Congenital Uterine Malformation by Experts (CUME): T-shaped uterus

A. Ludwin[1,2*], M. A. Coelho Neto[3, 4], I. Ludwin[1,2], C.O. Nastri[3], W. Costa [3,4], M. Acién [5], J.L. Alcazar[6], B. Benacerraf[7], G. Condous[8], A. DeCherney[9], R-L. De Wilde[10], M.P. Diamond[11], M.H. Emanuel[12,13], S. Guerriero[14], W. Hurd[15], D. Levine[16], S. Lindheim[17], A. Pellicer[18], F. Petraglia[19], E. Saridogan[20], W.P. Martins[3,4*]

1) Department of Gynecology and Oncology, Jagiellonian University, Krakow, Poland; 2) Ludwin & Ludwin Gynecology, Private Medical Center, Krakow, Poland; 3) SEMEAR Fertilidade, Reproductive Medicine, Ribeirao Preto, Brazil. 4) Department of Obstetrics and Gynaecology, Faculty of Medicine of Ribeirão Preto, University of Sao Paulo (DGO-FRMP-USP), Ribeirao Preto, Brazil. 5) San Juan University Hospital/Miguel Hernández University, Alicante, Spain. 6) Department of Obstetrics and Gynecology, University of Navarra, Pamplona, Spain. 7) Harvard Medical School, Brookline, MA, United States. 8) Acute Gynaecology, Early Pregnancy and Advanced Endosurgery Unit, Nepean Hospital, Sydney Medical School Nepean, University of Sydney, Sydney, NSW, Australia. 9) Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland, United States. 10) Carl-von-Ossietzky-University Oldenburg, Oldenburg, Germany. 11) Department of Obstetrics & Gynecology, Augusta University, Augusta, GA, United States. 12) Department of Gynaecology and Reproductive Medicine, University Medical Center Utrecht, Utrecht, The Netherlands. 13) Department of Gynaecology, University Hospital Ghent, Ghent, Belgium. 14) Department of Obstetrics and Gynecology, University of Cagliari, Cagliari, Italy. 15) Division of Reproductive Endocrinology and Infertility Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, United States. 16) Department of Radiology Beth Israel Deaconess Medical Center, Boston, MA, United States. 17) Department of Obstetrics & Gynecology, Wright State University, Boonshoft School of Medicine, Dayton, OH, United States. 18) Instituto Valenciano de Infertilidad, Valencia, Spain. 19) University of Florence, Florence, Italy, 20) University College London Hospital, London, United Kingdom.

***Corresponding authors:**

Professor Artur Ludwin, Department of Gynecology and Oncology, Jagiellonian University Kopernika 23, 31-501 Krakow, Poland, E-mail: ludwin@cm-uj.krakow.pl

Professor Wellington P Martins, Department of Obstetrics and Gynecology, Ribeirao Preto Medical School, University of Sao Paulo (DGO-FMRP -USP) Av. Bandeirantes, 3900 – 8 andar - HCRP - Campus Universitario; City: Ribeirao Preto; State: Sao Paulo; Country: Brazil; Postal code: 14048–900. E-mail: wpmartins@gmail.com

Running Title: T-shaped uterus

Keywords: Mullerian ducts, Uterine Anomalies, T-shaped uterus, Dysmorphic uterus Agreement

CONTRIBUTION

What does this work add to what is already known?

The prevalence, clinical relevance, and appropriateness of the management of T-shaped uterus are not properly assessed due to the lack of objective criteria for its diagnosis. We identified three uterine morphometric measurements that have good diagnostic test accuracy and moderate reliability.

What are the clinical implications of this work?

One can now use three measurements with cut-offs (lateral indentation depth ≥ 7 mm, lateral indentation angle ≤ 130° and T-angle ≤ 40°), which have good accuracy and moderate reliability to confirm the diagnosis when there is any suspicious of T-shape uterus by subjective assessment of the uterine cavity on the coronal plane.

ABSTRACT

Objectives: To assess whether is there any uterine measurement that is reliable and accurate to distinguish between T-shaped and normal/arcuate uterus considering the most voted option by 15 experts as the reference standard.

Methods: This was a prospectively designed multi-rater reliability/agreement study with elements of diagnostic accuracy study performed between Nov-2017 and Dec-2018 in a sample of 100 3D datasets of different uteri acquired in consecutive women with the presence of lateral uterine cavity indentations between 2014-2016. Fifteen blinded representative experts (5 clinicians, 5 surgeons, and 5 imaging specialists) provided their independent opinion whether that the uterus was T-shaped or not regarding anonymized images of the coronal plane of each uterus. Two other blinded experienced observers performed 15 measurements using the originally acquired 3D data-sets. The agreement between experts was assessed by kappa and percent agreement. The inter-observer reliability of measurements was assessed using the concordance correlation coefficient (CCC). The diagnostic test accuracy was assessed using the area under ROC curve (AUROC) and the best cut-off value was assessed using Youden's index, using the most voted option by the 15 experts as the reference standard. Sensitivity, specificity, negative and positive-likelihood ratio (LR- and LR+) and post-test probability were calculated.

Results: There were 20 T-shaped and 80 normal/arcuate uteri using CUME as reference standard (at least 8 votes). Single experts recognized from 5 to 35 (median = 19) T-shaped uteri by subjective judgments. The agreement among experts was 82% with kappa = 0.43. We identified three measurements with good diagnostic test accuracy considering CUME as reference standard: lateral indentation angle (AUROC=0.95), lateral indentation depth (AUROC=0.92), and T-angle (AUROC=0.87). From these three measurements, T-angle was the one with the best inter-observer reproducibility: CCC = 0.87 vs 0.82 vs. 0.62 (T-angle, lateral indentation depth and angle respectively). The best cut-offs values for these measurements were: lateral indentation angle ≤ 130° (sensitivity = 75%, specificity = 96%, LR- = 0.21, LR+ = 21, positive post-test probability = 83%), lateral indentation depth ≥ 7 mm (sensitivity = 85%, specificity = 78%, LR- = 0.06, LR+ = 4.2, positive post-test probability = 49%), and T-angle ≤ 40° (sensitivity = 65%, specificity = 91%, LR- = 0.23, LR+ = 6.4, positive post-test probability = 64%). We suggest considering as borderline T-shaped when only 2 of these 3 criteria (PPV = 50%) are present and definitely T-shaped uterus when meeting all the three criteria (PPV = 93%), with 75% and 93% post-test probability for the definitions, respectively.

Conclusions:

The diagnostic of T-shaped uterus is not easy and the agreement among top-experts is only moderate, and single expert judgment is commonly insufficient for accurate diagnosis. The study has identified three measurements with cut-offs that had good diagnostic test accuracy and fair to moderate reliability (lateral indentation depth ≥ 7 mm, lateral indentation angle ≤ 130° and T-angle ≤ 40°), and when applicated together they have provided high post-test probability of this condition. Based on the CUME criteria of T-shaped uterus, the prevalence, clinical implication and the management, as well as assessment of post-surgical morphologic outcomes of this condition may be determined with enough accuracy, reliability and with a known probability of disease after negative and positive test results. The CUME definition of T-shape uterus may help on the development of interventional randomized controlled trials, observational studies, and diagnostics of uterine morphology in every day practice, and therefore could be adopted by guidelines on uterine anomalies to enrich their classification systems.

INTRODUCTION

In the mid-20th century uterine cavity close to a (i) triangle; (ii) with small lateral indentations; or (iii) with even greater fundal and lateral indentations were considered as the three basic normal uterine variants (Figure 1) [1, 2]. T-shaped uterus was first described as a Diethylstilbestrol (DES) related congenital uterine anomaly with a uterine cavity shape resembling the letter “T” based on the hysterosalpingography findings [3, 4 5-7]. Considering DES-related etiology and morphology, T-shaped uterus has been included as a separate class of anomalies with no strict definition/morphometric criteria by Buttram and Gibbons [6], and the American Fertility Society (AFS) classification [8]. The last women from those 1 to 1.5 million who had been exposed to DES in utero (withdrawn from use during pregnancy in 1971) are no longer at the reproductive age [9, 10], while T-shaped uterus is still being identified in young women [7, 11]. In the era without DES, the T-shaped uterus may be still seen rarely, being of primary origin [12] or secondary to intrauterine adhesions syndrome [7], tuberculosis [13], or adenomyosis [14], but the quality of evidence is not good. It is mentioned that T-shaped uterus may be related to reproductive failure, as infertility [7, 12, 15-18] and recurrent miscarriage [7, 12, 16-18].

In 2013, the ESHRE-ESGE published a guideline in which T-shaped uterus was also subjectively defined: a narrow uterine cavity due to thickened lateral walls [19]. It did not mention objective criteria to diagnose T-shaped uterus nor how the measurements of the uteri should be performed. In available modifications of AFS system supplemented by morphometric criteria: Salim’s modification did not include T-shaped uterus as an anomaly [20], Ludwin’s modification did not provide measurable criteria for T-shaped uterus [11], and the recent official definitions published by ASRM are only for normal/arcuate, septate and bicornuate uterus [21]. Until now, there are no objective and measurable criteria on how to identify T-shaped uterus, and it is an important barrier to build evidence on prevalence, clinical implications and management of this condition.

Recently, the Congenital Uterine Malformations by Experts (CUME) initiative and original methodology has been developed to elaborate measurable criteria and definitions for normal and abnormal uterine morphologies, build common language and unify the classification systems based on reliability/agreement and accuracy testing [22].

