Transvaginal three-dimensional ultrasound for preoperative assessment of myometrial invasion in patients with endometrial cancer: a systematic review and meta-analysis

Tatiana Costas¹, Rocío Belda², Juan Luis Alcázar³

¹Department of Obstetrics and Gynecology, University Hospital Salamanca, Salamanca, ²Department of Obstetrics and Gynecology, General University Hospital, Valencia, ³Department of Obstetrics and Gynecology, University Hospital of Navarre, School of Medicine, University of Navarra, Pamplona, Spain

Abstract

Aim: The aim of this meta-analysis is to evaluate the diagnostic accuracy of three-dimensional transvaginal ultrasound subjective assessment (3D-TVS) in the preoperative detection of deep myometrial invasion (MI) in patients with endometrial cancer, using definitive frozen section diagnosis after surgery as the reference standard. Material and methods: A search for studies evaluating the role of 3D-TVS for assessing myometrial invasion in endometrial cancer from January 1990 to November 2020 was performed in PubMed/MEDLINE and Web of Science. The Quality Assessment of Diagnostic Accuracy Studies 2 evaluated the quality of the studies (QUADAS-2). All analyses were performed using MIDAS and METANDI commands. Results: Nine studies comprising 581 women were included. The mean prevalence of deep MI was 39.8%. QUADAS assessment showed that most studies had a high risk for the patient selection domain. Overall, the pooled estimated sensitivity, specificity, positive likelihood ratio and negative likelihood ratio of 3D-TVS for detecting deep MI were 84% (95% CI, 73-90%), 82% (95% CI, 75-88%), 5 (95% CI, 3.1-7.1) and 0.20 (95% CI, 0.11-0.35), respectively. Conclusions: 3D-TVS has an acceptable diagnostic performance for detecting MI in women with endometrial cancer.

Keywords: endometrial cancer; meta-analysis; myometrial invasion; systematic review; three-dimensional transvaginal ultrasound

Introduction

Cancer of the corpus uteri, mainly endometrial cancer (EC), is the sixth most frequent form of cancer in women worldwide, being the most common cause of gynaecologic malignancy in developed countries, with an incidence of 12.9/100,000 women and a mortality rate of 2.4/100,000 women. The most important prognosis features for EC are FIGO stage (International Federation of Gynaecology and Obstetrics), myometrial infiltration (MI), histological type and differentiation grade. Among these, MI ≥50% is associated with both pelvic lymph node involvement and extension into the parametrium [1]. Even in low-grade endometroid tumours, lymph node metastasis risk increases from 4% to 15% if the tumour invades more than 50% of the myometrium [2]. These women are at a higher risk of recurrence and might benefit from an extended surgical staging with systematic pelvic and para-aortic lymphadenectomy in order to tailor adjuvant therapy. It is important to avoid overtreatment, as most of these women are elderly with comorbidities. Lymphadenectomy shows no survival benefit in low-risk endometrial cancer patients and, in fact, results in increased complications and higher morbidity [3-7]. For this reason, many authors advocate intraoperative gross or frozen section evaluation of the myometrial invasion...
in clinical stage I endometrial carcinoma to aid the decision of whether lymphadenectomy should be performed or not. Frozen section has shown to be highly accurate (94%) for determining myometrial invasion [8]. However, it may be time consuming and it is not performed in all hospitals. Therefore, a method that could reliably assess myometrial invasion preoperatively would be helpful to provide individual tailoring of the surgical approach [9].

There is no consensus about the optimal imaging technique for evaluation of myometrial and cervical invasion. Two-dimensional transvaginal ultrasound (2D-TVS) and magnetic resonance imaging (MRI) have shown accuracies of 69-74% and 66-90% respectively [10-13]. Three-dimensional transvaginal ultrasonography (3D-TVS) has several advantages when evaluating myometrial infiltration [14]. However, although several studies have shown the utility of this technique in myometrial evaluation of EC, the routine implementation in clinical practice has not been established so far. Theoretically, from an efficient point of view, the assessment of pre-surgical staging with 3D-TVS reveals a notorious benefit in terms of economic approach as compared to MRI. Since the advent of 3D-TVS, several studies have been published aiming at evaluating the role of this technique for detecting the depth of myometrial infiltration in endometrial cancer.

