Access this article online

Quick Response Code:

Website: www.pogsjournal.org

DOI: 10.4103/pjog.pjog_53_23

International Endometrial Tumor Analysis (IETA) terminology for the evaluation of endometrial color flow and vascular pattern in women with abnormal uterine bleeding – a reproducibility study among ob-gyn ultrasound subspecialists in a tertiary training hospital

Nina Rojana Lim Yu[1], Regina Rosario Panlilio-Vitriolo[2]

Abstract:

BACKGROUND: The International Endometrial Tumor Analysis (IETA) group developed a catalog of standardized terms to describe findings that may be associated with uterine pathology. However, there is a lack of reliability studies for these descriptors in the literature.

OBJECTIVE: The objective of this study was to estimate interobserver and intraobserver reliability with regard to the IETA group descriptors for endometrial vascular characteristics in women with abnormal uterine bleeding.

MATERIALS AND METHODS: Five nonexpert and five expert raters assessed stored still images of transvaginal ultrasound examinations obtained from 68 women with abnormal uterine bleeding and endometrial thickening. Endometrial vascularity was evaluated using the IETA group descriptors for color flow and vascular pattern. Interobserver agreement was estimated by comparing the assessments of the nonexpert and expert raters. Intraobserver agreement was estimated by repeating the raters' assessment after 4 weeks. Interrater agreement to the subjective assessment of an expert investigator was also computed.

RESULTS AND CONCLUSION: The reproducibility of assigning IETA color score is good regardless of the degree of expertise of the rater, although the experts displayed better interobserver reliability (κ = 0.74 vs. 0.57) and intraobserver reliability (κ = 0.84 vs. 0.63). However, the reproducibility of describing IETA vascular patterns is significantly worse for both expert and nonexpert raters in both interobserver reliability (experts κ = 0.49 vs. 0.34) and intraobserver reliability (experts κ = 0.65 vs. 0.42). Both expert and nonexpert raters exhibited acceptable agreement with the reference standard, with experts performing better for both color score (κ = 0.79 vs. 0.70) and vascular pattern (κ = 0.63 vs. 0.44).

Keywords:

Abnormal uterine bleeding, endometrium, International Endometrial Tumor Analysis, reproducibility, vascular characteristics

[1]Department of Obstetrics and Gynecology, Manila Doctors Hospital, University of the Philippines - Philippine General Hospital, Manila, Philippines, [2]Division of Ultrasound, Department of Obstetrics and Gynecology, University of the Philippines - Philippine General Hospital, Manila, Philippines

Address for correspondence:
Dr. Nina Rojana Lim Yu, 2427, Bayani St., Brgy 882, Sta. Ana, Manila, Philippines.
E-mail: ninarojana@gmail.com

Submitted: 20-Aug-2023
Revised: 15-Nov-2023
Accepted: 19-Nov-2023
Published: 27-Dec-2023

How to cite this article: Yu NR, Panlilio-Vitriolo RR. International Endometrial Tumor Analysis (IETA) terminology for the evaluation of endometrial color flow and vascular pattern in women with abnormal uterine bleeding – a reproducibility study among ob-gyn ultrasound subspecialists in a tertiary training hospital. Philipp J Obstet Gynecol 2023;47:260-70.

Introduction

A 2020 report by the Philippine Obstetrical and Gynecological Society indicates that abnormal uterine bleeding is the most common gynecological diagnosis in the Philippines, accounting for 26.6% of all gynecological admissions.[1]

While the most common causes of abnormal uterine bleeding vary according to age and reproductive status, evaluation for endometrial cancer must done in women aged 45 years old and above. The use of 45 years old as the threshold for increased concern regarding endometrial cancer is supported by the fairly low risk of endometrial hyperplasia and carcinoma before age 45 years, which then increases with advancing age.[2] This potentially lethal neoplasm also accounts for abnormal uterine bleeding in approximately 10% of postmenopausal women.[3] Transvaginal ultrasound is a noninvasive method to evaluate women with abnormal uterine bleeding for endometrial hyperplasia or cancer when the endometrium is homogeneous. In women with postmenopausal bleeding and not on hormonal replacement therapy, an endometrial thickness of ≤ 4 mm is associated with a low risk of endometrial disease.[4] However, measurement of endometrial thickness in premenopausal women is not helpful in the evaluation of abnormal uterine bleeding.[2] Other than endometrial thickness, there are several other sonographic features of the endometrium and uterine cavity that may aid in the assessment for risk of cancer in patients with abnormal uterine bleeding. The International Endometrial Tumor Analysis (IETA) group was formed at the World Congress on Ultrasound in Obstetrics and Gynecology (OB-GYN) in Chicago last 2008, with the aim of developing a standardized catalog of terms and definitions to describe ultrasound features in the uterine cavity and endometrium on grayscale sonography, color flow imaging, and sonohysterography. The relationship between these features and the presence or absence of endometrial neoplasia is not yet well established. However, they may form the basis for prospective studies to predict the risk of different endometrial pathologies based on their sonographic features.[5] Other sonographic features linked to endometrial cancer risk include endometrial thickness, internal endometrial echo structure, characteristic of the subendometrial halo, determination of interruption of texture and endomyometrial junction, and Doppler analysis of vascularity. Additional parameters defined in the IETA consensus statement include the presence of localized lesions to diagnose the presence of a polyp, myoma, or a possible neoplastic pathology.[6] To date, there are no locally published studies evaluating the reliability and agreement of the descriptors in IETA consensus statement. Therefore, in the local setting, there is a need to assess the interobserver and intraobserver variability of the IETA color score and vascular pattern classification among different examiners with different levels of expertise. This is an important step in the validation of the IETA terms and definitions as a reporting system that may eventually be used to aid in determining endometrial cancer risk.