The present study is aimed to evaluate whether there are any sufficiently reliable and accurate measurable criteria for classification of T-shaped uterus, reflecting the diagnosis made most often by representative experts as a reference standard.

METHODS

Study design

This was a prospectively planned multi-rater reliability/agreement and diagnostic accuracy study to propose objective measurable diagnostic criteria to differentiate T-shaped uterus from normal/arcuate uterus. The Guidelines for Reporting Reliability and Agreement Studies (GRRAS) and STrengthening the Reporting of Observational studies in Epidemiology (STROBE) statement were used in the design and reporting of the study. The standardized design and principles of methodology (the CUME methodology) of the study were similar as recently published for developing a reasonable reference definition of normal/arcuate and septate uterus [22]. The target reference standard used in the CUME methodology is multi-expert reference, and three-dimensional ultrasound is used as a reference diagnostic tool. The target population is women of reproductive age, with target uterine appearance. The study was a part of a large prospective project on screening, diagnosis and classification of congenital utero-vaginal malformations by ultrasound. The whole project and this study were approved by the local ethics committee (KBET/236/B/2013). Written informed consent was obtained from all patients.

We planned to include one hundred 3D data-sets showing uteri with the presence of any lateral internal indentation without fundal indentation > 10 mm, which would be performed by assessing pre-acquired data-sets from consecutive women. Two independent observers prepared the images of a coronal view of all uteri, and 15 invited experts independently provided their subjective assessment of the images. Two raters, who were at level three of experience according to the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB, www.efsumb.org), with at least 15 years of experience in 3D ultrasound independently performed specific target measurements of each uterus using the same unprocessed 3D datasets. The field, years of experience, country of residence of the included experts and observers involved in performing of measurements are presented in Table S1.

Main and secondary outcome measures

Our primary objective and outcome were to assess the level of agreement among experts in the recognition of T-shaped uterus using subjective judgments from a review of coronal view images from the three-dimensional ultrasound. The secondary aims/outcomes were: (i) to evaluate the inter-observer reliability/agreement of potentially useful measurements developed to establish objective diagnostic criteria for T-shaped uterus and to distinguish T-shaped uterus from normal/arcuate; (ii) to assess the diagnostic test accuracy and the best cut-off values for these measurements in comparison with experts' opinion as reference standard; (iii) to define T-shaped uterus using the most accurate and reliable objective measurable criteria with reasonable post-test probability of condition for use the definition of T-shaped uterus in trials, observations, and practice.

Settings and Patients

The study included 3D data-sets of uteri of women in reproductive age (>18 and <45 years), who were recruited for detailed evaluation in a single recruitment center, a private clinic (Ludwin and Ludwin Gynecology, Krakow, Poland) specialized in the management of congenital uterine anomalies. Women with previously unknown pregnancy recognized during the evaluation, menopause, malignancy, benign lesions, distorted uterine cavity shape and wall, and those with previous uterine surgeries were excluded from the main project. Obvious uterine morphologies that were not considered as T-shaped uteri (namely normal/arcuate uterus without lateral indentation, uterine agenesis, unicornuate, bicornuate, didelphys and septate uteri) were ineligible for this study, because only uteri with presence of lateral indentation without external and internal vertical deformity were representative for our target population (women with lateral distortion of uterine cavity shape) as subjects for differentiation between T-shaped uterus and normal/arcuate uterus. Asymmetrical uteri with a lateral indentation on one side were not excluded.

All primary patients´ data and 3D ultrasound datasets, as well as this study data, were collected prospectively and aggregated in a large database of the main project (Jagiellonian University, Department of Gynecology and Oncology, Krakow, Poland). However, the data collected in the primary study, including demographic data, as well as measurements, previous real-time and off-line diagnoses were not used in this study, and did not influence the study results. Only recorded unprocessed 3D ultrasound volumes were used. Data collection for this study was performed using online software (Google forms) and database (Dropbox).

Ultrasound scanning

The technique of ultrasound scanning was the same as previously described [11, 22]. Two- and three-dimensional transvaginal ultrasound (or transrectal at intact hymen) was performed for evaluation of uterus. Ultrasound systems (Voluson E8 Expert BT13, GE Healthcare Ultrasound, Milwaukee, WI, USA) with volumetric intravaginal probes (GE RIC 5–9 MHz) were used for the 3D data-sets acquisition in a standardized manner (sagittal view of the uterus; maximum sweep angle of 120° and maximum quality; the approximate angle between the ultrasound beam and the uterine axis = 90°; holding breath and refrain from moving by patients during 3D volumes acquisition). Two different 3D datasets of each uterus were obtained, recorded and stored for all women, after confirmation that the entire uterus is captured in the 3D volume. If needed, the acquisition was repeated in the transverse plane, which is more appropriate than acquisition in sagittal view for uteri with higher transverse diameter, yet the whole uterus with cervix may be only partially visible in the coronal view. Women with regular cycles were examined between days 17 and 25 of the menstrual cycle since the secretory phase facilitates the visualization of the endometrial cavity. Women with irregular cycles, amenorrhea and on hormonal contraception were examined independently of the menstrual cycle, yet outside the menstruation. All the ultrasound examinations were performed in real time by one of two experienced sonologists (I.L. and A.L, level three of experience by EFSUMB, at least 15 years of experience with 3D-US, respectively).

Subject selection and preparation

The manner of subject selection was based on consecutive inclusion of 3D dataset of uteri from the target population of women with specific features of the uterus (presence of uterine cavity lateral indentations). 3D datasets of consecutive women included in the main project were specifically re-assessed for this study by an experienced sonologist (I.L.) until the inclusion of 100 uteri fitting eligibility criteria. The presence of lateral indentation was recognized subjectively, yet fundal indentations were measured, when present to confirm that they are lower than 1 cm.

A single obtained 3D volume of the uterus of each woman was anonymized, numbered, stored and sent to two observers from another institution (MACN and WCB, both with 3 years of experience with 3D-ultrasound) who prepared the images of coronal view for the invited experts, using two different ultrasound techniques:

- OmniView combined with VCI: Omniview is a rendered view of the coronal plane by curved rendered mode, which allows the identification of the internal os in the coronal plane of the uterus; and VCI (volume contrast imaging) which is a tool that enhances the contrast between tissues and organs [23].

- HD-Live: which also allows a curved rendered mode using and enhances the depth perception and generates a realistic appearance [24].

These imaging modalities were chosen because they were preferred by independent expert opinions in the first CUME study [22]. Each one of the observers (MACN and WCB) prepared half of the images and all the images were reviewed by the other two authors (AL and WPM). They confirmed that the less experienced observers obtained the correct coronal view of each uterus with visible both uterine cavity horns, cervical canal, and internal cervical os, and if needed corrections were requested. The two images for each uterus were combined into a single image (set side by side; the example is presented in Figure 2) and sent to the fifteen experts.

The prepared images of coronal view of the uteri were sent to the invited experts for subjective judgment, because in this level of expertise (especially clinicians, and surgeons) commonly only images from imaging methods are consulted for final diagnosis based on subjective opinion.

Unprocessed 3D volumes were evaluated off-line by two expert raters (A.L and W.P.M.) independently, who performed the measurements.

Sample size

The sample size of subjects (N = 100), raters for subjective judgments (N = 15), raters for preparing the images to be sent to the experts (N=2) and raters for performing the measurements (N = 2) is standardized for our CUME exploratory studies for specific definitions and based on the available guidelines for reliability/agreement studies [25, 26]. However, there are reasonable arguments respected by experts in the field of reliability studies that a sample > 50 subjects, > 3 raters is rarely worth of attempts to reach more precise reliability coefficients [27]. We arbitrary increased our target sample size to 100 subjects to achieve an interval of confidence of 95% to the concordance correlation coefficient (CCC) of 0.2, considering the expected CCC of at least 0.80 by two raters. The same explanation for the target sample size of expert raters was the same as in the previous study: we arbitrary selected the sample size of expert raters (15 raters; 5 raters for field) to provide more generalizable results when compared to the participation of only three experts.

Principles for the selection of experts for subjective judgements

The principles and eligible criteria for experts were the same as previously described for building the CUME group based on reliable information about expertise [22]. We invited to join the CUME group: (i) Editors in Chief/Deputy Editors/Members of Editorial Board of journals with the highest impact in the fields: Gynecology and Reproductive Medicine/Imaging/Gynecological Surgery; (ii) Presidents or Members of Executive Committee of the targeted societies in the field; (iii) globally well-known experts in these fields due to their publications about uterine anomalies; and (iv) with at least 50 publications in the field of Obs/Gyn/Surg/Imaging; and (v) consent to participate. Experts involved in previous consensuses on the measurable criteria for congenital uterine malformations with known potential personal or financial conflict of interest regarding promotion of specific criteria, society or surgical procedures (surgery of T-shaped uterus), close friends, and co-workers from the same institutions were not considered as eligible. Hence, we have used the same initial and supplemental list of experts that was previously used in the first CUME study [22]. If somebody from the initial list did not agree to participate or overdue time for initial response (7 days), we consecutively invited another expert from the supplemental list, who was representative for the same field. To avoid bias, the experts were blinded to the opinion of the other experts. *Rating process*

A dichotomous question (T-shaped vs. Not T-shaped) was used in an online questionnaire form (https://goo.gl/forms/s947iFtit5T7rwBf2), which was created for this study and the link was sent to the members of the expert panel (CUME group).

