the first sign of endometrial cancer, accurate diagnostic work-up is necessary in these women. Despite many studies on the different diagnostic measures in women with PMB, there is no consensus on the best diagnostic pathway.\textsuperscript{1,4} In many guidelines the measurement of endometrial thickness by transvaginal sonography (TVS) is used as a first step in the diagnostic pathway to distinguish women with a low and a high risk of having endometrial cancer. Clark et al found that a strategy with TVS as the initial test with a cut-off of 4 mm followed by endometrial sampling was the most cost-effective.\textsuperscript{5} In situations where ultrasound is not directly available, endometrial sampling can be used as the first step.\textsuperscript{6}

The meta-analysis by Dijkhuizen et al was the first meta-analysis on the diagnostic accuracy of endometrial sampling in women with postmenopausal bleeding.\textsuperscript{7} Several years after that, two other meta-analyses were published.\textsuperscript{8,9} These meta-analyses found that sensitivity, which is crucial to rule out endometrial cancer, was around 99%. However, in these studies (blind) dilatation and curettage (D&C) had been used as reference standard. Nowadays, D&C is almost completely replaced by hysteroscopy as a reference standard.\textsuperscript{10} Also, only a small proportion of women in these meta-analyses was postmenopausal.

In view of this, we decided to conduct a systematic review and meta-analysis to study the diagnostic accuracy of endometrial sampling in women with PMB regarding the diagnoses of endometrial cancer and atypical hyperplasia compared to two different reference standards: blind D&C and the current reference standard: hysteroscopy with histology or hysterectomy.\textsuperscript{10}

Methods

Identification of studies

In April 2015, we performed a computerized search in MEDLINE, EMBASE and Science Direct® to identify all studies on the diagnostic accuracy of endometrial sampling published between January 1965 and March 2015. The search was limited to studies in humans; language restrictions were not applied. We used all known synonyms for the following keywords: postmenopausal bleeding AND endometrial
sampling. We included observational studies on the evaluation of the diagnostic accuracy of endometrial sampling in women with PMB. References cited in the selected articles were checked for further relevant articles not identified by the electronic searches. The search strategy can be found in the Appendix.

Selection criteria
This review focused on diagnostic studies in which the histology results of endometrial sampling were compared with the results of a reference standard. The articles had to study women with postmenopausal uterine bleeding, the diagnostic test of interest was endometrial sampling (histology), the reference standard had to be endometrial histological findings from (blind) D&C, diagnostic hysteroscopy with histology by targeted biopsy or D&C or hysterectomy.

Identified articles were merged into a common file, duplicates were deleted, and results were divided between two reviewers (NvH and MMP) who independently examined the assigned articles and classified each as “exclude”, “include”, or “unsure.” Initial screening began with a title screen. Subsequently, abstracts were retrieved and screened to determine eligibility. Finally, full text articles were retrieved and screened for inclusion. A third reviewer (MB) settled discrepancies. For articles, which included both pre- and postmenopausal women, but did not report separately on the postmenopausal group, we sent an email to the corresponding author to ask for the data on postmenopausal women. For articles which were published before 1997 and therefore no email address of the corresponding author was mentioned, we searched the internet (Google, PubMed) for an email address to contact the corresponding author. We calculated the agreement on the selection of studies between the reviewers.

Quality assessment
Two reviewers (NvH and MMP) independently assessed the methodological quality of each selected paper using the QUADAS-2 tool for diagnostic studies, modified to conform to this review. Disagreements were resolved via consensus and if necessary via a third reviewer (MB).

We decided a priori the criteria of each study for low risk of bias in each of the four main domains of the Quadas-2 tool: patient selection, index test, reference standard, and flow and timing. For patient selection, the in- and exclusion criteria had to be clearly stated, and the patient sample had to be consecutive. For the index test, the independent assessment of the pathologist for endometrial sampling
The accuracy of endometrial sampling

without knowledge of the results of the reference test had to be clearly stated, and the histology results had to be pre-specified. For the reference standard (D&C or hysteroscopy) it had to be clearly stated that results were interpreted without knowledge of the result of endometrial sampling. For patient flow and timing, the time between endometrial sampling and reference test, if all patients received the same reference test and if all patients were included in the analysis, had to be clearly stated. Applicability was based on patients with PMB, endometrial sampling as the index test, D&C or diagnostic hysteroscopy with histology as reference test.