A recent study based in Sweden[7] determined that interrater reliability and intrarater reliability when using IETA terminology were limited. This may have implications when assessing the association between a particular ultrasound feature and a specific histological diagnosis, because lack of reproducibility reduces the reliability of the association between a feature and the outcome.

The simplicity and uniformity provided by the IETA group consensus statement are appealing, and these descriptors may be used in the development of scoring systems that correlate sonographic findings to specific pathologies and their prognosis. However, sonography is an innately operator-dependent modality and may be prone to inconsistencies between different observers. Furthermore, as of the time of writing, there are no locally published studies evaluating the reproducibility of the sonographic descriptors outlined by the IETA group. This study aims to evaluate the reproducibility of the IETA group color Doppler descriptors, specifically their interobserver and intraobserver reliability among raters of varying degrees of expertise. For a reporting system to be acceptable, it must exhibit good interobserver reliability to be able to consistently distinguish patients with different findings from each other, despite the subjective differences in interpretation by various observers. It also must exhibit good intraobserver reliability to be able to display consistent results across multiple readings by the same observer. These specific criteria can be assessed by measuring the interobserver and intraobserver reliability between raters on their usage of the IETA group descriptors. If the reporting system for color flow score and color pattern exhibits good reliability, future research may explore its potential to be correlated with specific endometrial pathologies such as carcinoma.

Objectives

General objective

- To determine the reproducibility of assessment of color score and vascular pattern in perimenopausal and postmenopausal women with abnormal uterine bleeding, using standard descriptors as defined by the IETA group.

Specific objectives

- To determine intraobserver and interobserver reliability, regarding the following IETA parameters

on color Doppler imaging, among multiple observers with different levels of expertise:
  - Color scores based on endometrial vascularity
  - Vascular patterns based on IETA descriptors.
- To determine agreement with a reference standard, regarding the following IETA parameters on color Doppler imaging, among multiple observers with different levels of expertise:
  - Color scores based on endometrial vascularity
  - Vascular patterns based on IETA descriptors.

Materials and Methods

Research design

This is an analytical, prospective cohort study on the intraobserver and interobserver reproducibility of selected endometrial color flow parameters among multiple OB-GYN ultrasound subspecialists with varying degrees of expertise.

Selection of participants and sampling design

This study was conducted in the Division of Ultrasound of the Department of OB-GYN at the Philippine General Hospital. Purposive sampling for test images was used, and these were retrieved from the records of patients who met the following criteria:

- Perimenopausal and postmenopausal women aged 45 years old and above
- Presented clinically with abnormal uterine bleeding
- Thickened endometrium measuring 5 mm and above (for postmenopausal women)
- Technically adequate color or power Doppler images in stored digital format. Images must have been acquired with the following machine settings:
  - Frequency of at least 5Mhz
  - Pulse repetition frequency of 0.3–0.9 kHz
  - Wall filter of 30–50 Hz
  - Color/power gain of 50 or less.
- Images obtained between 2018 and 2020.

The study participants were invited from the medical staff of the Division of Ultrasound, which included five fellows-in-training with <2 years of relevant experience and five consultants with at least 10 years of relevant experience. The fellows served as the nonexpert raters, while the consultants served as the expert raters.

Sample size

For the Cohen's kappa statistic, the computed minimum number of test images required in a study with at least two raters, to detect a statistically significant κ, with a pretest κ of 0.8, power of 90%, and alpha at 0.05, was 68. For a fixed number of test images to be rated, increasing the number of raters above 3 did not detrimentally affect the statistical power of the test.

Description of the machine used

Images used for the study were obtained with the Samsung Medison Accuvix A30, Samsung UGEO H60, and GE Voluson P6, each using a 5–7 MHz transvaginal probe with similar Doppler settings at the time of image acquisition.

Outcomes of interest

The primary outcomes of interest were the interobserver and intraobserver reliability between raters, with regard to their grading of the color score and assignment of the vascular pattern of test images, following the standardized IETA descriptors. The secondary outcome of interest was the correlation of the ratings of the fellows and consultants with that of the supervising investigator, whose subjective expert evaluation of the images served as the reference standard.

Study procedure

Test images were selected by the principal and supervising investigators from digitally stored department records that met the inclusion criteria. The supervising investigator is an expert in gynecologic imaging with more than 20 years of relevant experience. Sixty-eight test images were selected and evaluated according to the IETA group criteria, equally distributing representatives for each vascular pattern and vascular score as much as possible. This initial evaluation was staged into two phases, minimizing the effect of intraobserver variability for the supervising investigator. The subjective evaluation of the supervising investigator served as the reference standard in this particular study, against which the evaluations of the participating fellow and consultant raters were compared.

The criteria for grading color flow were as follows:

- Grade 1 – no flow
- Grade 2 – minimal flow
- Grade 3 – moderate flow
- Grade 4 – abundant flow.

The criteria for describing vascular pattern were as follows:

- Single dominant vessel without branching
- Single dominant vessel with branching
- Multiple vessels with focal origin at the myometrial–endometrial junction
- Multiple vessels with multifocal origin at the myometrial–endometrial junction
- Scattered vessels without visible origin at the myometrial–endometrial junction
- Circular flow
- No detectable color signals.

The selected test images were then compiled and used to develop an online testing tool. This testing tool was

proofread and tested by the supervising investigator before it was distributed to the raters. Two groups of raters participated in this study – five fellows-in-training with <2 years of relevant experience to serve as the nonexpert raters and five consultants with at least 10 years of relevant experience to serve as the expert raters. The nonexpert and expert raters were instructed to thoroughly study the IETA consensus statement and the representative images for color score and vascular morphology features, before reviewing our test images.