Clinical definitions for experts

Clinical definitions for T-shaped uterus regarding additional morphologic features, benchmarks, potential reproductive outcomes of this morphology were not considered as appropriate to define and recognize T-shaped uterus because there are no objective criteria on how to identify T-shaped uterus and there is no evidence regarding whether T-shaped uterus (not DES-related) is associated with worse reproductive outcomes. We have also avoided any suggestions that T-shaped uterus is congenital uterine malformation because there is no evidence that this morphology is a result of embryologic disorders. The experts were asked to subjectively identify uterine cavities as (i) T-shaped or (ii) not T-shaped. The experts had no access to any measurement or clinical information; with the exception that none of the included women had been exposed to DES.

*Measurements and their selection*Two experts (A.L., W.M.P), named observer 1 and observer 2, who were blinded to the opinion of the other experts, independently proposed potentially usefulness measurements for differentiation of T-shaped uterus from normal/arcuate uterus. After discussion, a standardized set of fifteen measurements (eleven types of measurements considering four symmetrical measurements performed on the left and right side) and their methodology was established. Measurements independent from measurements of uterine cavity features and without direct association with uterine cavity, as uterine wall thickness were excluded from further consideration as methodologically unsound, because the uterine cavity shape and deformity degree are commonly independent from uterine wall thickness.

The proposed measurements were described in Table 1. Detailed scheme and definitions of each measurement is presented in Figure 3. Symmetrical measurements from left and right side were used for calculation the average value of these measurements. Selected measurements were used for creation indexes: uterine cavity area/uterine cavity circle area, and uterine cavity circumference/uterine cavity circle area.

The fundal internal indentation depth was defined as the distance between the intercornual line (line connecting the highest point of the endometrial cavity in each side of the uterus) and the lowest point of the internal fundal indentation [22, 28]. The lateral indentation depth was measured as the distance between two lines (the line that connects the most lateral point of the uterine cavity to the internal os and the apex of the lateral indentation). The circle area of the uterine cavity was measured as a circle touching these three points: most lateral points of the uterine cavity and the internal os. The uterine cavity area was manually traced including all the external limits of the endometrial cavity and the inferior limit was the internal os. The uterine cavity circumference was measured by manual tracing. The uterine cavity length was measured as being the distance between the intercornual line to the internal os. The uterine cavity width was defined as the distance between the most lateral points of the uterine cavity. The fundal indentation angle was an angle of three points including the lowest point of the internal fundal indentation and two points on the myometrium/endometrium interface 5 mm distant from the first one. The lateral indentation angle was calculated based on three points: the apex of lateral indentation and two points on the myometrium/endometrium interface 5 mm distant from the first one. The cornual angle was calculated using these three points: the most lateral point of the uterine cavity and two points on the myometrium/endometrium interface 5 mm distant from the first one. The T-angle was calculated using 3 points: the most lateral point of the uterine cavity, the most contralateral point of the uterine cavity and the apex of ipsilateral lateral indentation.

We determined the position of the internal os by subjective impression based on three criteria: the region of echogenicity transition (myometrium is whither and cervix is darker), the position of uterine vessels, and narrowest point of the “hourglass sign” formed between uterine and cervical cavities.

Two observers (A.L. and W.P.M.) independently manipulated the same initial dataset of each uterus to obtain the midcoronal plane of the uterus using Omniview with VCI mode to perform the proposed measurements. The Omniview with VCI was selected for the measurements because it is easier to manipulate than HDlive and it facilitates the identification of the internal os.

Bayes' theorem and T-shaped uterus definition

Assessment of reliability/agreement and accuracy and finding of the best measurements was the first step in the development of T-shape uterus definition to select the best considerable criteria. Finally, Bayes' theorem integrated by Fagan [29] was used to quantify the post-test probability of T-shaped uterus and establish T-shaped uterus definition considering the reasonable post-test probability of T-shaped uterus for women in the target population. This probability means the probability of having the condition [29]. This approach is referenced in evidence-based medicine and clinically applied epidemiology for interpretation of diagnostic accuracy studies [30].

Statistical analysis

Analyses were conducted using GraphPad Prism version 6 (GraphPad Software Inc., San Diego, CA, USA), IBMSPSS Statistics 22 (IBM Corp., Armonk, NY, USA) and Stata version 13.0 (StataCorp LP, College Station, TX,USA). Agreement between all experts of the CUME panelists, and agreement between experts of each group of expertise (separately for clinicians, surgeons and sonologists) were expressed by kappa index (Fleiss kappa for multiple raters) and proportion of agreement (percent agreement: p_o) [31]. The margins of variability/error between the two experts who performed the measurements were estimated by limits of agreement. Inter-observer reliability and agreement of the uterine measurements were expressed by concordance correlation coefficient (CCC) and limits of agreement (LoA) and Coefficient of Repeatability (CoR). The interpretation of agreement/reproducibility was used as follows [32]:

Poor: Kappa < 0.2; ICC/CCC < 0.7; CoR (relative differences) > 50%

Fair: Kappa 0.2-0.4; ICC/CCC 0.7-0.9; CoR (relative differences) 20-50%

Moderate: Kappa 0.41-0.60; ICC/CCC 0.90-0.95; CoR (relative differences) 10-20%

Good: Kappa 0.61-0.80; ICC/CCC 0.95-0.99; CoR (relative differences) 5-10%

Very good: Kappa 0.81-1.00; ICC/CCC > 0.99; CoR (relative differences) <5%.

The area under the receiver–operating characteristics curve (AUC) was used to estimate diagnostic test accuracy, and Youden's index was used to find the best cut-off value, using the most voted option by the 15 experts as the reference standard. Sensitivity, specificity, negative and positive-likelihood ratio (LR- and LR+) were calculated. A Fagan nomogram generated online by Schwartz's calculator (http://araw.mede.uic.edu/cgi-bin/testcalc.pl) was used to estimate how much most accurate criteria and their combination change the probability of uterus being T-shaped (posterior probability) in target population [29]. Prior probability was based on prevalence of T-shaped uterus according to diagnosis made most often by experts in target population (suspected to have T-shaped uterus). Agreement, diagnostic accuracy of each of the invited experts (subjective judgments of single expert), diagnoses based on the most accurate measurements and combinations of these measurements in comparison with the most voted option as the reference were expressed by kappa, and their accuracy were compared using McNemar test. Statistical significance was defined by $P \leq 0.05$.

RESULTS

3D datasets of 2179 uteri of consecutive women were assessed as potentially eligible to select the sample of eligible 100 subjects of 3D datasets of uteri with the presence of lateral internal indentation that were included in the study and analyzed.

Diagnosis of T-shaped uterus

There were 20 T-shaped and 80 normal/arcuate uteri using as reference standard the most voted option (8 or more votes) by the 15 experts. The median of diagnoses per expert was 19, while the minimal and maximal numbers of votes for T-shaped uterus were 5 and 35, respectively. Number of votes for T-shaped uterus per expert, including division on sonologists, clinicians, and surgeons is presented in Table 2.

Agreement between experts (*subjective* evaluation)

From the 1500 votes, there were 305 votes for T-shaped uterus (20%) with moderate level of agreement (kappa = 0.43) and the observed agreement was 82% (Table 3). There were differences in the agreement expressed by kappa among the groups of experts: sonologists and clinicians had moderate level of agreement (0.45 and 0.48, respectively), while surgeons achieved fair level (0.31). Percent agreement (p_o) was close to 80%, and the highest percent agreement was between expert sonologists (86%) in comparison to surgeons (80%) and clinicians (79%). Clinicians more often diagnosed T-shaped uterus (28%) than surgeons (17%) and sonologists (16%).

Inter-observer reliability/agreement of measurements

Results of all tested measurements by the observers, and the inter-observer reliability/agreement of all these measurements are shown in Table S2 and S3, respectively. The reliability of the measurements with reasonable accuracy for the diagnosis of T-shaped uterus was the following: CCC = 0.82 for average lateral cavity indentation depth, CCC = 0.62 for average lateral cavity indentation angle, and CCC = 0.87 for T-angle (Table 4). Relative and absolute differences between the measurements of the two observers for the most accurate measurements are shown in Table 4 and Figure 4.

Diagnostic test accuracy of measurements

The diagnostic test accuracy of all measurements is shown in Table S4. We found that the lateral indentation angle (AUROC=0.95), the lateral indentation depth (AUROC=0.92) and T-angle (AUROC=0.87) have good diagnostic test accuracy compared to subjective impression by experts (Table 5). The best cut-offs values for these measurements were (Table 6): lateral indentation angle ≤ 130° (sensitivity = 80%, specificity = 96%, LR- = 0.21, LR+ = 21), lateral indentation depth ≥ 7 mm (sensitivity = 95%, specificity = 78%, LR- = 0.06, LR+ = 4.2), and T-angle ≤ 40° (sensitivity = 80%, specificity = 87%, LR- = 0.23, LR+ = 6.4). Table S5 contains the best cut-offs for the selected measurements and their agreement with experts´ opinion.