On the other hand, more hospitals, especially in developed countries, tend to implement 3D-TVS in their daily practices, revealing an increasing need of training and investigation of the actual performance of this technique. To the best of our knowledge, not a single meta-analysis analysed the diagnostic performance of 3D-TVS for detecting myometrial invasion in women with EC.

Thus, the purpose of this systematic review and meta-analysis is to evaluate the diagnostic accuracy of 3D-TVS in the preoperative detection of deep myometrial invasion in patients with endometrial cancer, using definitive frozen section diagnosis after surgery as the reference standard.

Material and methods

Protocol and registration

This systematic review and meta-analysis has been made according to the PRISMA Statement [15]. All methods for inclusion and exclusion criteria, data extraction and quality assessment were specified in advance. The protocol was not registered.

Data sources and searches

Two electronic databases, PubMed/Medline and Web of Science were screened in November 2020 by three of the authors (TC, RB and JLA) to identify eligible studies. “Endometrial cancer”, “ultrasound” and “three-dimensional” were used as search terms. No other methodological filters in the database were included to avoid possible omission of relevant studies, according to the recommendation of Leeflang et al [16].

Study selection and data collection

Two authors (TC and RB) screened the titles identified by the searches to exclude irrelevant articles, not strictly related to the topic. Abstracts of these articles were revised and some of these articles were excluded due to no relevant content (not relevant topic) or non-English languages. Full text publications were independently analysed by two authors (TC and RB) who read and collected data of all the publications.

Inclusion criteria were: 1) Use of 3D-TVS as index test to assess myometrial invasion preoperatively with subjective impression estimation of myometrial invasion (less than 50% of myometrium or more) by the investigator/s; 2) Reference standard was surgical pathological data; 3) Presence of results sufficient to construct the 2x2 table of diagnostic performance as minimum data requirement. To avoid inclusion of duplicate cohorts from at least two pair of studies reported by the same authors, the study period of each study was examined. If dates overlapped, the latest study published or the study that could provide us a 2x2 table was chosen. Disagreements arising during the process of study selection and data collection were resolved by consensus among three authors (JLA, TC and RB).

The PICOS (Patients, Intervention, Comparator, Outcome, Study design) criteria used for inclusion and exclusion of studies are shown in Table I. Diagnostic accuracy and additional useful information on patients and procedures were retrieved from a selected primary study written independently by the same authors (JLA, TC and RB).

Risk of bias in individual studies

Quality assessment was conducted using the tool provided by QUADAS (Quality Assessment of Diagnostic Accuracy Studies –2). The QUADAS-2 format includes four domains: 1) patient selection, 2) index test, 3) reference standard, 4) flow and timing. For each domain, the risk of bias and concerns regarding applicability was analysed and rated as low, high and 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 (JLA, TC and RB) evaluated the methodological quality independently. Disagreements were solved by discussion among these authors. The assessment of quality was based on whether the study described the study’s design as well as inclusion and exclusion criteria for the patient selection domain. The evaluation of
quality for the index test domain was based on whether the study reported on how the index test (3D-TVUS) was performed and interpreted. The evaluation of quality for the reference standard domain examined the reference standard used in each study and if sonographers and pathologists were blind or not to the index test. For the flow-and-timing domain’s evaluation, a description of the time elapsed from the index test assessment to the reference standard result was evaluated.

Statistical analysis

We extracted or derived information on diagnostic performance of 3D-TVS. Although other techniques were used in the studies (such as RMI or 2D-TVS), only the data from 3D-TVS evaluations were recollected. Primary outcome was pooled sensitivity, specificity, positive predictive value, negative predictive value and for instance, true positive, true negative, false positive and false negative values were obtained. Post-test probabilities were calculated and plotted on Fagan nomograms.