This evaluation was carried out in two phases. For the initial phase, each rater independently assigned the color scores and described the vascular patterns of 68 cases, using the IETA descriptors. This was done by means of the testing tool which was delivered through a secure online platform (password-protected Google™ Forms). The testing tool was in a multiple-choice format, and the choices for each item included all the color score or vascular pattern descriptors defined by the IETA consensus statement [Supplementary File 1]. The results of the initial phase were used to assess interobserver reliability, as well as each individual rater's agreement to the reference standard. For the final phase, which was done after 4 weeks, each rater was sent the testing tool once more, containing the same set of test images, but with the images rearranged compared to the initial phase of testing. The results of this final phase were used to assess intraobserver reliability.

Each rater evaluated the test images independent to and blinded from each other. They were given only one opportunity to evaluate the test images during both the initial and final phases. The results were then returned to the principal investigator for statistical analysis and data interpretation.

Statistical analysis plan

Intraobserver reliability analysis was performed by comparing the scores for each rater in their initial and final phases of testing and quantified by calculating the Cohen's weighted κ index value with its 95% confidence interval (CI). Interobserver reliability analysis was done by separately comparing the initial phase scores between the raters from the nonexpert group and the expert group and quantified by calculating the Fleiss' multirater κ index value with its 95% CI. The weighted κ index values of each individual rater compared to the reference standard were also computed. As enumerated in a review of reliability studies by Hernaez,[8] which was similarly adapted by Alcázar *et al.*,[9] a κ value of 0–0.20 indicates poor reliability, 0.21–0.40 indicates fair reliability, 0.41–0.60 indicates moderate reliability, 0.61–0.80 indicates good reliability, and 0.81–1.00 indicates very good reliability. The statistical software package SPSS™ Statistics for Windows version 26.0 (released 2019 by IBM Corp., Armonk, New York, USA) was used for data analysis.

Ethical considerations

Approval from the review ethics board was secured before conduction of the study. The ultrasound images and reports that were used were all obtained from department records. Reports were stored in the password-protected computer database of the department for a duration of 5 years and can only be accessible in the department itself. In compliance with the Data Privacy Act of 2012, permission from the head of the division was requested before access and use of these data. Only the principal investigator and supervising investigators facilitated data collection in coordination, with the administrative staff in charge of data safe-keeping and storage. Processing of data collected also followed the implementing rules and regulations under the Data Privacy Act of 2012. Since this research only used imaging records, and no active patient participation is required, a waiver of informed consent was requested. This was further justified by anonymizing the included cases to maintain patient privacy. No other patient demographic/information, save for their age, were included in this study, and there was a negligible risk to patient privacy. The outcome of the study could prove useful in validating the IETA group descriptors in the assessment of endometrial pathology.

No reward or compensation of any form was given to the participating sonologists or the patients whose imaging records were used.

Since this study only involved patient imaging records, and no new sonographic examinations had to be performed, no patients were subjected to additional risks. Likewise, participating sonologist raters were not rewarded for participation nor penalized for nonparticipation in any manner.

The investigators in this study have no conflicts of interest to declare.

Results

Grayscale and color Doppler still images were obtained from 68 peri- and postmenopausal women who presented with abnormal uterine bleeding and exhibited thickened endometrium on grayscale ultrasound. The reference standard for color score and vascular pattern, as evaluated by the supervising investigator, is detailed in Tables 1 and 2, respectively. It should be noted that the samples representing each category for color score and vascular pattern were not equally distributed, and this variation has been accounted for in the statistical analysis.

Overall interobserver reliability for assigning the color score was better among the expert raters (0.74) compared to the nonexpert raters (0.57). The expert raters have consistently displayed good to very good reliability across all color score categories [Table 3]. Specific interobserver reliability for nonexpert and expert raters for color score was best for absent flow (0.69 and 0.89) and abundant flow (0.65 and 0.85). The relatively lower kappa values for assigning minimal and moderate flow indicated a greater degree of subjectivity for assigning scores that lie away from the extremes.

Table 1: Prevalence of color scores on test images obtained from 68 women with abnormal uterine bleeding and thickened endometrium, as per the reference standard

Color score	Number (*n*=68)	Prevalence (%)
1 (absent flow)	13	19.1
2 (minimal flow)	26	38.3
3 (moderate flow)	23	33.8
4 (abundant flow)	6	8.8

Table 2: Prevalence of vascular patterns on test images obtained from 68 women with abnormal uterine bleeding and thickened endometrium, as per the reference standard

Vascular pattern	Number (*n*=68)	Prevalence (%)
Single dominant vessel with branching	7	10.3
Single dominant vessel without branching	12	17.6
Multiple vessels with focal origin	5	7.4
Multiple vessels with multifocal origin	16	23.5
Scattered vessels without visible origin	15	22.1
Circular flow	0	0
No detectable color signals	13	19.1

Overall interobserver reliability for assigning the vascular pattern was significantly worse for both nonexpert (0.34) and expert raters (0.49), although the latter still displayed better reliability [Table 4]. Specific interobserver reliability was best for no detectable color signals (0.71 and 0.88). However, interobserver reliability was particularly poor among both nonexperts and experts for multiple vessels with focal origin (0.08 and 0.19) and single dominant vessel with branching (0.15 and 0.18). Furthermore, two nonexpert raters have erroneously assigned a circular flow pattern, despite there being no such cases present in the testing tool.