The criteria, its frequency and diagnosis

The number of most accurate measurements, which meet the cut-offs of T-shaped uterus using the most voted option as reference were following:

- No criteria: N = 56, all of them were normal/arcuate uterus
- Only one criteria: N=20, 2 = T-shaped, 18 = normal/arcuate uterus
- Two criteria: N = 10, 5 = T-shaped, 5 = normal/arcuate uterus
- Three criteria: N = 14, 13 = T-shaped, 1 = normal/arcuate uterus

Posttest probability of T-shaped uterus by most accurate measurements and their combination

The estimated post-test probability of T-shaped uterus based on the known pre-test probability and the LR+ (for a presence of the criterion) or LR- (for an absence of the criterion) values using the Fagan nomogram for target population is presented for each criterion (diagnosis of T-shaped uterus = 1 of out 3 criteria), and if they used in combination (diagnosis of T-shaped uterus = 3 of out 3 criteria) in Figure 5. Post-test probability of uterus being T-shaped for presence of lateral indentation angle ≤ 130°, lateral indentation depth ≥ 7 mm, and T-angle ≤ 40° is 83%, 49% and 64%, respectively, and when the lack of these criteria, it is 6%, 5% and 9%, respectively. Post-test probability of uterus being T-shaped for presence of combination of three criteria is 93%, and when the lack of this combination (negative test results), the post-test probability is 8%.

Comparison of single experts' opinion, with measurements, their combination, and CUME reference

The diagnostic accuracy measures, agreement expressed by kappa and post-test probability of diagnoses made by single expert, single criterion (1 out of 3) and combination of 3 criteria relative to the most voted option as CUME reference is shown in Table 7. The present of at least 2 criteria or all 3 criteria had better than most of experts (13 and 11, respectively) with the most voted option (Kappa = 0.77 and 0.72). The proportion of T-shaped uterus from several single experts was significantly different than the observed using the most voted option (McNemar test, P < 0.05), whereas combination of measurements were not significantly different (McNemar test, P > 0.05).

Final reasoning and T-shaped uterus definition

We assume according to Bayes' theorem that a true positive diagnosis with specific morphometric criteria with good accuracy reflects the probability of a patient having a condition.

Using a single criterion by CUME, the probability that women to having T-shaped uterus is 50-80%, while the probability that the uterus is T-shaped if the CUME criteria are not present is very low (Figure 5). The post-test probability of T-shaped uterus based on a single criterion of T-shaped uterus by CUME should be considered as insufficient for diagnosis. For instance, if we use the internal indentation depth, 50% of women from those initially suspected to have T-shaped uterus will have a T-shaped uterus, and 50% of women will have a normal uterus. As an example, if we include both women in a surgical randomized controlled trial, we will treat both women with T-shaped uterus and with normal uterus. We may not find the real benefit of the intervention for women with T-shaped uterus, and women with a normal uterus will receive iatrogenic treatment. Moreover, in our sample 2/20 women (10%), 5/10 women (50%) and 13/14 women (93%) that one criterion, two criteria, and three criteria, respectively, had T-shaped uterus by the most voted option. Based on the study results we suggest three definitions of the normal/arcuate uterus, borderline T-shaped uterus and T-shaped uterus (Table 8). Also, with the suggested diagnosis based on three measurements it is possible to have a better agreement with the most voted option than did most of single experts who voted (Table 7).

DISCUSSION

This study shows that the diagnosis of T-shaped uterus is challenging. The level of agreement between top experts asked to subjectively distinguish between T-shaped and not T-shaped uterus based on the assessment of coronal view of the uterus from three-dimensional ultrasound is moderate (kappa = 0.43) and percent agreement is equal 80%. To our knowledge, this is the first study performed to objectively define T-shaped uterus. The findings of this study are also relevant to establish a definition of a main morphologic variant of normal uterus regarding lateral uterine cavity indentation.

The study has identified three measurements that had good diagnostic accuracy and fair to moderate inter-observer reproducibility: lateral indentation angle (AUROC = 0.95, CCC = 0.62, CoR = 14%), lateral indentation depth (AUROC = 0.92, CCC = 0.79, CoR = 26%) and T-angle (AUROC = 0.87, CCC = 0.87, CoR = 18%). Based on the study results we suggest three definitions of the normal/arcuate uterus, borderline T-shaped uterus and T-shaped uterus for research (Table 8) and we establish the CUME definition of T-shaped uterus as follow: lateral indentation angle < 130^o; lateral indentation depth > 7 mm and T-angle ≤ 40^o (Figure 6 and Figure 7, and Figure S1).

With the suggested diagnosis based on three measurements it is possible to have a better agreement with the most voted option than did most of single experts who voted. Reproducibility of these measurements is fair to moderate [32], even considered well-trained observers. Therefore, it is important to have some training before start measuring and trying to develop technical refinements to improve it (for example using an average of three measurements or by the creation of software for automated measurements).

There are some important limitations of this study: (i) although the present study was able to define objective measurable criteria to diagnose T-shaped uterus with moderate agreement according to top experts opinion, the clinical relevance of the non-DES-related T-shaped uterus is unknown; (ii) measurements were performed by highly trained raters and using a single 3D ultrasound dataset for each uterus which may not reflect what happens in the “real world”, in which many sources of variability are present what may lead to a negative impact on diagnostic accuracy and reliability of diagnoses of uterine anomalies; (iii) we used only data-sets with subjectively detected lateral indentation without any other subjectively detected anomaly; (iv) for the lateral indentation depth, internal os was used as a reference point, and sometimes the identification of the internal os might be difficult to determine. Therefore, the observed results should not be generalized for screening purposes or for uterus with other anomalies, despite that the sample represents a wide spectrum of uteri, which may be considered as normal/arcuate by ASRM classification, and normal uterus artificially labeled as septate uteri by the ESHRE/ESGE classification[11, 22, 33]. The decision to include specific cases might have led to some selection bias.

The study may have an important contribution in the development of common language for uterine anomalies and further development of single global classification system with the unification of the criteria of divided local classifications as the ESHRE-ESGE and ASRM. Moreover, the study results can help on understanding and validating the quality of evidence about T-shaped uterus, and performing meaningful and better studies on this condition than so far. Someone may state that non-DES-related T-shaped uterus is not a congenital anomaly class, created by surgeons [19], as a group with competing interests to spread or justify lucrative surgical procedures on such condition by labeling normal uterus to ‘congenital malformation’. Further efforts should be targeted to recognize whether T-shaped uterus is a true uterine anomaly with clinical consequences.

The reliability of all studies regarding prevalence, clinical implication, and surgical outcomes may be interpreted in the context of this CUME study. So far, the diagnosis has been only subjective. However, the study has shown that the agreement of subjective judgments is not very high, even among top experts (Figure 8). An ideal solution would be having a subjective judgment from several experts, but this is impractical in clinical practice. Therefore, the proposed CUME criteria might be used as a surrogate of multi-expert opinion to achieve similar results, and with better accuracy and agreement with reference than most single expert opinion.

The definition of T-shaped uterus proposed by ESHRE/ESGE, that is a ‘narrow uterine cavity due to thickened lateral walls with a correlation 2/3 uterine corpus and 1/3 cervix’ [19] is arbitrary and was not based on scientific support. By using this definition, it is impossible to differentiate T-shaped uterus from the normal uterus, and from uterus infantilis, because the reference value of normal rine wall thickness and any cut-offs for differentiation of T-shaped uterus with other uterine morphologies were not provided. Interestingly, the ESHRE/ESGE definition was used retrospectively in studies [12, 17, 34], however, its application in real clinical practice to differentiate between T-shaped uterus and normal uterus is impossible. Before the ESHRE-ESGE classification T-shaped uterus was recognized according to DES etiology and subjective impression by AFS-1988 [8, 11].

Using the results of the present study, it will be much easier to study the prevalence of T-shaped uterus and its clinical relevance. Until these studies are available, it is suggested not to treat T-shaped uterus as an anomaly, but as a variant of normal uterus, to avoid unnecessary interventions in infertile women or in women with previous miscarriage. Such studies are crucial in the era of arbitrary creating of over-definition of diseases, and such fake disease mongering [35].

Conclusions

The diagnostic of T-shaped uterus is not easy and the agreement among top-experts is only moderate, and single expert judgment is commonly insufficient for accurate diagnosis. The present study has identified three measurements with cut-offs that had good diagnostic test accuracy and fair to moderate reliability (lateral indentation depth ≥ 7 mm, lateral indentation angle ≤ 130° and T-angle ≤ 40°), and when applicated together they have provided high post-test probability of this condition. Based on the CUME criteria of T-shaped uterus, the prevalence, clinical implication and the management of this condition may be determined with enough accuracy, reliability and with a known probability of disease after negative and positive test results. The CUME definition of T-shape uterus may help on the development of interventional randomized controlled trials, observational studies and diagnostics of uterine morphology in every day practice, and therefore could be adopted by guidelines on uterine anomalies to enrich their classification systems.

ACKNOWLEDGMENTS

We want to thank Marzena Stryjak (Ludwin & Ludwin Gynecology, Krakow, Poland) for their help in data management.

FUNDING

This study was supported by Jagiellonian University (grant no. K/ZDS/005568).