For all articles, each domain was assessed in terms of risk and bias, and the first three domains were also assessed in terms of applicability for this review. Each item was labelled 'low', 'high', or 'unclear'. Studies, which scored 'high bias' on more than one of four items, were excluded. And only studies, which scored 'low' on all three items of concerns on applicability, were included in this review. We included all applicable studies on this subject, regardless of the number of postmenopausal women included and regardless if data were collected prospectively or retrospectively.

Data extraction
For studies, which included pre- and postmenopausal women, we used only those calculations and conclusions concerning the latter. From each article we extracted (if available): the reference standard that was used, the number of women who underwent endometrial sampling, the number of women in whom endometrial sampling was not possible, failed or showed insufficient material for a pathologic diagnosis, the number of women who underwent both endometrial sampling and the reference standard, the number of cases with endometrial cancer; atypical hyperplasia or endometrial disease. Hyperplasia without atypia was considered a benign result. Endometrial disease was defined as benign endometrial polyps in one study, and as polyps, hyperplasia and cancer together in most other studies. For this meta-analysis we decided to define endometrial disease as endometrial cancer; atypical hyperplasia and benign endometria polyps together as endometrial disease.

Data analysis
For each study, we calculated the percentage of women in whom endometrial sampling failed to provide a diagnosis, either due to the possibility to obtain tissue (for example, because of cervical stenosis) or due to the fact that the sample that was obtained was insufficient for the pathologist to establish a diagnosis. We described the number of endometrial cancers in women with a failed endometrial sampling.
For studies that had numbers available, we constructed 2 x 2 tables and calculated sensitivity and specificity for the diagnosis of endometrial cancer, atypical hyperplasia or endometrial disease. Sensitivity and specificity were calculated for the cases in which both endometrial sampling as well as the reference test was successful. We calculated the weighted sensitivity, taking into account the size of each study, compared to the two different reference strategies. When a 2x2 table could be constructed, we plotted the sensitivity against the ‘1 - specificity’ in a receiver-operating curve (ROC).

Results

Study selection
Our systematic search identified 499 titles. After exclusion of studies, which did not exist online anymore and exclusion of duplicates, we identified 377 articles, of which 65 articles were found to be relevant (Figure 1). After reading these 65 articles in full-text, we could include 11 studies that reported on postmenopausal women only, 2 articles that described data on postmenopausal women separately in a total population of perimenopausal women and 17 articles that compared the results of endometrial sampling with histology findings from (blind) D&C or diagnostic hysteroscopy in a combined population of pre- and postmenopausal women.

In none of the 17 studies that reported on a combination of pre- and postmenopausal women, we were able to contact the corresponding author. In 10 studies these authors did not respond, while contact details were not available for the other 7 studies. Therefore, we had to exclude these 17 studies from the meta-analysis. The initial agreement of the two reviewers (NvH and MMP) regarding eligibility was 94% (weighted kappa 0.88 (95% CI 0.76-0.99)).
The accuracy of endometrial sampling

Figure 1. Study selection flowchart
* The reference list for excluded studies is available from the corresponding author.

Search MEDLINE
n=223

Search EMBASE
n=231

Search Science Direct
n=13

Cross-references
n=32

122 studies excluded:
  7 studies not available
  115 duplicates

116 studies selected
based on title

65 studies selected
based on abstract

35 studies excluded*:
  4 patients not with PMB
  5 review article/letter
  2 on cytology instead of histology
  10 no reference standard
  14 not on diagnostic accuracy end sampling

30 studies eligible for inclusion:
  13 data on PMP women
  17 data combined pre/post menopausal

17 studies excluded:
  10 authors no email available
  7 authors contacted: no response

1 study on PMP women excluded after quality assessment

12 studies included
Chapter 7

Table 1. Study characteristics of the twelve included studies on the diagnostic accuracy of endometrial sampling