The mean intraobserver reliability between the initial and final evaluations for assigning the color score is good for the nonexpert raters (0.63) and very good for the expert raters (0.84) [Table 5]. Individual intraobserver reliability ranges from moderate to good for the nonexperts, and from good to very good for the experts.

The mean intraobserver reliability between the initial and final evaluations for assigning the vascular pattern is worse for both the nonexpert raters (0.42) and expert raters (0.65) [Table 6]. Individual intraobserver reliability

Table 3: Interobserver reliability among different raters for assigning the International Endometrial Tumor Analysis color score, expressed as the multirater kappa index

Color score (*n*=68)	Nonexpert (*n*=5)		Expert (*n*=5)	
	Prevalence	Multirater κ (95% CI)	Prevalence	Multirater κ (95% CI)
1	20.3	0.691 (0.688–0.693)	18.8	0.894 (0.892–0.897)
2	37.4	0.541 (0.539–0.544)	34.1	0.666 (0.664–0.669)
3	29.4	0.476 (0.473–0.478)	34.7	0.650 (0.647–0.652)
4	12.9	0.648 (0.645–0.650)	12.4	0.851 (0.848–0.853)
	Overall interobserver reliability	0.573 (0.571–0.574) - moderate reliability	Overall interobserver reliability	0.738 (0.736–0.739) - good reliability

CI: Confidence interval

Table 4: Interobserver reliability among different raters for assigning the International Endometrial Tumor Analysis vascular pattern, expressed as the multirater kappa index

Vascular pattern (*n*=68)	Nonexpert (*n*=5)		Expert (*n*=5)	
	Prevalence	Multirater κ (95% CI)	Prevalence	Multirater κ (95% CI)
Single dominant vessel with branching	8.5	0.152 (0.149–0.154)	7.9	0.175 (0.173–0.178)
Single dominant vessel without branching	17.1	0.418 (0.418–0.420)	14.7	0.590 (0.587–0.592)
Multiple vessels with focal origin	11.2	0.082 (0.079–0.084)	7.6	0.188 (0.185–0.190)
Multiple vessels with multifocal origin	22.1	0.299 (0.296–0.301)	27.6	0.471 (0.468–0.473)
Scattered vessels without visible origin	19.1	0.211 (0.208–0.213)	23.8	0.368 (0.366–0.370)
Circular flow	1.5	0.087 (0.084–0.089)	0.0	N/A
No detectable color signals	20.6	0.712 (0.710–0.715)	18.2	0.882 (0.879–0.884)
	Overall interobserver reliability	0.341 (0.340–0.342) - fair reliability	Overall interobserver reliability	0.491 (0.489–0.492) - moderate reliability

CI: Confidence interval, N/A: Not available

ranges from fair to moderate for the nonexperts and from moderate to good for the experts. This is similar to the decrease in overall interobserver reliability for assigning vascular pattern among all raters.

The mean interobserver agreement with the reference standard for assigning the color score is good for both the nonexpert raters (0.70) and expert raters (0.79) [Table 7]. Individual reference agreement ranges from good to very good for both the nonexpert and expert raters.

The mean interobserver agreement with the reference standard for assigning the vascular pattern is worse for both the nonexpert raters (0.44) and expert raters (0.63) [Table 8]. Individual reference agreement ranges from fair to moderate for the nonexperts and from moderate to good for the experts. This echoes the previously observed trend of decreased interobserver and intraobserver reliability in assigning the vascular pattern as compared to assigning the color score.

Discussion

Transvaginal color Doppler enables noninvasive assessment of uterine and endometrial vascularization. The role of this technique in differentiating benign from malignant endometrial pathologies has been

Table 5: Intraobserver reliability for all raters for assigning the International Endometrial Tumor Analysis color score, expressed as the weighted kappa index

Nonexpert	Weighted κ (95% CI)	Expert	Weighted κ (95% CI)
A	0.776 (0.762–0.790)	A	0.855 (0.843–0.867)
B	0.518 (0.499–0.537)	B	0.877 (0.866–0.888)
C	0.758 (0.743–0.773)	C	0.868 (0.856–0.880)
D	0.436 (0.416–0.456)	D	0.806 (0.792–0.820)
E	0.683 (0.666–0.700)	E	0.772 (0.757–0.787)
Mean intraobserver reliability	0.634 - good reliability	Mean intraobserver reliability	0.836 - very good reliability

CI: Confidence interval

Table 6: Intraobserver reliability for all raters for assigning the International Endometrial Tumor Analysis vascular pattern, expressed as the weighted kappa index

Nonexpert	Weighted κ (95% CI)	Expert	Weighted κ (95% CI)
Fellow A	0.548 (0.532–0.564)	Consultant A	0.795 (0.782–0.808)
Fellow B	0.453 (0.437–0.469)	Consultant B	0.609 (0.593–0.625)
Fellow C	0.447 (0.430–0.464)	Consultant C	0.651 (0.635–0.667)
Fellow D	0.353 (0.336–0.370)	Consultant D	0.612 (0.596–0.628)
Fellow E	0.305 (0.289–0.321)	Consultant E	0.598 (0.582–0.614)
Mean intraobserver reliability	0.421 - moderate reliability	Mean intraobserver reliability	0.653 - good reliability

CI: Confidence interval

Table 7: Agreement with the reference standard for all raters for assigning the International Endometrial Tumor Analysis color score, expressed as the weighted kappa index

Nonexpert	Weighted κ (95% CI)	Expert	Weighted κ (95% CI)
Fellow A	0.852 (0.839–0.865)	Consultant A	0.812 (0.798–0.826)
Fellow B	0.605 (0.587–0.623)	Consultant B	0.833 (0.820–0.846)
Fellow C	0.753 (0.738–0.768)	Consultant C	0.668 (0.651–0.685)
Fellow D	0.557 (0.538–0.576)	Consultant D	0.770 (0.755–0.785)
Fellow E	0.734 (0.719–0.749)	Consultant E	0.854 (0.842–0.866)
Mean agreement with reference	0.700 - good agreement	Mean agreement with reference	0.787 - good agreement