References

1. Schultze GKF, Erbsloh J. *Gynakologische Rontgendiagnostik*: Stuttgart, Ferdinand Enke; 1954.
2. Erbsloeh J. Normal shapes, variations and malformations in hysterosalpingography. *Radiologe* 1975; **15**: 2-10.
3. Kaufman RH, Adam E. Genital tract anomalies associated with in utero exposure to diethylstilbestrol. *Isr J Med Sci* 1978; **14**: 353-362.
4. Kaufman RH, Binder GL, Gray PM, Jr., Adam E. Upper genital tract changes associated with exposure in utero to diethylstilbestrol. *Obstet Gynecol Surv* 1977; **32**: 611-613.
5. Acien P, Acien MI. The history of female genital tract malformation classifications and proposal of an updated system. *Hum Reprod Update* 2011; **17**: 693-705.
6. Buttram VC, Jr., Gibbons WE. Mullerian anomalies: a proposed classification. (An analysis of 144 cases). *Fertil Steril* 1979; **32**: 40-46.
7. Fernandez H, Garbin O, Castaigne V, Gervaise A, Levaillant JM. Surgical approach to and reproductive outcome after surgical correction of a T-shaped uterus. *Hum Reprod* 2011; **26**: 1730-1734.
8. AFS. The American Fertility Society classifications of adnexal adhesions, distal tubal occlusion, tubal occlusion secondary to tubal ligation, tubal pregnancies, mullerian anomalies and intrauterine adhesions. *Fertil Steril* 1988; **49**: 944-955.
9. Al Jishi T, Sergi C. Current perspective of diethylstilbestrol (DES) exposure in mothers and offspring. *Reprod Toxicol* 2017; **71**: 71-77.
10. van Gils AP, Tham RT, Falke TH, Peters AA. Abnormalities of the uterus and cervix after diethylstilbestrol exposure: correlation of findings on MR and hysterosalpingography. *AJR Am J Roentgenol* 1989; **153**: 1235-1238.
11. Ludwin A, Ludwin I. Comparison of the ESHRE-ESGE and ASRM classifications of Mullerian duct anomalies in everyday practice. *Hum Reprod* 2015; **30**: 569-580.
12. Di Spiezio Sardo A, Florio P, Nazzaro G, Spinelli M, Paladini D, Di Carlo C, Nappi C. Hysteroscopic outpatient metroplasty to expand dysmorphic uteri (HOME-DU technique): a pilot study. *Reprod Biomed Online* 2015; **30**: 166-174.
13. Ahmadi F, Zafarani F, Shahrzad GS. Hysterosalpingographic Appearances of Female Genital Tract Tuberculosis: Part II: Uterus. *Int J Fertil Steril* 2014; **8**: 13-20.
14. Puente JM, Fabris A, Patel J, Patel A, Cerrillo M, Requena A, Garcia-Velasco JA. Adenomyosis in infertile women: prevalence and the role of 3D ultrasound as a marker of severity of the disease. *Reprod Biol Endocrinol* 2016; **14**: 60.
15. Barranger E, Gervaise A, Doumerc S, Fernandez H. Reproductive performance after hysteroscopic metroplasty in the hypoplastic uterus: A study of 29 cases. *BJOG: An International Journal of Obstetrics and Gynaecology* 2002; **109**: 1331-1334.
16. Boza A, Akin OD, Oguz SY, Misirlioglu S, Urman B. Surgical correction of T-shaped uteri in women with reproductive failure: Long term anatomical and reproductive outcomes. *J Gynecol Obstet Hum Reprod* 2019; **48**: 39-44.
17. Ducellier-Azzola G, Lecointre L, Hummel M, Pontvianne M, Garbin O. Hysteroscopic enlargement metroplasty for T-shaped uterus: 24 years' experience at the Strasbourg Medico-Surgical and Obstetrical Centre (CMCO). *European Journal of Obstetrics Gynecology and Reproductive Biology* 2018; **226**: 30-34.
18. Sukur YE, Yakistiran B, Ozmen B, Sonmezer M, Berker B, Atabekoglu C. Hysteroscopic Corrections for Complete Septate and T-Shaped Uteri Have Similar Surgical and Reproductive Outcome. *Reprod Sci* 2018; 1933719118756774.

19. Grimbizis GF, Gordts S, Di Spiezio Sardo A, Brucker S, De Angelis C, Gergolet M, Li TC, Tanos V, Brolmann H, Gianaroli L, Campo R. The ESHRE/ESGE consensus on the classification of female genital tract congenital anomalies. *Hum Reprod* 2013; **28**: 2032-2044.
20. Salim R, Woelfer B, Backos M, Regan L, Jurkovic D. Reproducibility of three-dimensional ultrasound diagnosis of congenital uterine anomalies. *Ultrasound Obstet Gynecol* 2003; **21**: 578-582.
21. Practice Committee of the American Society for Reproductive Medicine. Electronic address Aao, Practice Committee of the American Society for Reproductive M. Uterine septum: a guideline. *Fertil Steril* 2016; **106**: 530-540.
22. Ludwin A, Martins WP, Nastri CO, Ludwin I, Coelho Neto MA, Leitao VM, Acien M, Alcazar JL, Benacerraf B, Condous G, De Wilde RL, Emanuel MH, Gibbons W, Guerriero S, Hurd WW, Levine D, Lindheim S, Pellicer A, Petraglia F, Saridogan E. Congenital Uterine Malformation by Experts (CUME): better criteria for distinguishing between normal/arcuate and septate uterus? *Ultrasound Obstet Gynecol* 2018; **51**: 101-109.
23. Jouannic JM, Rosenblatt J, Demaria F, Jacobs R, Aubry MC, Benifla JL. Contribution of three-dimensional volume contrast imaging to the sonographic assessment of the fetal uterus. *Ultrasound Obstet Gynecol* 2005; **26**: 567-570.
24. Kagan KO, Pintoffl K, Hoopmann M. First-trimester ultrasound images using HDlive. *Ultrasound Obstet Gynecol* 2011; **38**: 607.
25. Cocchetti DV. Sample size requirements for increasing the precision of reliability estimates: problems and proposed solutions. *J Clin Exp Neuropsychol* 1999; **21**: 567-570.
26. Donner A, Rotondi MA. Sample size requirements for interval estimation of the kappa statistic for interobserver agreement studies with a binary outcome and multiple raters. *Int J Biostat* 2010; **6**: Article 31.
27. Cicchetti DV. The precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. *J Clin Exp Neuropsychol* 2001; **23**: 695-700.
28. Ludwin A, Ludwin I, Coelho Neto MA, Nastri CO, Bhagavath B, Lindheim SR, Martins WP. Septate uterus by updated ESHRE/ESGE, ASRM and CUME definitions: association with infertility, previous miscarriage, and warnings for women and healthcare systems, and associated cost analysis. *Ultrasound Obstet Gynecol* 2019;
29. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med* 1975; **293**: 257.
30. Caraguel CG, Vanderstichel R. The two-step Fagan's nomogram: ad hoc interpretation of a diagnostic test result without calculation. *Evid Based Med* 2013; **18**: 125-128.
31. Kottner J, Gajewski BJ, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS). *Int J Nurs Stud* 2011; **48**: 659-660.
32. Martins WP, Nastri CO. Interpreting reproducibility results for ultrasound measurements. *Ultrasound Obstet Gynecol* 2014; **43**: 479-480.
33. Knez J, Saridogan E, Van Den Bosch T, Mavrelos D, Ambler G, Jurkovic D. ESHRE/ESGE female genital tract anomalies classification system-the potential impact of discarding arcuate uterus on clinical practice. *Hum Reprod* 2018; **33**: 600-606.
34. De Bruyn C, Greet M, Willem O, Rudi C. The impact of hysteroscopic surgery for dysmorphic uterus on reproductive and obstetric outcomes: Pilot study. *Gynecological Surgery* 2013; **10**: S101-S102.
35. Brodersen J, Schwartz LM, Heneghan C, O'Sullivan JW, Aronson JK, Woloshin S. Overdiagnosis: what it is and what it isn't. *BMJ Evid Based Med* 2018; **23**: 1-3.

Figure legends

Figure 1 Three basic types of the normal uterine cavity according to experience from hysterosalpingography (Schultze and Erbsloh, 1954). (A) triangle shaped cavity, (B) with small fundal and lateral indentations, and (C) with larger fundal and lateral indentations

Figure 2 Images of the presentation of two uteri sent to the experts in two 3D ultrasound modalities: HDlive on the left and Omniview with Volume Contrast Imaging (VCI) on the right. The diagnosis made most often by experts was used as a reference standard and recognition of normal uterus (top row: more votes than not T-shaped: 14/15) and T-shaped uterus (bottom row: more votes for T-shaped, 15/15).

Figure 3 Set of tested measurements for recognition of T-shaped uterus. There were: internal fundal indentation depth (a); right internal lateral indentation depth (b); left internal lateral indentation depth (c); circle area of the uterine cavity (ellipse) (d); uterine cavity area (area trace) (e); uterine cavity circumference (f); uterine cavity length from intercornual to internal os (g); uterine cavity width (intercornual) (h); fundal angle and right and left lateral indentation angle (i); right and left cornual angle (j); right T-angle (k); left T-angle (l). Symmetrical measurements from the left and right side were used for calculation the average value of these measurements. Selected measurements were used for creation indexes: uterine cavity area/uterine cavity circle area, and uterine cavity circumference/uterine cavity circle area.

Figure 4 Bland-Altman plots for the absolute (top row) and relative (bottom row) difference observed between measurements of the two observers for the most accurate measurements (lateral indentation depth, lateral indentation angle, and T-angle: average measurements of left and right side).