Heterogeneity of all studies was graphically explored, drawing forest plots of sensitivity and specificity. We then formally assessed the presence of heterogeneity for sensitivity and specificity using Cochran’s Q test and the I² Index. A test for heterogeneity examines the null hypothesis that all studies are evaluating the same effect (Higgins et al [17]).

Cochran’s Q statistic is computed by summing the squared deviations of each study’s estimate from the overall meta-analytic estimate, weighting each study’s contribution in the same manner as in the meta-analysis. A p-value <0.1 indicates heterogeneity. The I² index describes the percentage of total variation across studies that is due to heterogeneity rather than chance. According to Higgins et al values of 25%, 50%, and 75% would be considered to indicate low, moderate and high heterogeneity, respectively [17].

Summary receiver-operating characteristics (sROC) curves for each approach were plotted to illustrate the relationship between sensitivity and specificity.

All analyses were performed using Meta-analytical integration of Diagnostic Accuracy Studies (MIDAS) and METANDI commands in STATA version 12 for Windows (Stata Corporation, College Station, TX, USA). p-value <0.05 was considered statistically significant.

Results

Search results

The electronic search provided 108 citations. A flowchart summarizing literature identification and selection is given in Figure 1. Eighty-three of them were excluded due to no relevant topic (i.e. 3D MRI) or being review papers, 12 were excluded after reading abstract (not relevant to the review) and one due to other language than English (Chinese n=1). The full text of the remaining twelve articles were examined. One of them [18] was excluded due to the impossibility of obtain 2x2 tables. Alcázar et al [19] and Mascilini et al [20] analysed myometrial invasion with TDS and tumour/uterine volume 3D ratio without subjective impression as a method. For that reason, they were also discarded.

Nine studies were finally taken into account [9,14,18,21-26] (Table I). All studies analysed myometrial invasion with “subjective impression method”. No additional relevant studies were found from references cited in the papers included in the review. A flowchart summarizing literature identification and selection is given in Figure 1.

Characteristics of included studies

The nine studies [9,14,18,21-26] published between 2009 and 2019, reported data from 581 patients, 231 women having myometrial invasion >50% in the definitive pathological samples. This was considered as the prevalence and, therefore, pre-test probability.

The mean patient age was reported in 8 out of the 9 studies, in Trujillo et al no reference to this item being made [22]. Patients’ mean age was 61.4 years, ponderated by the number of patients in each study. The prevalence of ≥50% myometrial invasion was 39.8%. Hormonal status differentiating between pre- and postmenopausal women was given in 3 out of 9 studies [14,21,26]. The differentiation between histological grades was referred in all papers except for Green et al [23] and Jantarasaen-
garam et al [14]. This last author excludes all G3 cases as being considered high-risk profile. All studies described data regarding tumour histological type (endometrioid and non-endometrioid) except for Christensen et al [24].

As mentioned before, in all studies selected for quantitative analysis, method to estimate myometrial invasion was the subjective impression on 3D-TVS by examiner/s.

Most of the publications studied in this meta-analysis make a comparison between subjective impression with 3D-TVS and MRI [18,22,24-26]. Two papers also compared subjective impression with 3D-TVS and objective methods with 3D-TVS [9,21]. Another paper compares 2D-TVS and 3D-TVS [23]. Only one publication uses the subjective impression of 3D-TVS as the only method assessed [14]. Pathological evaluation of the removed uterus was considered as a reference standard in all papers.

**Methodological quality of included studies**

A graphical display of the risk evaluation of bias and concerns regarding applicability of the selected studies, according to predefined criteria, is shown in figure 2.

