

Accuracy and Reliability of Volume Measurements by 3D Ultrasound: A Prospective Study Performed by Novices

Tho Quynh Nguyen, Melanie Flores, Melchor Ortiz, Soyoung Lee, Zuber D Mulla, Sanja Kupesic Plavsic

ABSTRACT

Purpose: This study evaluates the accuracy and reliability of off-line volume measurement using virtual organ computer-aided analysis (VOCAL™) and two-dimensional (2D) ultrasound by novices.

Methods: Thirty medical students, 15 sonographers and three experts were recruited to measure a liver, a gestational sac and a breast cancer lesion using 2D ultrasound and three-dimensional (3D) VOCAL™ method. The VOCAL™ measurements produced by the experts were used as the gold standard for comparison. Accuracy and reliability were assessed by analysis of variance followed by a mean separation procedure.

Results: The t-test revealed with statistical significance that in measuring the liver, students outperformed sonographers regardless of 2D or VOCAL™ methods; also, VOCAL™ generated more accurate measurements than 2D ultrasound in both groups. In regards to the gestational sac, students continued to perform well with both 2D and VOCAL™ while the sonographers were accurate by using VOCAL™ only. Both groups of novices were not accurate with measuring the breast lesion. VOCAL™ increased the novices' interobserver reliability but not intraobserver reliability.

Conclusion: Students and sonographers can measure regularly shaped structures, such as the gestational sac accurately and reliably using VOCAL™. VOCAL™ increased interobserver reliability in all groups of operators.

Keywords: 3D ultrasound, 2D ultrasound, VOCAL™, Novice.

How to cite this article: Nguyen TQ, Flores M, Ortiz M, Lee S, Mulla ZD, Plavsic SK. Accuracy and Reliability of Volume Measurements by 3D Ultrasound: A Prospective Study Performed by Novices. *Donald School J Ultrasound Obstet Gynecol* 2012;6(3):219-224.

Source of support: Nil

Conflict of interest: None

INTRODUCTION

The use of ultrasound to measure the size of organs and bodily structures has been in clinical practice for more than 30 years. More recently three-dimensional (3D) ultrasound measurements have been proven feasible and reproducible by a variety of studies in obstetrics and gynecology, neonatology and cardiology.¹⁻³ Compared to two-dimensional (2D), 3D ultrasound results in smaller inter-observer and intraobserver variability.^{4,5}

3D ultrasound examination has four main steps: Data acquisition, 3D visualization, volume/image processing and storage of the volume or image data. Once the organ volume

is stored, the sonographer can navigate through it in all three planes with tomographic precision. Digital volume storage allows retrospective analysis of the volumes and independent review by a second examiner. The 3D ultrasound examination may be broken up into smaller tasks, and simple ones may be done by less experienced operators.

Some studies focused on exploring ways to utilize novices in performing various parts of the 3D ultrasound examination. A recent study, comparing ovarian follicle counts between an ultrasound novice (an obstetrics-gynecology resident) and an experienced sonographer found that the novice's count using 3D ultrasound is similar to that of the sonographer's.⁶ The computer-based volume measurement of stored 3D data sets is a potential area where novices can be of use. Since the data acquisition process is not involved, these novices do not need to be trained to operate ultrasound machines. It has been shown that complete novices without extensive training were able to perform simple ultrasound tasks with excellent results. In another study, a medical student with no knowledge of ultrasound, after attending an online course on nuchal translucency measurement, was able to measure nuchal translucency thickness with an intraoperator reliability approximating the expert's level.⁷

The objective of this study is to evaluate the accuracy and reliability of offline volume measurement by novices using 2D method and 3D ultrasound assessment by the virtual organ computer-aided analysis (VOCAL™, GE Healthcare, Kretztechnik, Zipf, Austria).