Figure 5 Fagan nomogram to determine the probability of a woman having T-shaped based on the most voted opinion of experts [29]. (A) Lateral indentation angle, (B) lateral indentation depth, (C) and T-angle as single diagnostic criteria, and (D) CUME definition of T-shaped uterus (presence of three criteria). The straight line runs from the patient's pre-test probability of T-shaped uterus in cases suspected as being T-shaped uterus by presence of lateral cavity indentation (that is, 20%) through the positive (red line) and negative likelihood ratio (blue line) for presence one of the criterion: lateral indentation angle < 130, lateral indentation depth > 7 mm, and T-angle > 40 and absence of these criteria, to the post-test probability of disease. The post-test probability of uterus being T-shaped is more than two-four times higher with the presence of criteria of lateral angle (80%), depth (50%) and T-angle (60%), and it is low (1–5%) for the absence of all criteria.

Figure 6 T-shaped uterus by CUME. The presence of three criteria: lateral indentation angle, lateral indentation depth, and T-angle with these cut-offs is defined as T-shaped uterus. The presence of 2 out of 3 criteria is defined as borderline T-shaped uterus, whereas the lack or presence only the single criterion is defined as normal uterus considering lateral uterine morphology. This definition is supported by the reliability and diagnostic test accuracy testing of measurements, and finally by the post-test probability of the condition >90% for target population considering the diagnosis made most often by top experts as reference.

Figure 7 How to perform the proposed measurements in real 3D-US using Omniview with VCI. (A) lateral indentation angle (right and left). (B) Lateral indentation depth (right and left). (C) T-angle (right and left).

Figure 8 Examples of uteri with the diagnoses made most often by experts and recognition of normal and T-shaped uterus with very good (12-15 of 15 votes) and good agreement (8-11 of 15 votes) with decreasing votes for normal uterus from top to bottom row.

Supplemental Figure 1. A manner of measurements to diagnose T-shaped uterus in three variants of uterus with lateral indentations: (left column) with concomitant small fundal indentation, and (middle column) with cornues of uterine cavity on the top of uterine cavity, and (right column) with cornues of uterine cavity below the top of uterine cavity. (Top row) T-angle, (Middle row) lateral indentation depth, (Bottom row) and lateral indentation depth and their manner of measurements.

Table 1 Description of tested measurements for differentiation of T-shaped uterus from normal/arcuate uterus.

Parameter	Unit	Description
Internal fundal indentation depth	Mm	Distance between two lines (intercornual line and nadir of fundal indentation)
Right internal LI depth	Mm	Distance between two lines (line 1 = internal os to most lateral point of the uterine cavity; line 2= line that crosses the apex of lateral indentation), right side
Left internal LI depth	Mm	Distance between two lines (line 1 = internal os to most lateral point of the uterine cavity; lina 2=line that crosses the apex of lateral indentation) left side
Circle area of the uterine cavity	cm²	Area of the circle touching 3 points (most lateral points of the uterine cavity and internal os)
Uterine cavity area	cm²	Area of the uterine cavity (manual drawing)
Uterine cavity circumference	Cm	Circumference of the uterine cavity (manual drawing)
Uterine cavity length	Mm	Distance between two lines (intercornual line and internal os)
Uterine cavity width	Mm	Width of the uterine cavity (distance between the two most lateral points of the endometrial cavity)
Fundal indentation angle	Degrees	Three points (apex of fundal indentation and two points on the myometrial/endometrial interface 5mm distant from this point for each side))
Right LI angle	Degrees	Three points (apex of lateral indentation and two points on the myometrial/endometrial interface 5mm distant from this point for each side)
Left LI angle	Degrees	Three points (apex of lateral indentation and two points on the myometrial/endometrial interface 5mm distant from this point for each side)
Right cornual angle	Degrees	Three points (most lateral point of the uterine cavity and two points on the myometrial/endometrial interface 5mm distant from this point for each side)
Left cornual angle	Degrees	Three points (most lateral point of the uterine cavity and two points on the myometrial/endometrial interface 5mm distant from this point for each side)
Right T-angle	Degrees	Three points (most lateral point of uterine cavity, contralateral most lateral point of uterine cavity and ipsilateral apex of lateral indentation)
Left T-angle	Degrees	Three points (most lateral point of uterine cavity, contralateral most lateral point of uterine cavity and ipsilateral apex of lateral indentation)
Cavity area to Circle area ratio	None	Area of the uterine cavity (manual drawing) divided by area of the circle touching 3 points (most lateral points of the uterine cavity and internal os)
Cavity area to Circumference Ratio	Cm	Area of the uterine cavity (manual drawing) divided by circumference of the uterine cavity (manual drawing)
Cavity area Circumference² Ratio	None	Area of the uterine cavity (manual drawing) divided by (circumference of the uterine cavity)²

LI = lateral indentation.

Table 2 Votes for T-Shaped uterus by subjective judgment of individual experts and field of expertise

	N	Median	p25-p75		Min-Max		Mean	SD	Sum
All	15	19	14	29	5	35	20.3	10.0	305
Sonologists	5	17	6.5	24	6	29	15.6	9.5	78
Clinicians	5	33	19.5	34.5	17	35	28.2	8.2	141
Surgeons	5	16	9.5	25.5	5	27	17.2	8.7	86

Number of uteri = 100, N = number of experts, p25-p75 = lower – upper quartile, Min = minimal number of votes, Max = maximal number of votes

Table 3 Agreement between experts recognizing T-shaped uterus using subjective judgment of coronal view of uterus from three-dimensional ultrasound

Expert	T-shaped selected n/N (%)	kappa	95% CI	Proportion of agreement*	95% CI
Clinicians (n=5)	141/500 (28%)	0.48	0.41-0.55	790/1,000 (79%)	76-81%
Sonologists (n=5)	78/500 (16%)	0.45	0.41-0.49	856/1,000 (86%)	83-88%
Surgeons (n=5)	86/500 (17%)	0.31	0.23-0.39	804/1,000 (80%)	78-83%
Overall (n=15)	305/1500 (20%)	0.43	0.39-0.47	8,566/10,500 (82%)	81-82%

Number of uteri = 100, Number of all experts = 15, Number of experts per group = 5, CI = Confidence interval

* Considering each subgroup of experts (n=5) there are 10 pairwise comparisons for each uterus (rater 1 vs. rater 2, rater 1 vs. rater 3, rater 1 vs. rater 4, …) resulting in a total of 1000 comparisons when assessing the 100 subjects. Considering all the fifteen experts, there are 105 pairwise comparisons for each uterus, resulting in a total of 10,500 comparisons.

Table 4 Inter-observer reliability/agreement of three selected most accurate measurements from potentially useful measurements to establish objective diagnostic criteria for T-shaped uterus

	Absolute differences		**Relative differences**		**CCC**	**95%CI**	
	LoA	**CoR**	**LoA**	**CoR**			
Lateral indentation depth	-1.0mm to 2.2mm	1.6mm	-17% to 36%	26%	0.82	0.75	0.87
Lateral indentation angle	-19° to 19°	19°	-14% to 14%	14%	0.62	0.48	0.73
T-angle	-9° to 7°	8°	-20% to 15%	18%	0.87	0.82	0.91

N = 100 uteri, LoA = limits of agreement; CoR = Coefficient of Repeatability; CCC = Concordance correlation coefficient, CI = Confidence interval

Table 5 Diagnostic test accuracy of three selected most accurate measurements from potentially useful measurements to establish objective diagnostic criteria for T-shaped uterus

	Area Under the Curve			Correlation with sum of votes
	Area	**95% CI**		**Spearman**
Lateral indentation depth	0.92	0.87	0.98	0.66
Lateral indentation angle	0.95	0.90	0.99	-0.73
T- Angle	0.87	0.78	0.95	-0.62

CI = Confidence interval

Table 6 Best cut-off points for the three most useful measurements to define T-shaped uterus

	Criteria	Sensitivity	95%CI	Specificity	95%CI	Youden index J
Lateral angle	≤130°	80% (16/20)	56-94%	96.3% (77/80)	89-99%	0.76
Lateral indentation depth	≥7mm	95% (19/20)	75-100%	77.5% (62/80)	67-86%	0.73
T- Angle	≤40°	80% (16/20)	56-94%	87.5% (70/80)	78-94%	0.68

CI = Confidence interval

Table 7 The diagnostic accuracy measures, agreement expressed by kappa and post-test probability of diagnoses made by single expert, single criterion (1 out of 3) and combination of 3 criteria relative to the most voted option as CUME reference