<table>
<thead>
<tr>
<th>Study design</th>
<th>Patient selection</th>
<th>Index test</th>
<th>Reference standard</th>
<th>Menopausal status</th>
<th>PMB n</th>
<th>HRT %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldberg 1982</td>
<td>Prospective</td>
<td>Unclear</td>
<td>Vabra &amp; Accurette</td>
<td>Blind D&amp;C</td>
<td>Post</td>
<td>40</td>
</tr>
<tr>
<td>Batool 1994</td>
<td>Prospective</td>
<td>Unclear</td>
<td>Pipelle</td>
<td>Blind D&amp;C</td>
<td>Post</td>
<td>70</td>
</tr>
<tr>
<td>Ben-Baruch</td>
<td>Retrospective</td>
<td>Unclear</td>
<td>Pipelle</td>
<td>Blind D&amp;C</td>
<td>Pre &amp; post</td>
<td>90</td>
</tr>
<tr>
<td>vdBosch 1995</td>
<td>Prospective</td>
<td>Consecutive</td>
<td>Pipelle</td>
<td>Hysteroscopy w/ histology</td>
<td>Post</td>
<td>140</td>
</tr>
<tr>
<td>vdBosch 1996</td>
<td>Prospective</td>
<td>Consecutive</td>
<td>Pipelle</td>
<td>Hysteroscopy w/ histology</td>
<td>Post</td>
<td>87</td>
</tr>
<tr>
<td>Giusa-Chifieri 1996</td>
<td>Prospective</td>
<td>Unclear</td>
<td>Novak</td>
<td>Hysteroscopy w/ histology</td>
<td>Post</td>
<td>80</td>
</tr>
<tr>
<td>Gupta 1996</td>
<td>Prospective</td>
<td>Consecutive</td>
<td>Pipelle</td>
<td>Hysteroscopy w/ histology</td>
<td>Post</td>
<td>76</td>
</tr>
<tr>
<td>De Silva 1997</td>
<td>Prospective</td>
<td>Consecutive</td>
<td>Pipelle</td>
<td>Hysteroscopy w/ histology</td>
<td>Post</td>
<td>50</td>
</tr>
<tr>
<td>Mortakis 1997</td>
<td>Not reported</td>
<td>Unclear</td>
<td>Pipelle</td>
<td>Hysteroscopy w/ histology</td>
<td>Pre &amp; post</td>
<td>78</td>
</tr>
<tr>
<td>Bunyavejchevin 2001</td>
<td>Prospective</td>
<td>Unclear</td>
<td>Pipelle</td>
<td>Blind D&amp;C</td>
<td>Post</td>
<td>30</td>
</tr>
<tr>
<td>Epstein 2001</td>
<td>Prospective</td>
<td>Consecutive</td>
<td>Endorette</td>
<td>Blind D&amp;C</td>
<td>Post</td>
<td>133</td>
</tr>
<tr>
<td>Spicer 2006</td>
<td>Prospective</td>
<td>Consecutive</td>
<td>Accurette</td>
<td>Hysteroscopy w/ histology</td>
<td>Post</td>
<td>136</td>
</tr>
</tbody>
</table>

D&C=dilatation and curettage; PMB=number of women with postmenopausal bleeding; Nr=not reported

Quality assessment

Table 2 presents quality assessment of the included studies.\textsuperscript{12-24} Quality assessment showed in four studies (25%) a ‘low’ risk of bias on all four items, three studies showed a ‘high’ risk of bias on one of the items, while eight studies had an ‘unclear’ risk of bias on the description of methods on patient selection, the index or reference test. All studies, except for one scored ‘low’ on the three items of applicability.\textsuperscript{12-23}

Based on a high concern of applicability of the index test and reference standard described in O’Connell et al, we decided to exclude this study.\textsuperscript{24} After study selection and quality assessment, we included 12 articles in this systematic review, reporting on 1,029 women with postmenopausal bleeding (Table 1 and 2, Figure 2).
The accuracy of endometrial sampling

Table 2. Risk of bias and concerns of applicability by study using a modified Quadas-2 tool.11

<table>
<thead>
<tr>
<th>Study</th>
<th>Patient selection</th>
<th>Index test</th>
<th>Reference standard</th>
<th>Flow and timing</th>
<th>Patient selection</th>
<th>Index test</th>
<th>Reference standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldberg, '82</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Batool, '94</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Ben-Baruch, '94</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>vdBosch, '95</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>vdBosch, '96</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Giusa-Chifferi, '96</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Gupta, '96</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>De Silva '97</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Mortakis '97</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>O'Connell '97</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Bunyavejchevin '01</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Epstein '01</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Spicer '06</td>
<td>Low</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

Figure 2. Overall risk of bias and applicability using a modified Quadas-2 tool.11

<table>
<thead>
<tr>
<th>Patient selection</th>
<th>Index test</th>
<th>Reference test</th>
<th>Flow and timing</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0%</td>
<td>25%</td>
<td>50%</td>
<td>75%</td>
</tr>
</tbody>
</table>

