The Value of Endometrial Volume as Estimated by Three-Dimensional Ultrasound for Detecting Endometrial Cancer in Postmenopausal Women: A Systematic Review and Meta-Analysis

Abstract

Objective: To analyse the diagnostic performance of endometrial volume calculated by three-dimensional (3D) ultrasound for diagnosing endometrial carcinoma in women with postmenopausal bleeding.

Methods: An extensive search of papers analysing the role of endometrial volume calculated by 3D ultrasound for diagnosing endometrial carcinoma in women with postmenopausal bleeding was performed in MEDLINE/PubMed and Web of Science from January 1996 to January 2020. Quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool.

Results: The extended search identified 318 citations but after exclusions, eight articles were included in the meta-analysis. The risk of bias for most studies was high for the four domains assessed in QUADAS-2. Overall, after excluding three studies that contributed significantly to heterogeneity, pooled estimated sensitivity and specificity for diagnosing endometrial cancer were 87% (95% confidence interval: 80–92%) and 60% (95% confidence interval: 51–68%), respectively. Heterogeneity was low or moderate.

Authors: Cristian Morales,1 Raul Diago,2 Salvador Cortés,3 Carla Peixoto,4 Maria Ángela Pascual,5 Stefano Guerriero,6 *Juan Luis Alcázar7

1. Department of Obstetrics and Gynecology, Hospital San Juan de Dios, Santiago, Chile
2. Department of Obstetrics and Gynecology, Hospital Universitario Severo Ochoa, Madrid, Spain
3. Department of Obstetrics and Gynecology, Hospital Universitario Virgen de la Arrixaca, Murcia, Spain
4. Department of Obstetrics and Gynecology, Centro Hospitalar São João, Porto, Portugal
5. Department of Obstetrics, Gynecology, and Reproduction, Hospital Universitario Dexeus, Barcelona, Spain
6. Department of Obstetrics and Gynecology, Policlinico Universitario Duilio Casula, Cagliari, Italy
7. Department of Obstetrics and Gynecology, Clinica Universidad de Navarra, Pamplona, Spain

*Correspondence to jalcazar@unav.es

Disclosure: The authors have declared no conflicts of interest.

Received: 11.03.20
Accepted: 14.04.20
Keywords: Cancer, endometrial, endometrial volume, three-dimensional (3D) ultrasound.
Citation: EMJ Radiol. 2020;1[1]:79-88.
INTRODUCTION

Endometrial carcinoma is the most frequent gynaecological malignancy in western countries, with most of the patients being postmenopausal. The main symptom of this disease is postmenopausal bleeding. The first approach to take in a woman who is symptomatic is to evaluate the endometrial thickness using two-dimensional ultrasound because a endometrial thickness <5 mm has a very high negative-predictive value (99.3%) when ruling out endometrial cancer, meaning that unnecessary biopsies can be avoided. In contrast, a thickened endometrium (>5mm) is a relatively nonspecific finding that can be found in many benign endometrial pathologies, such as cyst atrophy, polyp, or non-atypical hyperplasia. In fact, the specificity reported is approximately 50.0%. In the last two decades, three-dimensional (3D) ultrasound has become available for the diagnosis of some gynaecological diseases. Currently, 3D ultrasound is considered the first-line imaging diagnostic technique for some gynaecological lesions, such as congenital uterine anomalies. Furthermore, extensive research using this technique has been reported in the fields of reproductive medicine and gynaecological oncology.

The estimation of endometrial volume using 3D ultrasound is accurate and reproducible among examiners. Specifically, the role of the endometrial volume for diagnosing endometrial carcinoma in women with postmenopausal bleeding has been evaluated in a small number of small-scale prospective studies since the first report on its use in 1996. However, the role of this technique as a diagnostic test in this clinical setting has not been clearly established.

The aim of this systematic review and meta-analysis is to evaluate the diagnostic performance of the endometrial volume calculated by 3D ultrasound for diagnosing endometrial carcinoma in symptomatic postmenopausal women.

METHODS

Protocol and Registration

This meta-analysis has been performed according to the PRISMA statement and the Synthesizing Evidence from Diagnostic Accuracy Tests (SEDATE) guidelines. The protocol was not registered, a decision made by the researchers to avoid delays in starting the meta-analysis. All inclusion and exclusion criteria for studies to be selected were defined, as well as how data extraction and quality assessment had to be performed before starting the data search. Because of the study’s nature and design, Institutional Review Board (IRB) approval was waived.

Data Sources and Searches

Two of the authors (SC and CM) screened two electronic databases, MEDLINE/PubMed and Web of Science, to identify potentially eligible studies published from January 1996 to January 2020. The search terms included and captured the concepts of “endometrial cancer”, “endometrial malignancy”, “three-dimensional ultrasound”, “postmenopausal bleeding”, and “endometrial volume”. The language limit was set to English.