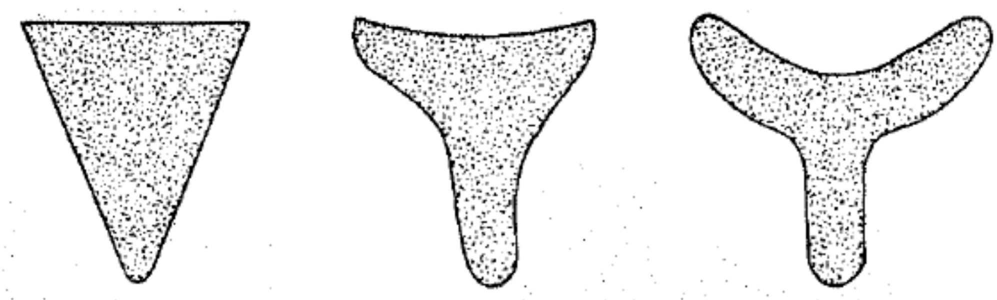
	tp	fn	tn	fp	OA	Sens	Spec	PPV	NPV	LR+	LR-	kappa	95% CI		p
Exp1	20	0	71	9	91%	100%	89%	69%	100%	8.9	0.00	0.76	0.61	0.90	0.004
Exp2	14	6	75	5	89%	70%	94%	74%	93%	11.2	0.32	0.65	0.46	0.84	1.00
Exp3	6	14	80	0	86%	30%	100%	100%	85%	>100	0.70	0.41	0.18	0.64	<0.001
Exp4	16	4	79	1	95%	80%	99%	94%	95%	64.0	0.20	0.83	0.69	0.98	0.375
Exp5	7	13	80	0	87%	35%	100%	100%	86%	>100	0.65	0.46	0.23	0.69	<0.001
Exp6	18	2	65	15	83%	90%	81%	55%	97%	4.8	0.12	0.57	0.40	0.75	0.002
Exp7	17	3	75	5	92%	85%	94%	77%	96%	13.6	0.16	0.76	0.60	0.92	0.727
Exp8	14	6	77	3	91%	70%	96%	82%	93%	18.7	0.31	0.70	0.52	0.88	0.508
Exp9	18	2	64	16	82%	90%	80%	53%	97%	4.5	0.13	0.55	0.38	0.73	0.001
Exp10	20	0	65	15	65%	100%	81%	57%	100%	5.3	0.00	0.63	0.48	0.79	<0.001
Exp11	14	6	70	10	84%	70%	88%	58%	92%	5.6	0.34	0.54	0.34	0.73	0.454
Exp12	9	11	75	5	84%	45%	94%	64%	87%	7.2	0.59	0.44	0.21	0.67	0.210
Exp13	13	7	77	3	90%	65%	96%	81%	92%	17.3	0.36	0.66	0.47	0.85	0.344
Exp14	5	15	80	0	85%	25%	100%	100%	84%	>100	0.75	0.35	0.12	0.58	<0.001
Exp15	20	0	73	7	93%	100%	91%	74%	100%	11.4	0.00	0.81	0.67	0.94	0.016
LI angle	16	4	77	3	93%	80%	96%	84%	95%	21.3	0.21	0.78	0.62	0.93	1.00
LI depth	19	1	62	18	81%	95%	78%	51%	98%	4.2	0.06	0.55	0.38	0.72	<0.001
Tangle	16	4	70	10	86%	80%	88%	62%	95%	6.4	0.23	0.61	0.42	0.79	0.18
≥ 1 criterion	20	0	56	24	76%	100%	70%	45%	100%	3.3	0.00	0.48	0.33	0.64	<0.001
≥ 2 criteria	18	2	74	6	92%	90%	93%	75%	97%	12.0	0.11	0.77	0.61	0.92	0.289
All 3 criteria	13	7	79	1	92%	65%	99%	93%	92%	52.0	0.35	0.72	0.54	0.90	0.070

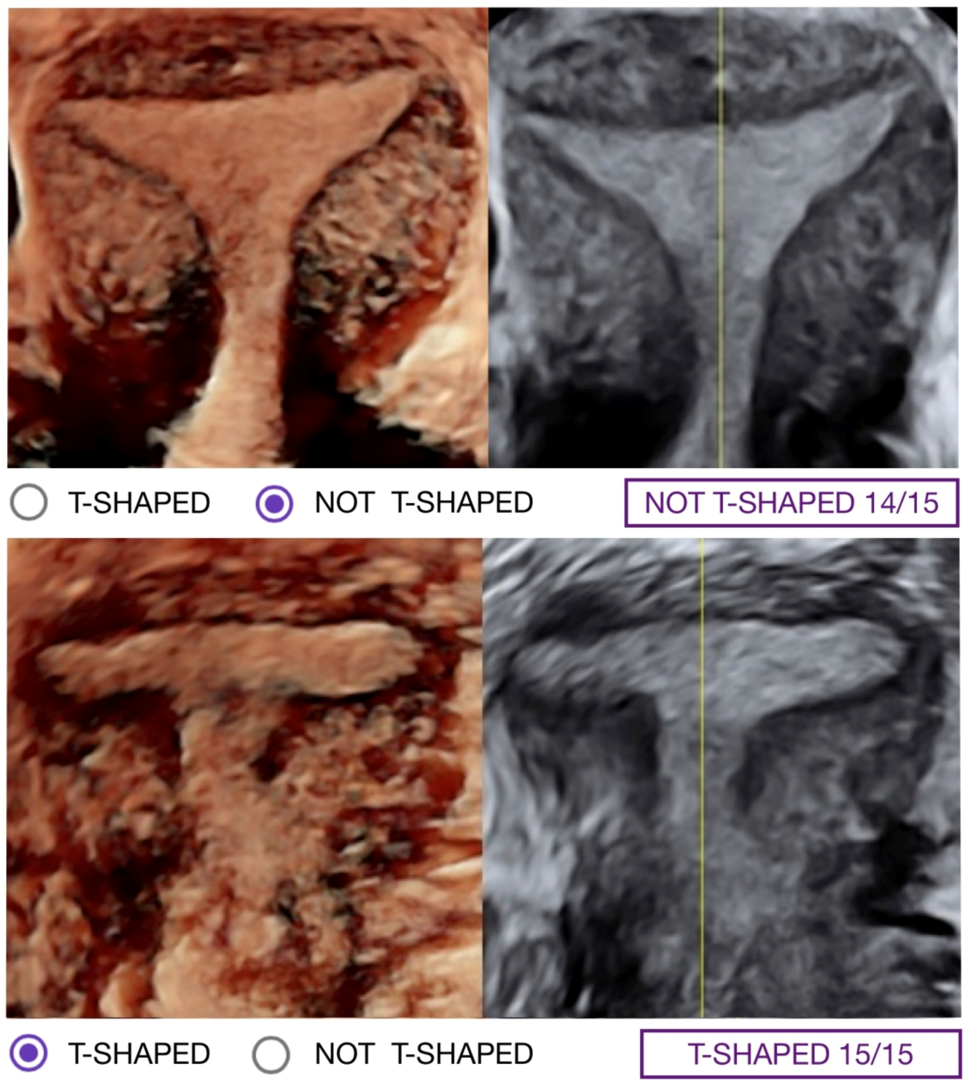
Number of uteri = 100, Exp = Expert; LI = lateral indentation; tp = true positive; fn = false negative; tn = true negative; fp = false positive; OA = observed agreement; Sens = sensitivity; Spec = specificity; PPV = Positive Predictive Value; NPV = Negative Predictive Value; LR+ = Positive Likelihood Ratio; LR- = Negative Likelihood Ratio

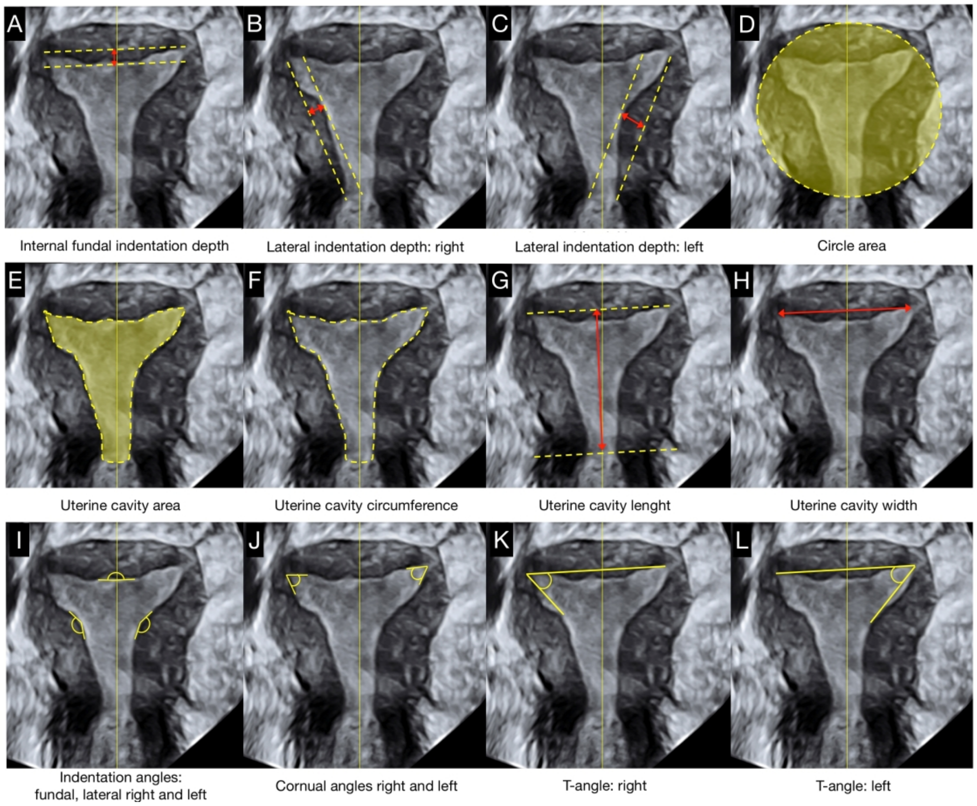
Table 8. Criteria and definitions proposed for classification of uterine morphology as normal/arcuate, borderline T-saped and T-shaped uterus by CUME