Proportion of studies with low, high or unclear risk of bias and applicability

High
Unclear
Low

Diagnostic accuracy of endometrial sampling in women with PMB

Table 1 and 3 show the findings of the 12 included studies.12-23 The Pipelle® device was used in eight studies,12,13,15,17,20-22 while the other studies reported on the use of AccuRette®,23 Endorette®,18 and Novak endometrial sampler®.14 One study reported on two different sampling methods: AccuRette® and Vabra®.19

Blind D&C was used as the reference standard in five studies17-21 while hysteroscopy with histology (by biopsy and/or curettage) was the reference standard
in seven studies. In three studies the diagnosis of endometrial cancer detected by endometrial sampling was confirmed by hysterectomy and not by hysteroscopy or D&C.

All 12 studies reported on the fraction of women in whom endometrial sampling failed, mostly due to cervical stenosis. The failure rates of endometrial sampling varied between 1% and 53%, with a weighted failure rate of 11%. Eight studies reported on the fraction of women in whom insufficient material was found at histology, which varied between 7% and 76%, with a weighted insufficient rate of 31%. In the article by Batool et al. the rate of insufficient samples was much higher than in the other studies (42/55). In 37 of these women with an insufficient sample, material was also insufficient for diagnosis by D&C, which might explain the high insufficient-rate.

The weighted percentage of women with endometrial (pre) cancer among those who had failed or insufficient sampling is 7% (range 0%-18% in seven studies). Goldberg et al. described a percentage of 18% endometrial cancer in women with insufficient or failed samples. This article from 1982 lacked detail on the small number of women (n=12) included.

Diagnosis of endometrial cancer

From all 12 articles we could extract data on the sensitivity and specificity regarding the diagnosis endometrial cancer (Table 3). The sensitivity of endometrial sampling was 100% in all five studies using blind D&C, but varied between 50-100% in the seven studies using hysteroscopy with histology as a reference standard, with a weighted sensitivity of 90%. Specificity was 99-100% regardless of the reference standard that was used. Figure 3A shows an ROC plot of the performance of the 12 studies that allowed the calculation of both sensitivity and specificity.

Diagnosis of (pre)cancer of the endometrium

With respect to the diagnosis of endometrial (pre-) cancer, i.e. atypical hyperplasia or endometrial cancer we could calculate sensitivity and specificity from the data in all five studies using D&C as a reference and in four studies using hysteroscopy as a reference (table 3). The weighted sensitivity in studies using D&C was 92% (range 71-100%), whereas the weighted sensitivity in studies using hysteroscopy as a reference standard was 82% (range 56-94%). Specificity was 99-100% in all studies. Figure 3B shows an ROC plot of the performance of the twelve studies that allowed the calculation of both sensitivity and specificity.
Table 3. Feasibility and diagnostic accuracy of endometrial sampling