METHODS

The protocol was approved by the Institutional Ethics Committee, and all volunteers gave written informed consent. A total of 48 volunteers participated in this prospective study. Thirty volunteers were medical students from the first and second years who have never been trained in ultrasound imaging. Another 15 volunteers were sonographers proficient in working with 2D ultrasound, but not familiar with 3D ultrasound and/or VOCAL™ analysis. All sonographers are currently employed by our institution and have completed at least 2 years of formal sonography education. In this article, both students and sonographers were considered novices. While the medical students were considered complete novices, sonographers were considered

novices in 3D ultrasound only. The last three volunteers were well-known radiologists from different institutions with over 20 years of experience in 3D ultrasound.

PROCEDURE

All three groups of volunteers (students, sonographers and experts) were asked to perform the same procedure done on a laptop computer. Three structures were selected for measurements by all participants: A liver surrounded by ascitic fluid, a twin gestational sac and a cancer lesion of the breast. All three structures were deidentified data sets obtained for research with permission from the patients. The 4D View software program (version 9.0 by GE Healthcare) was used for all the measurements. The 4D View program displays the offline ultrasound image in longitudinal, transverse and coronal planes simultaneously (Fig. 1).

The procedure included two parts: 2D volume measurement using formula for an ellipsoid, and 3D volume assessment by VOCAL™. In measuring the three structures of interests by 2D method, participants were instructed to obtain two largest distances on the first still image on the longitudinal plane and one largest distance on the transverse plane. Essentially, all participants measured three largest distances corresponded to the length, width, and depth of the structures of interest on still images. Each measurement was repeated three times to assess for individual variation. From the measurements of distances, the structure's volume was determined by applying the ellipsoid formula ($\text{length} \times \text{width} \times \text{depth} \times 0.524$). Since, there were a total of three repeats, three 2D volume values were obtained for each ultrasound structure by each participant.

The second part of the procedure involved 3D measurements done by the VOCAL™ method. The students and sonographers were given a Bamboo Pen (made by

Wacom, Japan) which is essentially a pen-like mouse accompanied by a 9×15 cm pressure-sensitive mouse pad. The experts used a regular mouse instead of the Bamboo Pen for reasons we will address later.

The students and sonographers were allowed 5 minutes to become familiar with using 4D View software and the Bamboo Pen. They were allowed another 45 minutes to perform the measurements of three structures as described above. The experts, however, only required about 20 to 25 minutes to complete the same tasks.

In obtaining VOCAL™ volume measurements, the contour of the structure was identified and delineated by the volunteer (Fig. 1). Twelve tracings were performed to obtain one calculation of the structure's volume. Each tracing was done after rotating the structure 15° around a fixed axis. From 12 tracings of the structure's border, the 4D View program produced a 3D image of the structure and subsequently determined its volume. This process was also repeated three times to assess intraobserver variation.

In summary, for each of the three structures, one participant produced three volumes calculated by the 2D method and three volumes measured by VOCAL™ (or the 3D method). The measurements obtained by the expert using VOCAL™ (3D) method were used as the gold standard for the evaluation of accuracy of novices' measurements. Within each structure, the mean volumes for the three groups were compared by doing an analysis of variance (ANOVA) using a completely random design. A least significant difference (LSD) t-test was then performed, if the ANOVA procedure revealed significant interactions among the participant groups and the ultrasound method used. The t-test allowed for evaluation of the accuracy of measurements done by the novice groups compared to the experts. The t-test compared the means of each subgroup (students using 2D, students using 3D, sonographers using 2D and sonographers using 3D ultrasound) with the mean of the expert group using 3D. The smaller the difference, the more accurate each group was. A confidence interval that included zero indicated that no statistically significant difference was found between the novice mean and the expert mean.

Intraobserver and interobserver variations were assessed by the mean square error (MSE) and standard deviation (SD) also obtained from the ANOVA. The SAS software package (SAS Institute, Inc., Cary, North Carolina) was used to carry out all ANOVA and LSD analyses.

Additionally, data of the volunteers' age, gender and self-assessed level of computer proficiency were collected. Level of computer proficiency was divided into 4 categories: low, average, above average and computer expert.