Study Selection and Data Collection

Two authors (CM and RD) screened the titles and abstracts identified by the search to exclude irrelevant articles. Then, full-text articles were selected to identify potentially eligible studies by applying set criteria:

- prospective and retrospective cohort studies that included patients with postmenopausal bleeding who underwent transvaginal ultrasonography examinations and included the calculation of endometrial volume using the virtual organ computer-aided analysis (VOCAL™) method;
- histological findings evaluated with endometrial samplings or hysterectomy;
- presence of data reported that would allow construction of a 2×2 table with a specific
cut-off of endometrial volume to estimate the diagnostic accuracy.

To avoid inclusion of duplicate cohorts from at least two studies reported from the same authors, the study period of each study was examined; if dates overlapped, the latest study published was selected. Additional articles were searched by reading the reference lists of those articles selected for full-text reading. The patient, intervention, comparator, outcome, and study design criteria used for inclusion and exclusion of studies were recorded.

The authors had intention to assess data based on individual patient information; therefore, they contacted the authors from the selected studies asking for specific data about some clinical characteristics of the patients, 3D ultrasound endometrial volume estimation results, and histologic data. This way, using the predefined endometrial volume, thresholds reported from the authors in the respective paper could be avoided. However, no responses were received from any of the authors. Therefore, the quantitative analysis using the respective threshold reported in each paper was performed.

Diagnostic accuracy results from the selected studies were retrieved independently by two authors (CM and RD). Disagreements arising during the process of study selection and data extraction were resolved by consensus among all four authors.

Risk of Bias in Individual Studies

A quality assessment of studies included in the meta-analysis was conducted by using the tool provided by the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2). The QUADAS-2 format includes four domains: 1) patient selection; 2) index test; 3) reference standard; and 4) flow and timing. For each domain, the risk of bias and concerns about applicability (the latter not applying to the domain of flow and timing) were analysed and rated as low, high, or unclear risk. The quality assessment was used to provide an evaluation of the overall quality of the studies and to investigate potential sources of heterogeneity.

Three authors (CM, RD, and JLA) evaluated the methodological quality independently. Disagreements were solved by discussion between these authors. The assessment of the quality was based on several issues, depending on the domain assessed. For the patient selection domain, the authors considered whether the study described the study’s design (in retrospective studies in which the reference test was already known by the researchers when performing the index test could not be elucidated, the worst case scenario was opted for and these were considered studies with a high risk of bias) as well as patients’ inclusion and exclusion criteria. For the index test domain, whether the study reported on the method of 3D volume acquisition and how the volume was calculated was considered, as well as how this was performed and interpreted. For the reference test domain, whether the study reported on the reference standard used (histology or not) and how sample was obtained was considered. Finally, for the flow-and-timing domain, the authors considered whether the study reported on the time elapsed from the index test assessment to the reference test (more than 4 weeks from index test to reference test was considered as high risk for bias).

Statistical Analysis

Information on the diagnostic performance of endometrial volume was extracted. A bivariate model was used to estimate the pooled sensitivity, specificity, positive likelihood ratio (LR), and negative LR. The LR were used to characterise the clinical utility of a test and to estimate the post-test probability of disease. Using 8% prevalence of endometrial cancer in women with postmenopausal bleeding (pretest probability), post-test probabilities were calculated by the positive and negative LR and plotted on a Fagan nomogram.

Heterogeneity for sensitivity and specificity was assessed by the Cochran’s Q statistic and the heterogeneous $I^2$ index. A p value <0.1 indicated heterogeneous, and $I^2$ values of <25%, 25–50%, and >75% were considered to indicate low, moderate, and high heterogeneity, respectively.

Forest plots of sensitivity and specificity of all studies were plotted. Considering that it could be a threshold effect, given that different studies used different thresholds for endometrial volume, a bivariate random effects modelling of
sensitivity and specificity was used to identify how much the threshold effect could explain heterogeneity, if found.

Hierarchical summary receiver operating characteristic curves were plotted to illustrate the relationship between sensitivity and specificity. Additionally, a binomial exact distribution for assessing within-study variability for sensitivity and specificity was used. Publication bias was assessed by the method of Deeks et al.\textsuperscript{15}

All analyses were performed with MIDAS and METANDI commands in Stata version 12.0 software for Windows (StataCorp, College Station, Texas, USA). A p value <0.05 was considered statistically significant.

RESULTS

Search Results

The electronic search provided 318 citations. After exclusion of 120 duplicate records, 198 citations remained. Of these, 166 were excluded because it was clear from the title or abstract that they were not relevant to the review (studies not related to the topic [n=146], reviews [n=9], articles published in non-English languages [n=8], and letters to Editor [n=3]).