Regarding risk of bias and the domain patient selection, all of them explain specifically patterns of inclusion. Three studies do not include advanced stage patients [14,22,24]. Christensen et al [24] included endometrial hyperplasia with atypia in preoperative histology into the group of patients studied. This last case can increase the risk of overestimating the accuracy of 3D-TVS for the estimation of <50% of MI. In contrast, the exclusion of advanced stage cases was not considered a risk of bias, in spite that this technique is relevant especially for early stages since true tumour stage can only be determined after surgery. One study excluded patients with leiomyomas greater than 3 cm and submucous myomas [21]. This study was considered as a high risk for bias in patient selection domain since selecting exclusively ideal patients for the validation of the 3D-TVS, can suppose a bias in the validity of the test. Two studies [9,21] excluded patients whose videos are incomplete and one study excluded patients with suboptimal image resolution (classified by score 3 of 4 of quality) [14]. Incomplete videos is understandably criteria for exclusion, but if only optimal image resolution is selected, the scenario left is made up of exclusively ideal situations and this can lead to an overestimation of the accuracy of the test.

Regarding the study design, prospective design with US prior to surgery is carried out in 5 of the 9 studies [9,14,18,21,26]. The rest of papers describe retrospective methodology. Christensen et al [24] evaluated the volume of the uterus 6 months after surgery, the US examiners being blinded to the pathological result. In 2 of the retrospective studies the examiner was blinded to the pathology [23,24] and in 2 studies no explanation was provided [22,25].

Concerning the domain index test, all studies describe clearly the index test as well as how it was performed and interpreted. However, Christensen et al [24], describing a 4-step subjective methodology (1- initial subjective evaluation, 2- TUI function, 3- render function and
4- final evaluation with all gathered information), did not clearly define the importance given to each of these sub methods, leaving a riddle to understand the real criteria used in this study. All studies adopted the same pre-specified threshold to define deep MI (≥50% of myometrial thickness in any of the three spatial orthogonal planes). Four out of nine studies disposed ≥2 TVS examiners, and the remainder had one expert examiner for all patients [9,23,25,26]. From the group which included ≥2 examiners, Christensen et al [24] disposed of examiners with limited experience in 3D TVS and Alcazar et al [9] mixed both experts and beginners. The rest of them were experts. In two of the retrospective cases, as mentioned before, the blind condition of examiners to histological final results is not clearly defined, considering, in consequence, an unclear risk of bias 22.25.

Concerning the domain reference standard, all studies stated that histology was analysed after uterus removal, but most of them did not describe how this was done (frozen section method was only mentioned in 2 of the studies [20,25]). Three papers do not mention if the pathologist was blind to the 3D-TVS conclusions, leaving the risk of bias in these cases unclear [14,21,26].

Time and flow were not specified in four of the studies included [21-23,25].

Concerning regarding applicability, all studies were considered as low concern for patient selection domain. The incidence of EC is high; however, advanced stages of this neoplasia are not usually susceptible for 3D-TVS as they already implicate a surgical attitude; EC is usually diagnosed in early stages, motivating the patients’ selection.

Index test has not been found to implicate a high concern of applicability in any of the studies of this meta-analysis, except for Christensen et al [24] as a consequence of the above mentioned 4-step method.

Reference standard concern of applicability was not found to be high in any case, as histological definitive sample is usually the reference test in all gynaecology clinics.

**Diagnostic performance of 3D-TVS for detection of deep myometrial invasion**

We analysed the overall sensitivity and specificity of 3D-TVS in all studies to determine pulled sensitivity, specificity, LR+ and LR– of subjective impression in 3D-TVS in detecting deep MI. Results were 84% (95% CI, 73-90%), 82% (95% CI, 75-88%), 5.0 (95% CI, 3.1-7.1) and 0.2 (95% CI, 0.1-0.3) respectively. Diagnostic odds ratio (DOR) was 23.0 (95% CI, 9.0-59.0). sROC curve is shown in figure 3. Area under the curve was 0.89 (95% CI: 0.86-0.92)

Heterogeneity was moderate for sensitivity (I² 65.5% (95% CI 41.4-90.0%)) and for specificity (I² 65.1 % (95% CI 40.3-89.9%)). Data are shown in figure 4.
Fagan nomograms show that a positive test significantly increased the pretest probability of MI 50% while a negative significantly decreased the pre-test probability of MI ≥ 50% (fig 5). We did not observe publication bias (fig 6).