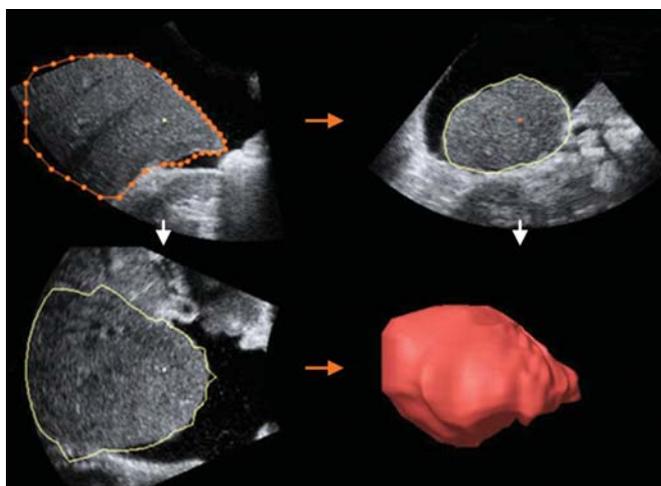
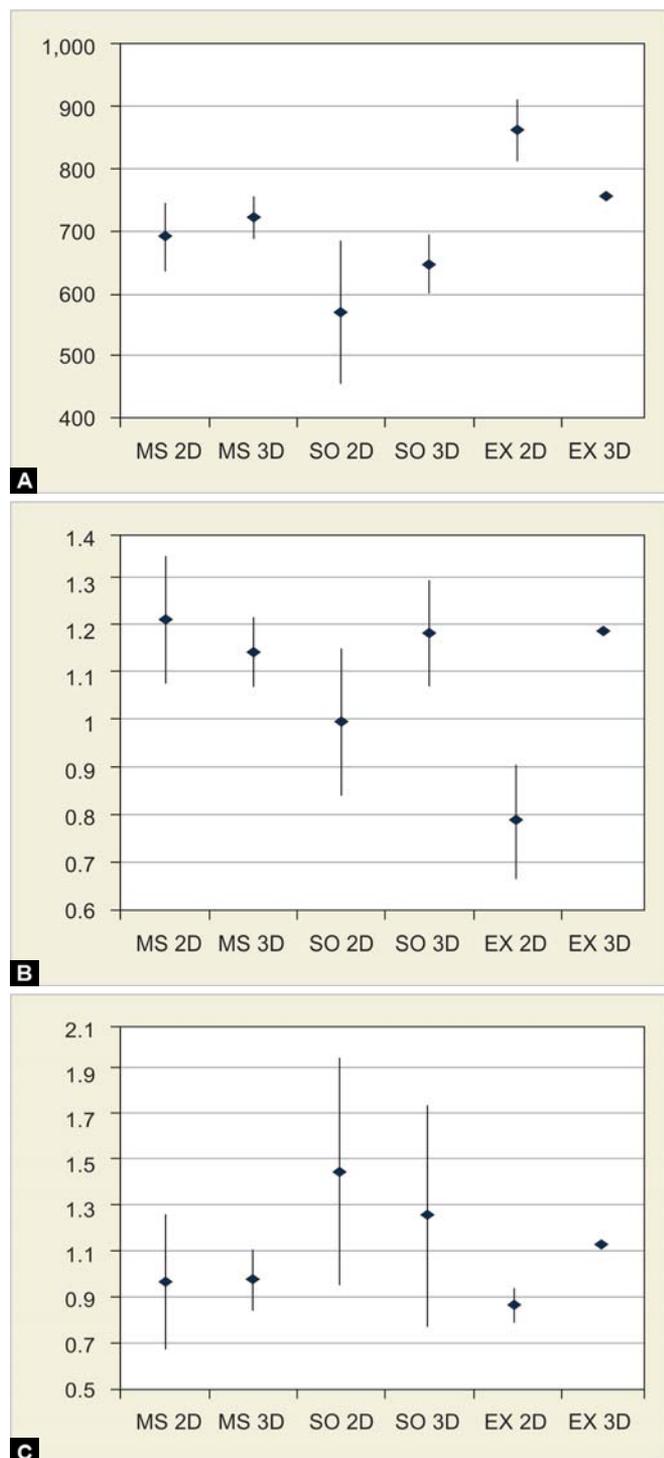