CI: Confidence interval

Table 8: Agreement with the reference standard for all raters for assigning the International Endometrial Tumor Analysis vascular pattern, expressed as the weighted kappa index

Nonexpert	Weighted κ (95% CI)	Expert	Weighted κ (95% CI)
Fellow A	0.524 (0.507–0.541)	Consultant A	0.690 (0.675–0.705)
Fellow B	0.439 (0.422–0.456)	Consultant B	0.691 (0.676–0.706)
Fellow C	0.507 (0.491–0.523)	Consultant C	0.685 (0.670–0.700)
Fellow D	0.370 (0.353–0.387)	Consultant D	0.498 (0.481–0.515)
Fellow E	0.379 (0.362–0.396)	Consultant E	0.600 (0.583–0.617)
Mean agreement with reference	0.444 - moderate agreement	Mean agreement with reference	0.633 - good agreement

CI: Confidence interval

assessed in several studies with controversial results. In a determined that Doppler analysis can contribute to the diagnosis of endometrial pathology in women with postmenopausal bleeding and thickened endometrium. Opolskiene *et al*.[10] demonstrated that the addition of Doppler analysis to a risk calculation model for endometrial malignancy using grayscale parameters improved its diagnostic performance significantly. Furthermore, a 2017 study of Alcázar *et al*.[11] has determined that evaluation of color score on Doppler analysis is a highly reproducible parameter regardless of the experience of the examiner, even compared to some other grayscale parameters.

The color score is a subjective semiquantitative assessment of the amount of blood flow present. A color score of 1 is given when no color flow signals are seen in the endometrium, a score of 2 is given when only minimal color can be detected, a score of 3 is given when moderate color is present, and a score of 4 is assigned when abundant color is detected [Figure 1].

The vascular pattern within is reported with respect to the presence or absence of dominant vessels or of other specific patterns. Dominant vessels are defined as one or more distinct vessels passing the endomyometrial junction. The dominant vessel may show branching within the endometrium, which may be described as either orderly or disorderly. Dominant vessels may present as a single vessel, otherwise referred to as the pedicle artery sign, with or without branching. Multiple dominant vessels may have a focal origin at the endomyometrial junction, or they may otherwise have a multifocal origin. Other vascular patterns within the endometrium include scattered vessels (dispersed color signals within the endometrium but without visible origin at the myometrial–endometrial junction) and circular flow [Figure 2].

In this study, we evaluated the reproducibility of the IETA descriptors for color score and vascular pattern among raters of varying degrees of expertise. We discovered that interobserver reliability for assigning the

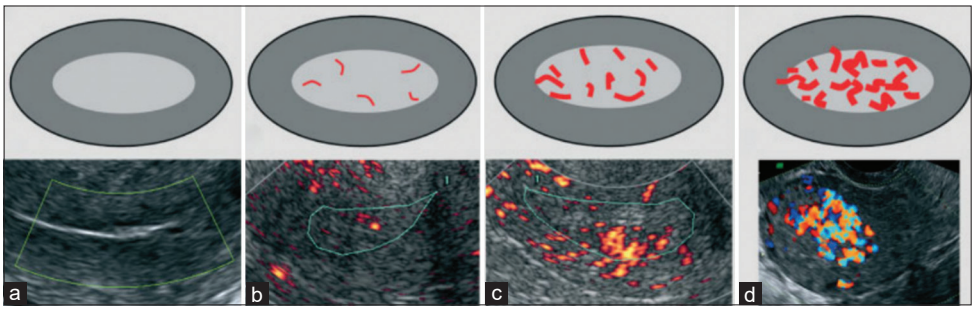


Figure 1: Color score of the endometrium on Doppler assessment based on the International Endometrial Tumor Analysis consensus statement [5] (a) absent flow (b) minimal flow (c) moderate flow (d) abundant flow

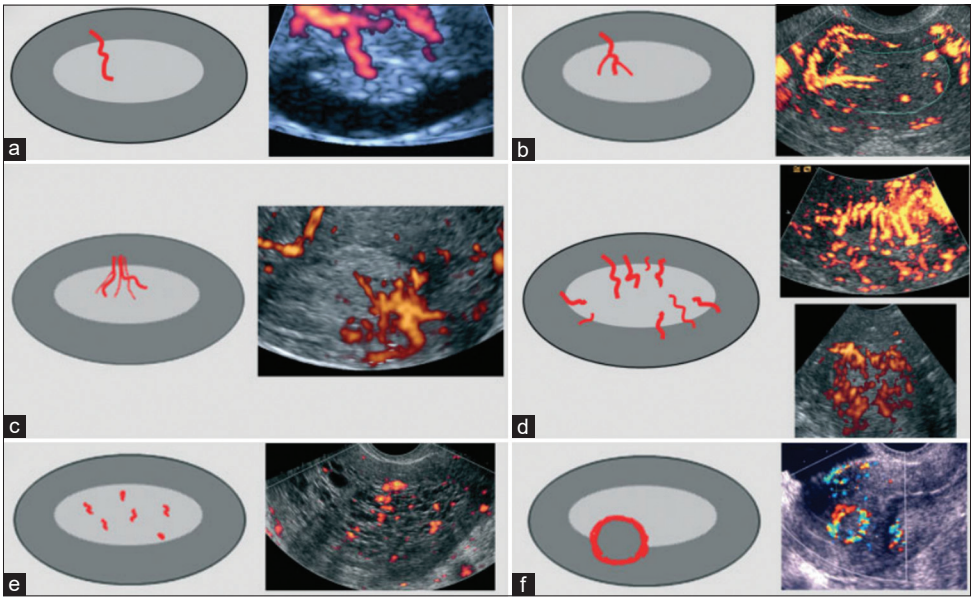


Figure 2: Vascular patterns on Doppler assessment based on the International Endometrial Tumor Analysis consensus statement[5] (a) Single dominant vessel without branching (b) Single dominant vessel with branching (c) Multiple vessels with focal origin (d) Multiple vessels with multifocal origin (e) Scattered vessels without visible origin (f) Circular flow