| Study and reference standard | Failed samples n (%) | Insufficient samples n (%) | (pre)cancer in failed/insufficient samples n (%) | Endometrial cancer | Cancer or atypical hyperplasia | Endometrial disease |
|-----------------------------|----------------------|----------------------------|-----------------------------------------------|-------------------|-------------------------------|----------------|||
| **Blind D&C**               |                      |                            |                                               |                   |                               |                ||
| Goldberg, '82               | 5/40 (13)            | 7/35 (20)                  | 2/12 (17)                                    | 3/3 (1.0)         | 25/25 (1.0)                   | 8/9 (0.89)     | nr |
| Vabra                       | 5/40 (13)            | 6/35 (17)                  | 2/11 (18)                                    | 3/3 (1.0)         | 26/26 (1.0)                   | 9/9 (1.0)      | nr |
| Batool, '94                 | 15/70 (21)           | 42/55 (76)                 | 0                                             | 3/3 (1.0)         | 8/8 (1.0)                     | nr             | nr |
| Ben-Baruch, '94             | 2/90 (2)             | 6/88 (16)                  | 0                                             | 9/9 (1.0)         | 36/36 (1.0)                   | 10/10 (1.0)    | nr |
| Bunyavejchevin '94          | 16/30 (53)           | 7/14 (50)                  | 1/23 (4)                                     | 2/2 (1.0)         | 5/5 (1.0)                     | 2/2 (1.0)      | nr |
| Epstein '01                 | 2/113 (16)           | 31/112 (28)                | nr                                            | nr                | 5/7 (0.71)                    | nr             | nr |
| **Hysteroscopy**            |                      |                            |                                               |                   |                               |                ||
| vdBosch, '95                | 2/106 (2)            | nr                         | 0                                             | 5/5 (1.0)         | 80/80 (1.0)                   | nr             | 20/39 (0.51) |
| vdBosch '96                 | 2/140 (1)            | nr                         | 0                                             | 6/6 (1.0)         | nr (0.99)                     | nr             | nr |
| Giusa-Chifieri, '96         | 8/80 (10)            | 6/72 (8)                   | 1/14 (7)                                     | 16/17 (0.94)      | 49/49 (1.0)                   | 17/18 (0.94)   | nr |
| Gupta, '96                  | 22/76 (29)           | nr                         | 0                                             | 3/3 (1.0)         | 51/51 (1.0)                   | nr             | nr |
| De Silva '97                | 9/50 (18)            | 6/41 (15)                  | 1/15 (7)                                     | 1/2 (0.5)         | 13/13 (1.0)                   | nr             | nr |
| Mortakis '97                | 2/78 (3)             | nr                         | 1 (7)                                        | 5/6 (0.83)        | 70/70 (1.0)                   | 8/11 (0.73)    | nr |
| Spicer '06                  | 15/136 (11)          | 65/136 (48)                | 3/80 (4)                                     | 2/2 (0.67)        | 53/53 (1.0)                   | 3/5 (0.56)     | nr |

nr = not reported; sens = sensitivity; spec = specificity; endometrial disease = polyp/hyperplasia/cancer; *for total number of studies
Chapter 7

Diagnosis of endometrial disease
As in most studies diagnostic accuracy regarding benign pathology was not described separately, we decided to extract data on the accuracy regarding the diagnosis of endometrial disease, i.e. endometrial cancer, hyperplasia and endometrial polyps together (table 3). The sensitivity of endometrial sampling was 29% in one study using blind D&C and the weighted sensitivity was 39% (range 21-69%) in five studies using hysteroscopy with histology as a reference standard. Specificity was again high, 98-100% regardless of the reference standard used. Figure 3C shows an ROC plot of the performance of the twelve studies that allowed the calculation of both sensitivity and specificity.

Figure 3A1 tm 3C2. Receiver operating curve (ROC) plots demonstrating the accuracy of endometrial sampling in diagnosing endometrial cancer, endometrial (pre) cancer or endometrial disease with D&C or hysteroscopy as a reference standard

A1. ROC plot for endometrial cancer with D&C as reference

A2. ROC plot for endometrial cancer with hysteroscopy as a reference

B1. ROC plot for endometrial (pre) cancer with D&C as reference

B2. ROC plot for endometrial (pre) cancer with hysteroscopy as a reference
The accuracy of endometrial sampling

Discussion

In this meta-analysis, we assessed the diagnostic accuracy of endometrial sampling regarding the diagnoses of endometrial cancer, endometrial (pre) cancer and endometrial disease (including endometrial polyps) in women with PMB, compared to two different reference strategies: D&C and hysteroscopy. Specificity of endometrial sampling is very high, irrespective of the type of disease or the reference test that was used. Sensitivities, on the other hand, are lower than anticipated based on existing meta-analyses, for all types of disease, but especially for atypical hyperplasia and endometrial disease, which includes endometrial polyps.

An important strength of this meta-analysis is that we performed a thorough search for articles on the diagnostic accuracy in women with PMB. By searching with all synonyms for PMB and endometrial sampling, we think we selected all articles on this subject. We also selected articles, which described only a subgroup of postmenopausal women and tried to contact the authors of these articles. Unfortunately, none of them responded. We included all eligible articles, regardless of the language used.

This article also has several limitations. Publication bias and the risk of missing potentially relevant articles are concerns with any systematic review. We attempted to mitigate this issue by using a robust search strategy, by checking cross-references and by consulting with a clinical librarian. Also, observer agreement regarding study selection was high. However, by performing this rigorous systematic search, we...
could only identify four more studies compared to the three existing meta-analyses on this subject. 7-9

Another weakness is that, because only a small number of studies is available and most studies are based on small samples, we had to draw conclusions based on a limited number of patients. Apart from the limited power, the relatively small number of studies and variability in methods also did not allow for more standard statistical analyses recommended for diagnostic test accuracy reviews, such as pooling sensitivity and specificity using the bivariate model or estimating summary ROCs.