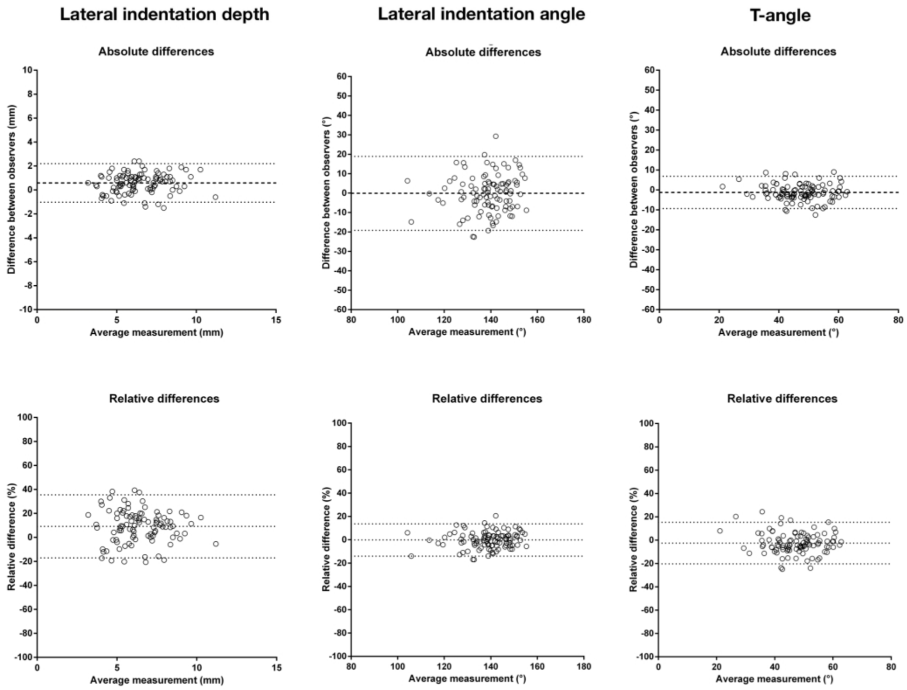
Criteria/Definitions*	Normal/arcuate	Borderline T-shaped	T-shaped
lateral indentation angle ≤ 130° lateral indentation depth ≥ 7 mm T-angle ≤ 40°	None or only 1 criterion	Only 2 criteria	All 3 criteria

* the lack of internal fundal indentation >1 cm according to CUME criteria for differentiation between normal/arcuate uterus

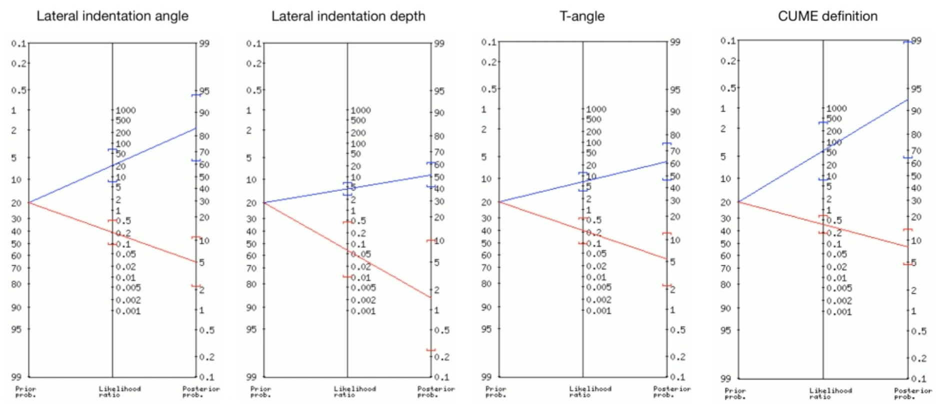


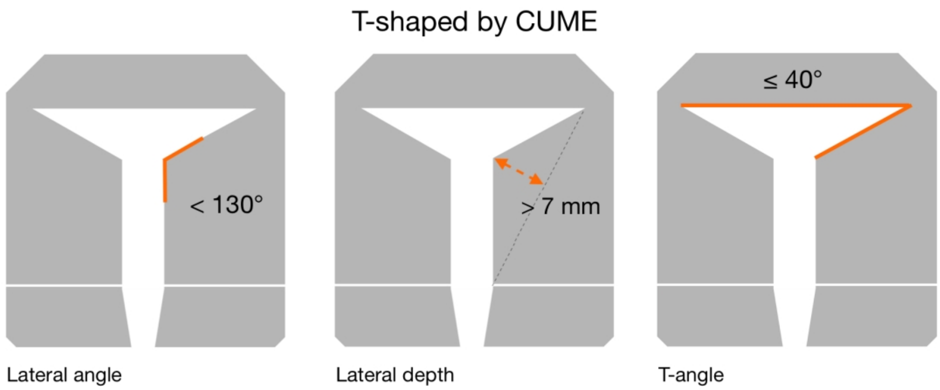

T-SHAPED
NOT T-SHAPED
NOT T-SHAPED 14/15
T-SHAPED
NOT T-SHAPED
T-SHAPED 15/15

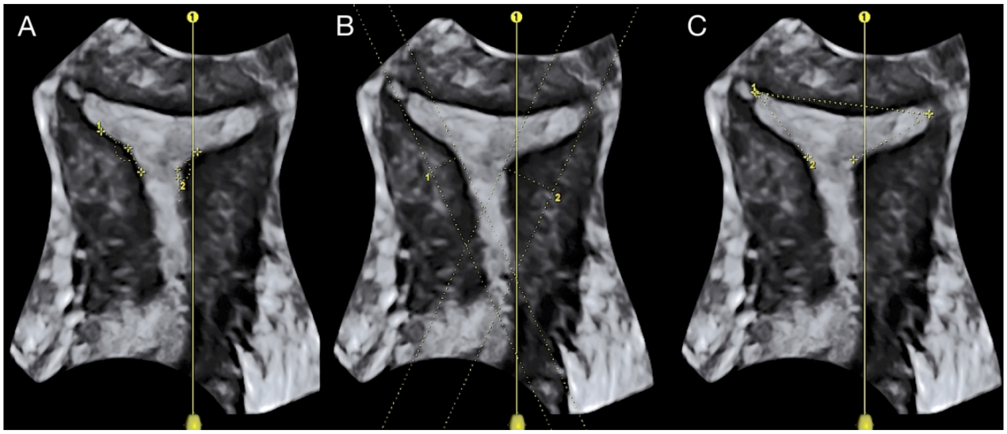

A
B
C
D
Internal fundal indentation depth
Lateral indentation depth: right
Lateral indentation depth: left
Circle area
E
F
G
H
Uterine cavity area
Uterine cavity circumference
Uterine cavity lenght
Uterine cavity width
I
J
K
L
Indentation angles:
fundal, lateral right and left
Cornual angles right and left
T-angle: right
T-angle: left

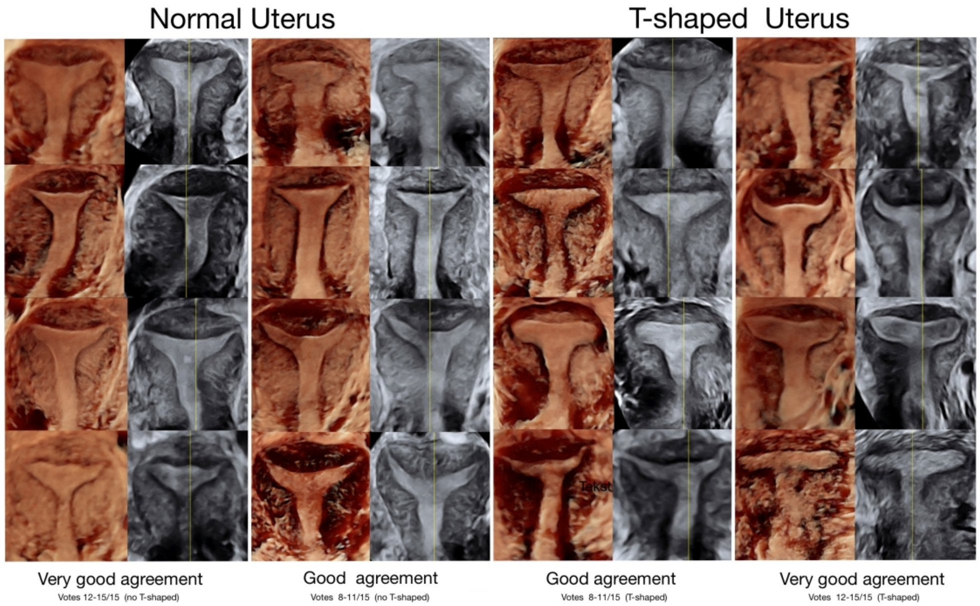


Lateral indentation depth
Lateral indentation angle
T-angle
Absolute differences
Relative differences
Difference between observers (mm)
Difference between observers (°)
Relative difference (%)
Average measurement (mm)
Average measurement (°)


Lateral indentation angle
Lateral indentation depth
T-angle
CUME definition
Prior prob.
Likelihood ratio
Posterior prob.


T-shaped by CUME
< 130°
> 7 mm
≤ 40°
Lateral angle
Lateral depth
T-angle


A
B
C


Normal Uterus
T-shaped Uterus
Very good agreement
Votes 12-15/15 (no T-shaped)
Good agreement
Votes 8-11/15 (no T-shaped)
Good agreement
Votes 8-11/15 (T-shaped)
Very good agreement
Votes 12-15/15 (T-shaped)