color score ranges from moderate to good depending on the degree of experience of the raters, while interobserver reliability for describing the vascular pattern is significantly poorer for both nonexpert and expert raters. Specific color score and vascular pattern parameters that exhibit better interobserver agreement tend to lie at the extremes, such as the absence of color flow or the presence of abundant color flow. We also confirmed that both interobserver reliability and intraobserver reliability were consistently better among the expert raters, compared to the nonexpert raters. Moreover, we determined that intraobserver reliability is consistently better than interobserver reliability, regardless of the experience of the rater. This strong intraobserver reliability suggests that the IETA descriptors provide a consistent standard for reporting endometrial vascular characteristics, although statistical validation of this consistency may be addressed in future studies.

Both the nonexpert and expert raters exhibited good agreement with the reference standard in assigning the color score. However, this overall agreement with the reference standard decreased when it came to describing the vascular pattern. The common pattern of decreasing reliability as far as the vascular pattern is concerned indicates a greater degree of subjectivity, and this has also been observed in previously published reliability studies.

To the best of our knowledge, there are only three other previously published studies that investigated the reliability of the color flow and vascular pattern descriptors that have been standardized by the IETA consensus statement. Alcázar *et al*. investigated the reproducibility of endometrial vascular patterns in a 2006 study[9] and that of endometrial color flow scores in a 2017 study.[10] In 2018, Sladkevicius *et al*. published an agreement and reliability study for all of the descriptors that are covered by the IETA terminology, including both grayscale and color Doppler parameters.[7] Although there are key differences in the study design and statistical analysis between these three studies and our current one, the primary outcomes of interest are directly comparable and are detailed in Table 9.

There are key differences in methodology between these studies that could have affected the specific values for interobserver and intraobserver reliability for each of the evaluated vascular parameters. The years of experience of the nonexpert raters in these previous studies ranged from <1 year to 3 years, while the years of experience of the expert raters ranged from 10 years (Alcázar *et al*., 2006) to 33 years (Sladkevicius *et al*., 2018). In both studies primarily authored by Alcázar, the nonexpert raters were fellows-in-training who have completed at least 1 year of specialist training, while the nonexperts in Sladkevicius study were newly certified specialists in gynecologic ultrasound. All of the expert raters involved in these studies are duly certified, practicing specialists in gynecologic imaging.

The vascular pattern reliability study of Alcázar *et al*. in 2006 had a much smaller number of raters, and the criteria for assigning vascular score have been collapsed into only three, as compared to the present study's seven. The color score reliability study of Alcázar *et al*. in 2017 was more similar in design to our own, with eight raters retrospectively evaluating stored video clips of postmenopausal patients with abnormal uterine bleeding. However, in contrast to our present study, the sample images have all been obtained by a single expert rater, and these have been equally distributed among each of the evaluated color score parameters. The authors of that study have surmised that these factors contributed to the unusually high degrees of interobserver and intraobserver reliability that they have observed in their investigation. Meanwhile, Sladkevicius *et al*. comprehensively evaluated all of the sonographic parameters that have been codified by the IETA group, and included all grayscale and color Doppler descriptors in their 2018 reliability study. Of the three cited studies, this most closely resembles the design and statistical power of our current investigation.

The most consistent trend observed across all four studies is the decrease in both interobserver reliability and intraobserver reliability when it comes to assigning the vascular pattern, as compared to assigning the color score. This may seem unusual at first, since assigning color score is innately more subjective than describing the vascular pattern. The IETA group did not specify any quantitative measurements for the color score and used qualitative descriptors such as "minimal," "moderate," and "abundant." Meanwhile, the IETA group provided specific defining sonographic features for each specific vascular pattern, because these may potentially correlate with certain endometrial pathologies. The "single dominant vessel" patterns are believed to correspond to the pedicle artery sign that is characteristic to benign polyps.[12] The "multiple vessel" patterns may be associated with neoangiogenic phenomena that occur in endometrial malignancies.[13] The "circular" pattern is typically associated with submucous fibroids.[14] Nevertheless, all four studies seem to indicate a greater degree of subjectivity in the evaluation of endometrial vascular patterns, in spite of the standardized features described by the IETA group.