The three existing meta-analyses focused on the diagnostic accuracy of endometrial sampling in a mixed population of pre- and postmenopausal women. 7-9 As the diagnostic accuracy of a test is strongly dependent on the prevalence (or pre-test probability) of a diagnosis, and the prevalence of endometrial cancer and atypical hyperplasia is much lower in pre- versus postmenopausal women, we think it is important to study this subject in a selected population of women with PMB. Therefore, we searched specifically for articles on the diagnostic accuracy of endometrial sampling in women with PMB and included only these studies, which reported data on postmenopausal bleeding separately.

Endometrial sampling fails in 42% of cases (either technical failure or insufficient material) and in 7% of these cases a (pre)cancer is found. This finding is in accordance with findings in other studies, which describe a failure or inconclusive rate of 16% to 50% and in 5 to 20% of these cases significant endometrial pathology is found. 25-27 Therefore, a case of a failed or inconclusive sample, should lead to further diagnostic work-up. Also, a benign result of endometrial sampling is not completely reassuring, as sensitivities are lower than anticipated based on previous literature. In the three existing meta-analyses (blind) D&C has been used as a reference standard, which is worrisome as D&C is known to miss 50-85% of focal intracavitary pathology. 28,29 As D&C could miss focal pathology, it could also possibly miss endometrial (pre) cancer in an endometrial polyp. Therefore, nowadays, D&C is almost completely replaced by hysteroscopy as a reference standard, both in clinical as well as in research settings. 10 It suggests that endometrial sampling, which is performed as a mini-curettage, as well misses a significant number of focal pathologies and therefore possibly also focal (pre) cancers. Because in women with atypical hyperplasia (which is regarded as endometrial (pre) cancer in 17-52% an underlying cancer is found at hysterectomy, 30 it is important to diagnose not only endometrial cancer but also atypical hyperplasia. Given the above findings,
The accuracy of endometrial sampling

further diagnostic work-up for focal intracavitary pathology in women with a failed, insufficient or benign result of endometrial sampling seems warranted.

The results of this systematic review suggest that the sensitivity of endometrial sampling is lower than was thought before for all types of disease, but especially for the diagnosis of atypical hyperplasia and endometrial disease in general. The question is if we can reassure patients without an endometrial polyp and a benign result of endometrial sampling. Is sensitivity of endometrial sampling especially low in women with an endometrial polyp? Unfortunately, we cannot answer these questions based on available literature. Therefore, more research on this subject is needed, using larger samples, given the prevalence of endometrial cancer and atypical hyperplasia (5-10%). Future research should therefore aim to gather information about large (prospective) cohorts of patients with PMB, to study the (cost-)effectiveness and diagnostic accuracy of the endometrial biopsy and hysteroscopy in the diagnostic pathway in women with PMB and a thickened endometrium on TVS.
Appendix

# Searches

1  uterine hemorrhage/ or metrorrhagia/
2  ((uterine or vaginal or abnormal) adj3 (hemorrhage or metrorrhagia or bleeding)).ti,ab.
3  1 or 2
4  Postmenopause/
5  (postmenopaus* or post-menopaus*).ti,ab.
6  4 or 5
7  3 and 6
8  endometrial sampling*.ti,ab,kw.
9  pipelle*.ti,ab.
10  ((biop* or sample* or aspiration) adj3 of endometri*).ti,ab,kw.
11  endometri* biop*.ti,ab,kw.
12  Biopsy/ and exp Endometrium/
13  8 or 9 or 10 or 11 or 12
14  7 and 13
15  exp “Sensitivity and Specificity”/
16  (Sensitiv* or Specific*).ti,ab.
17  (predict or ROC-curve or receiver-operator*).ti,ab.
18  (likelihood or LR*).ti,ab.
19  exp Diagnostic Errors/
20  (inter-observer or intra-observer or interobserver or intraobserver or validity or kappa or reliability).ti,ab.
21  reproducibility.ti,ab.
22  (test adj2 (re-test or retest)).ti,ab.
23  “Reproducibility of Results”/
24  accuracy.ti,ab.
25  Diagnosis, Differential/
26  validation studies.pt.
27  (failure* or success* or inadequate* or inconclusive*).ti,ab.
28  15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27
29  14 and 28
The accuracy of endometrial sampling

References