The full text of the remaining articles was read. A further 24 studies were excluded: two studies did not use the VOCAL method; 14 studies included only women with a thickened endometrium, >4 mm;\textsuperscript{17-20,22,23} three studies excluded patients with previous gynaecologic disease such as fibroids or polyps;\textsuperscript{16,21,23} and one study pooled the hyperplasia with atypia and endometrial cancer in the same group.\textsuperscript{16} Concerning the index test domain, all the studies used the VOCAL rotational method to calculate the endometrial volume. In seven studies, the method of the index test as well as how it was performed was adequately described. One study did not describe the angle rotation step used.\textsuperscript{22} However, five studies\textsuperscript{16,17,19,20,23} were considered at high risk because they used a 30° rotation step for endometrial volume acquisition, and it has been shown that this approach is less reliable than using 9° or 15°.\textsuperscript{24,25} Only two studies used less than 30° rotation step, and they were considered as having low risk for bias regarding the index test.\textsuperscript{18,21}

For the reference standard domain, all studies were considered low risk because all patients were studied with endometrial sampling and posterior histologic diagnosis. Regarding the flow and timing domain, in four studies the time elapsed between the index test and reference standard was less than 1 week,\textsuperscript{17,18,20,23} but in four studies it was unclear.\textsuperscript{16,19,21,22}

Concerning applicability, for the patient selection domain, index test, and referent test, all studies were considered low risk for applicability because they used an adequate technique (transvaginal ultrasound) in the adequate clinical setting (postmenopausal bleeding) with an adequate reference standard (endometrial biopsy).
Diagnostic Performance of Endometrial Volume for Detection of Endometrial Cancer

The pooled sensitivity, specificity, positive LR, and negative LR of endometrial volume for detecting endometrial cancer were 87% (95% confidence interval [CI]: 77–93%), 69% (95% CI: 54–82%), 2.8 (95% CI: 1.9–4.2), and 0.19 (95% CI: 0.12–0.30), respectively. The diagnostic odds ratio was 15.0 (95% CI: 9.0–24.0). Significant heterogeneity was found for sensitivity ($I^2 = 74.48%$; Cochran Q = 27.43; p < 0.001) and specificity ($I^2 = 93.45%$; Cochran Q = 106.89; p < 0.001). Bivariate modelling showed that a threshold effect explained this heterogeneity with three studies involved.\(^{16,17,19}\)

After excluding these three studies, pooled sensitivity, specificity, positive LR, and negative LR of endometrial volume for detecting endometrial cancer were 87% (95% CI: 80–92%), 60% (95% CI: 51–68%), 2.2 (95% CI: 1.7–2.7), and 0.22 (95% CI: 0.13–0.36), respectively. The diagnostic odds ratio was 9.9 (95% CI: 5.1–19.3), but no heterogeneity was found for sensitivity and moderate heterogeneity was found for specificity (Figure 2). A hierarchical summary receiver operating characteristic curve for the diagnostic performance of endometrial volume for detecting endometrial malignancy is shown in Figure 3.

The Fagan nomogram shows that an increased endometrial volume increased the pretest probability of endometrial malignancy, from 8% to 16%, whereas a normal finding decreased the pretest probability, from 8% to 2%. No publication bias was observed (p = 0.43).

**DISCUSSION**

Most women with postmenopausal bleeding have a benign aetiology, and fewer than 8-10% will be diagnosed with endometrial cancer.\(^{2,26}\) Two-dimensional ultrasound is the first step in the evaluation of women with postmenopausal bleeding with the measure of the endometrial thickness because it has been shown to be the most cost-effective strategy in this clinical setting.\(^{27,28}\)
Several meta-analyses assessing the diagnostic performance of endometrial thickness for detecting endometrial cancer in women with postmenopausal bleeding, and even in asymptomatic postmenopausal women, have been reported. In women with postmenopausal bleeding, the most recent meta-analysis has demonstrated that an endometrial thickness <5 mm is effective to rule out endometrial cancer, with a high sensitivity (96.2%) and negative predictive value (99.3%), but rather low specificity (51.5%).

The advent of 3D ultrasound allowed an accurate estimation of organ or structure volume. There are different approaches for the estimation of organ volume, such as the use of the prolate ellipsoid measuring the three orthogonal diameters of the structure, using a distance and the perimeter of an ellipse, a spherical method, or the so-called VOCAL method. The latter processing system of the 3D volume allows calculation of the volume using a rotational method, with different rotation angles (9°, 15°, 30°).