Sladkevicius *et al*.[7] proposed an interesting explanation to this conundrum. They believed that the main reason for the poorer reliability in assigning the vascular pattern

Table 9: Comparison with previous reliability studies for endometrial color Doppler parameters in women with abnormal uterine bleeding

Parameter	Current study	Alcázar *et al*. (2006)[a]	Alcázar *et al*. (2017)[b]	Sladkevicius *et al*. (2018)[c]
Interobserver reliability for color score				
Nonexpert raters	0.57 - moderate reliability	No data	0.85 - very good reliability	0.69 - good reliability
Expert raters	0.74 - good reliability		0.84 - very good reliability	0.77 - good reliability
Intraobserver reliability for color score				
Nonexpert raters	0.63 - good reliability		0.90 - very good reliability	0.78 - good reliability
Expert raters	0.84 - very good reliability		0.86 - very good reliability	0.82 - very good reliability
Interobserver reliability for vascular pattern				
Nonexpert raters	0.34 - fair reliability	0.41 - moderate reliability	No data	0.32 - fair reliability
Expert raters	0.49 - moderate reliability	0.61 - good reliability		0.35 - fair reliability
Intraobserver reliability for vascular pattern				
Nonexpert raters	0.42 - moderate reliability	0.52 - moderate reliability		0.48 - moderate reliability
Expert raters	0.65 - good reliability	0.78 - good reliability		0.53 - moderate reliability

[a]Evaluated average paired weighted κ values rather than pooled multirater κ values. Since this study predates the IETA consensus statement (2010), they used a modified, consolidated version of the same set of vascular pattern descriptors that the IETA group eventually formalized, [b]Evaluated average paired weighted κ values rather than pooled multirater κ values, [c]Evaluated ICC values rather than multirater κ values. In this context, the ICC can be interpreted in the same manner as Fleiss' κ. ICC: Intraclass correlation coefficient, IETA: International Endometrial Tumor Analysis

is the greater number of descriptive categories for the raters to choose from. The IETA group has defined six distinct vascular patterns, in contrast to color flow which only has four distinct scores. This assertion is supported by the relatively better reliability for vascular patterns that we observed in the study by Alcázar *et al*.[9] In their study, they consolidated the six vascular pattern categories into just three – single-vessel pattern (with or without branching), multiple-vessel pattern (focal or multifocal origin), and scattered-vessel pattern. This reduction in available options appeared to have led to greater consistency for all raters, despite their differences in degree of expertise.

The strengths of our present study are the inclusion of as many as five raters for each degree of expertise and the statistically significant number of test images utilized. Moreover, this study appears to be the first attempt to evaluate the reliability of the IETA group descriptors for both color score and vascular pattern in the Philippine setting.

The limitations of our study include the use of digitally stored static images instead of real-time ultrasound images. This precludes the modification of machine settings to optimize the assessment of color Doppler signals. The test images were also obtained by a wide range of sonologists, and we were unable to correct for the inherent operator-dependent differences in sonographic technique because these data were collected retrospectively. The combination of these factors means that we cannot necessarily generalize our findings to live scanning.

This consensus statement on the terminology, definitions, and measurements for describing the sonographic features of the endometrium and uterine cavity was developed by the IETA group for the primary purpose of enabling consistency in reporting imaging findings, including diagnostic and research applications.[5] In turn, this would then facilitate the interpretation and comparison of results between imaging studies. A potentially useful application for a reproducible and consistent reporting system is the development of a risk model to predict endometrial malignancy in high-risk patients based on pertinent grayscale and color Doppler findings.

The IETA group color flow descriptors, in spite of their innate subjectivity, exhibit good interobserver and intraobserver reliability regardless of the raters' degree of expertise. However, the interobserver reliability and intraobserver reliability with regard to the use of IETA vascular pattern descriptors are underwhelming. The use of fewer categories for vascular pattern could potentially improve reliability, as a higher number of categories are associated with difficulty in achieving interrater agreement. Moreover, specialized training workshops on endometrial vascular imaging can be introduced to minimize the disparity in the reliability of the evaluations between nonexpert and expert raters. We believe that these steps could improve the reliability of the vascular

pattern descriptors, and increase its applicability in the potential development of a risk model for endometrial malignancy.

Since the sample images are retrospectively collected from stored still images rather than videos or stored three-dimensional volumes, the evaluation of several IETA group grayscale parameters is limited. The assessment of the endometrial–myometrial junction, subendometrial halo, and the presence of polyps or masses requires a complete sweep of the endometrium on imaging, utilizing a standardized and comprehensive sonographic technique. Thus, critical grayscale findings may have been missed on stored still images.

Endometrial thickness is an objective finding that is easily measured and is expected to have good reproducibility. Endometrial echogenicity, while subjective, relies on an internal imaging benchmark (comparison with myometrial echogenicity) – thus, it is expected to have adequate reproducibility as well. The same cannot be said for the color Doppler parameters that were defined by the IETA group. Given the existing limitations, the color Doppler parameters of color flow score and vascular pattern are those that warrant a reliability study the most. The aforementioned grayscale parameters may be better evaluated by a well-controlled prospective imaging study.

Moreover, this study focuses on the assessment of internal reliability, thus validation with external reference standards for diagnosis such as biopsy or ancillary imaging lies beyond this study's scope. The assessment of external reliability may serve as an opportunity for future research.

Conclusion and Recommendation

In conclusion, our findings suggest that the reproducibility of assigning IETA color score is good regardless of the degree of expertise of the rater. However, the reproducibility of describing IETA vascular patterns is significantly worse for both nonexpert and expert raters. All raters, regardless of their degree of expertise, also exhibited greater agreement with the reference standard when evaluating color score, as compared to the vascular pattern.

Future reliability studies would greatly benefit from a prospective design that utilizes real-time images, to allow for the standardization of sonographic technique and technical specifications. This would also facilitate correlation with surgical findings when available, providing external validation to the usefulness of these descriptors. Given the acceptable reproducibility of the IETA group descriptors for color score, we recommend further investigation into its ability to predict endometrial malignancy, possibly in combination with well-established grayscale parameters that are related to cancer, such as endometrial thickness and endometrial echogenicity.