Discussion

EC is considered a neoplasia with good prognosis in terms of survival when the diagnosis is made at an early stage. The differences between the survival rates observed in stage 1 compared to those seen in stage 3 are overwhelming: according to the American Cancer Society the survival rate in early stages reaches 95% but it decreases to 17% in distant spreading [27]. Thus, the importance to find tests that lead to a better staging has become a challenge for the gynecological community.

As there is a really improvement of the rates in early cancer detection, new needs emerge from the elevated number of patients waiting for attending. In consequence, the efficiency of tests used for pre-surgical staging started to be an important object of investigation. In this context, MRI and 2D-TVS are currently considered the imaging techniques of choice for estimating MI before surgery [28]. Alcázar et al [9] in 2009 were the first authors that studied the performance of 3D-TVS. Since then, several studies using 3D-TVS for detecting MI have been reported.

We found that the overall diagnostic performance of 3D-TVS in detecting deep MI in women with EC have A pooled sensitivity of 84% and pooled specificity of 81%, subjective method being the reference measurement technique.

To the best of our knowledge, this is the first meta-analysis that evaluates subjective impression in 3D-TVS as the method of estimation of MI in women with EC reported to date.

The principal limitation in our meta-analysis is the low number of papers reported. The total sum of patients studied of 581. In this group of studies, moderate heterogeneity has been also found. Besides, the small sample the data reported should be interpreted carefully. In addition, there has not been a proper analysis of the objective methods used by some authors. However, we do consider that this could have increased the heterogeneity and we decided to exclude them. Due to the reduced number of papers published that describe objective methods, sub-analysis was non-viable.

Regarding generalizability, Alcazar et al published a meta-analysis in which 2D-TVS showed 78% (95% CI, 72-83) sensibility and 81% (95% CI, 71-87) specificity [29]. In this same study, objective methodology with
Karlson and Gordon’s methods for the estimation of myometrial infiltration showed a sensitivity of 84-85% and 82-80% specificity, respectively. In 2017, Alcázar compared 2D-TVS and MRI by meta-analysis. In this case, with a sample of 560 patients they reached a sensitivity of 75% and specificity of 86% for 2D-TVS [28].

Subjective 3D-TVS has, then, slightly superior results than those obtained by subjective methods in 2D-TVS and very similar results from objective methods of 2D-TVS.

MRI has been considered a good imaging test for evaluating MI. Recent studies demonstrated a pooled sensitivity ranging from 81% to 90% and pooled specificities ranging from 82% to 89%. No differences were found in these studies between contrast enhanced MRI and diffusion weighted MRI [28,30-32].

When comparing data from several meta-analyses about MRI with the data we have obtained in our meta-analysis regarding 3D-TVS, we observe that, roughly, diagnostic performance seems to be similar [28,30-32].

Objective methods in 3D-TVS have been studied. Alcazar describes TDS method as an objective method [9]. In this case, sensibility reaches 81.7% and specificity 41.5%. Ergenoglu validates this method demonstrating sensibility 89% and specificity 61% [21]. Another objective method described described by Mascilini et al [20] is tumor volume/uterine volume ratio. With this method, sensitivity reported in some studies was 90.5% and 60%, respectively, whereas, specificity was 28.5% and 85%, respectively [19,20].

However, we did not find a representative number of prospective studies dedicated exclusively to the comparison of 2D-TVS and 3D-TVS methods for MI assessment in EC. Also, lack of information has been identified regarding the objective methods used for 3D-TVS assessment. We think it is important to dedicate the investigation to the objective methods used in 3D-TVS evaluation for improving the quality of comparison with subjective 3D-TVS methods.

In conclusion, subjective impression in 3D-TVS seems to be an appropriate method for the estimation of MI of EC in early stages. However, as we have observed a high risk of patient selection bias, it is necessary to increase the number of prospective well-designed studies in this area.

Conflict of interest: none

References

15. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) website. Available at: http://www.prisma-statement.org/