The assessment of endometrial volume as measured by 3D ultrasound for detecting endometrial cancer in women with postmenopausal bleeding was first reported in 1996. In this study, Gruboeck et al. reported a series of 97 women with postmenopausal bleeding (11 had cancer).
They showed that an endometrial volume greater than 13 mL had a sensitivity and specificity of 100.0% and 98.8%, respectively, for diagnosing endometrial cancer. However, no further study was reported in the subsequent 10 years. Since 2007, several studies have been published addressing this issue, all of them using the VOCAL method.

In the present meta-analysis, the authors have evaluated the diagnostic performance of endometrial volume as estimated by 3D ultrasound using the VOCAL method to predict the presence of endometrial malignancy in women with postmenopausal bleeding. In the meta-analysis, it was observed that pooled sensitivity and specificity of the endometrial volume were 87% (95% CI: 80–92%) and 60% (95% CI: 51–68%), respectively, after excluding some papers that were identified as potential source of heterogeneity for a threshold effect.

The main strength of this study is that, to the best of the authors’ knowledge, this is the first

Figure 3: Hierarchical summary receiver operating characteristic curve for endometrial volume.
Area under the curve: 0.87 (95% confidence interval: 0.84–0.90).
HSROC: hierarchical summary receiver operating characteristic.
meta-analysis reported addressing this topic. Long et al.\textsuperscript{2} have reported a recent meta-analysis assessing the diagnostic performance of endometrial thickness for detecting endometrial cancer in women with postmenopausal women.\textsuperscript{2} In this meta-analysis, four studies were reported on, comprising data from 434 women, analysing 3D endometrial volume in this clinical setting. Out of these four studies, three have been included in this present meta-analysis\textsuperscript{18,19,23} and one was not because, from this paper, 2x2 tables could not be extracted.\textsuperscript{40} However, they did not perform an analysis about endometrial volume because of the small sample size.

However, the authors do consider there are some limitations that preclude drawing definitive conclusions regarding the role of endometrial volume as estimated by 3D ultrasound to detect endometrial cancer in women with postmenopausal bleeding.

First, the collected sample can be considered as relatively small as compared with that reported on meta-analyses focussed on endometrial thickness. The data presented here are based on 981 women derived from just eight studies, while meta analyses about endometrial thickness report data from 2,896 to 17,339 patients.\textsuperscript{29-32}

On the other hand, as stated above, studies that used the VOCAL method for analysing the 3D volumes obtained during the exam, and estimating the endometrial volume, were selected because this method has been reported as the most accurate to estimate the volume of the endometrium.\textsuperscript{41} Raine-Fenning et al.\textsuperscript{24} described that employing a rotation step of less than 30° was associated with a significantly smaller variance in measurements and a significantly greater mean endometrial volume. In this meta-analysis, most of the studies (in fact, all of them except those from Cho et al.\textsuperscript{21} and Alcázar et al.\textsuperscript{18}) used a 30° rotation step. This fact could be considered as a source of bias from the technical point of view, since the rotation angle used was not the most optimal for calculating the endometrial volume.

Furthermore, it is important to consider that in six studies the inclusion criteria were only patients with a thickened endometrium, >4 mm, and this also may lead to a selection bias, leaving out from the analysis some cases of endometrial cancer present in symptomatic women with a thin endometrium. It should be borne in mind that 25–34% of the Type II endometrial cancer could be present in patients with thin or indistinct endometrium.\textsuperscript{42-43} The authors have no information about endometrial volume in these cases.

In addition, there was high heterogeneity among the studies relating to different cut-off values used for endometrial volume (1.35–5.3 mL). The authors of the papers included were contacted in an attempt to perform a meta-analysis based on individual patient data, but none answered. Therefore, it is difficult to be precise about the specific cut-off value of endometrial volume to rule out endometrial cancer.

In most of the studies, endometrial hyperplasia with or without atypia cases were pooled in the benign group. There were no precise data for differentiating between the hyperplasia with and without atypia using the endometrial volume, so the authors had to be careful in the interpretation of that point, considering that almost 25% of patients with hyperplasia with atypia had a coexistent endometrial cancer in the final histology.\textsuperscript{44,45}

Nevertheless, the authors could not compare the diagnostic performance of endometrial volume and endometrial thickness, of which is the current standard.\textsuperscript{46-48} Thus, it cannot be elucidated whether endometrial volume is better than endometrial thickness.

\textbf{CONCLUSION}

In conclusion, endometrial volume as estimated by 3D ultrasound using VOCAL software has a moderate diagnostic performance for detecting endometrial malignancy in women with postmenopausal bleeding. A rough comparison with the results from a recent meta-analysis focussed on endometrial thickness suggests that endometrial volume appears inferior to endometrial thickness.\textsuperscript{2} However, a formal meta-analytical comparison has not been performed so far. There is clear room for future research in this topic because better-designed prospective studies are needed.
References