Acknowledgment

I would like to express my sincere gratitude to my advisor, Dr. Vitriolo, for her guidance, encouragement, and support throughout this research project. She has been a constant source of inspiration and motivation for me and has helped me overcome many challenges and difficulties. I am also grateful to my other consultants for their valuable feedback and suggestions that improved the quality of my work. I would also like to thank my husband, colleagues, and friends, who have been supportive and collaborative during this journey.

Authorship contributions

Nina Rojana L. Yu - Involved in conceptualization, methodology, formal analysis, investigation, writing-original draft, writing-review and editing, visualization.

Regina Rosario Panlilio-Vitriolo - Involved in conceptualization, validation, investigation, resources, writing-review and editing, supervision, project administration.

Financial support and sponsorship

This study is entirely funded by the primary author.

Conflicts of interest

There are no conflicts of interest.

References

1. Annette MM, Maria Antonia EH, Helen RA, Antoinette UA, Jennifer CP, Ina SI, *et al*. POGS report on obstetrical and gynecological indicators of health care. Philipp J Obst Gynecol 2020;46:29-37.
2. Levens ED, Decherney AH. Diagnosis of abnormal uterine bleeding in reproductive-aged women. Pract Bull 2012;120:197-206.
3. Prendergast EN, Misch E, Chou YA, Roston A, Patel A. Insufficient endometrial biopsy results in women with abnormal uterine bleeding. Obstet Gynecol 2014:123 Suppl 1:180S-1S.
4. ACOG committee opinion no. 440: The role of transvaginal ultrasonography in the evaluation of postmenopausal bleeding. Obstet Gynecol 2009;114:409-11.
5. Leone FP, Timmerman D, Bourne T, Valentin L, Epstein E, Goldstein SR, *et al*. Terms, definitions and measurements to describe the sonographic features of the endometrium and intrauterine lesions: A consensus opinion from the international endometrial tumor analysis (IETA) group. Ultrasound Obstet Gynecol 2010;35:103-12.
6. Sharon B, Mendoza JP, Sigue AJ, Gorgonio NM. Creation of a scoring system to determine endometrial cancer risk using the international endometrial tumor analysis (IETA) features. Philipp J Obstet Gynecol 2017;41:9-17.
7. Sladkevicius P, Installé A, Van Den Bosch T, Timmerman D,

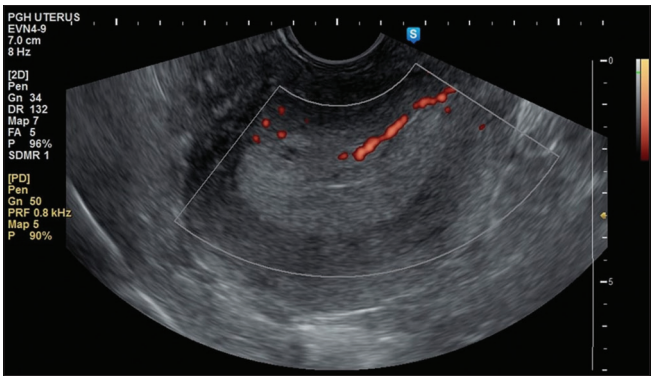
Benacerraf B, Jokubkiene L, *et al.* International endometrial tumor analysis (IETA) terminology in women with postmenopausal bleeding and sonographic endometrial thickness ≥ 4.5 mm: Agreement and reliability study. Ultrasound Obstet Gynecol 2018;51:259-68.

8. Hernaez R. Reliability and agreement studies: A guide for clinical investigators. Gut 2015;64:1018-27.
9. Alcázar JL, Ajossa S, Floris S, Bargellini R, Gerada M, Guerriero S. Reproducibility of endometrial vascular patterns in endometrial disease as assessed by transvaginal power Doppler sonography in women with postmenopausal bleeding. J Ultrasound Med 2006;25:159-63.
10. Opolskiene G, Sladkevicius P, Valentin L. Ultrasound assessment of endometrial morphology and vascularity to predict endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness>or=4.5 mm. Ultrasound Obstet Gynecol 2007;30:332-40.
11. Alcázar JL, Pascual MÁ, Ajossa S, de Lorenzo C, Piras A, Hereter L, *et al.* Reproducibility of the international endometrial analysis group color score for assigning the amount of flow within the endometrium using stored 3-dimensional volumes. J Ultrasound Med 2017;36:1347-54.
12. Timmerman D, Verguts J, Konstantinovic ML, Moerman P, Van Schoubroeck D, Deprest J, *et al.* The pedicle artery sign based on sonography with color Doppler imaging can replace second-stage tests in women with abnormal vaginal bleeding. Ultrasound Obstet Gynecol 2003;22:166-71.
13. Abulafia O, Sherer DM. Angiogenesis of the endometrium. Obstet Gynecol 1999;94:148-53.
14. Cil AP, Tulunay G, Kose MF, Haberal A. Power Doppler properties of endometrial polyps and submucosal fibroids: A preliminary observational study in women with known intracavitary lesions. Ultrasound Obstet Gynecol 2010;35:233-7.

Supplementary File 1

Sample item from online testing tool.

(Authors' note: The entire set of test images used for the online testing tool, as well as the full data collection forms, are available in digital format. The corresponding author can be contacted for any requests on getting the full version.)



What is the color flow score exhibited in the above image?

A. Color score 1: Absent flow
B. Color score 2: Minimal flow
C. Color Score: 3: Moderate flow
D. Color Score 4: Abundant flow.

What is the vascular pattern exhibited in the above image?

A. Single dominant vessel with branching
B. Single dominant vessel w/o branching
C. Multiple vessels with visible focal origin
D. Multiple vessels with multifocal origin at the myometrial–endometrial junction
E. Scattered vessels without visible origin at the myometrial–endometrial junction
F. Circular flow
G. No detectable color signals.