Fig. 1: Image of the liver in transverse, longitudinal and coronal planes as well as the constructed 3D image as seen in the 4D View software

RESULTS

Accuracy

Figures 2A to C provided a general visual of how the means and SDs of individual groups compared with the experts’



Figs 2A to C: Mean and SD of volume measurements of (A) Liver, (B) gestational sac, (C) breast cancer lesion. Students outperformed sonographers in measuring the liver. Sonographers were not accurate in measuring the gestational sac using 2D method. Both novice groups were not accurate in measuring the breast cancer lesion. MS: Medical students; SO: Sonographers; EX: Experts; 2D: Two-dimensional method; 3D: Three-dimensional or VOCAL™ method

Table 1: Comparisons of means of novice groups with the mean of expert group using 3D ultrasound. ANOVA and t-test revealed the differences in means with 95% confidence intervals included in parentheses. Confidence intervals which include 0 indicate no significant differences between the means

Liver	MS 2D	63.5 (-10.5; 137.5)
	MS 3D	33.6 (- 40.4; 107.6)
	SO 2D	185.1 (107.8; 262.3) *
	SO 3D	109.2 (31.9; 186) *
Gestational sac	MS 2D	-0.024 (-0.165; 0.118)
	MS 3D	0.045 (-0.097; 0.187)
	SO 2D	0.194 (0.046; 0.342) *
	SO 3D	0.006 (- 0.142; 0.153)

Difference between means significant at 0.05 level are indicated by* MS: Medical students; SO: Sonographers

(note, the experts’ SDs were so small that they were not visible on the graphs). In measuring the liver, the students outperformed the sonographers regardless of the 2D or 3D methods; also, the 3D method proved to be more accurate than the 2D method in both groups (Table 1). The ANOVA revealed that for both the liver and gestational sac, significant interactions between the group of observer and the method of measurement used were found; however, there is no such interaction found in measurements of the breast lesion. Since, these interactions reflect the presence of an influence of either the observer group or the method of measurement on the measured volumes of the liver and gestational sac, we could use a t-test to evaluate the volumes of the liver and gestational sac (Table 1).

In measuring the gestational sac, the sonographers were more accurate than the students only by using VOCAL™. The sonographers’ 2D measurements were significantly different from the gold standard, while the students continued to perform well with both 2D and 3D. While 3D method improved the accuracy of the measurements of the sonographers tremendously, it did not effect the same changes on the students.

Interobserver and Intraobserver Reliability

The ANOVA provided the SDs of volume measurements within each group, which were used to assess the interobserver reliability. The student group showed a smaller degree of interobserver variability than the sonographer group regardless of the structure analyzed (Table 2). Furthermore, based on the ratio between SDs of 2D vs 3D, VOCAL™ method increased interobserver reliability. The SDs of volumes were 1.6 to 2.2 times greater in 2D than they were in 3D for the medical students group. In the sonographer group, the SDs of volumes were 1.0 to 2.4 times greater in 2D than in 3D (Table 2).

The MSEs obtained were calculated from the three repeats of each measurement. The smaller MSE was the

Table 2: Interobserver reliability calculated as SDs of volume measurement by medical students and sonographers

SDs		Medical students	Sonographers
2D	Liver	54.1	115.2
	GS	0.135	0.157
	Br	0.296	0.504
3D	Liver	34	47.2
	GS	0.075	0.114
	Br	0.136	0.490
2D/3D ratio	Liver	1.6	2.4
	GS	1.8	1.4
	Br	2.2	1.0

GS: Gestational sac; Br: Breast lesion

Table 3: Intraobserver reliability calculated as MSEs of the volume measurement by medical students and sonographers

MSE		Medical students	Sonographers
2D	Liver	907	1128
	GS	0.0053	0.0029
	Br	0.0092	0.0261
3D	Liver	1471	1543
	GS	0.0016	0.0031
	Br	0.0102	0.0365
2D/3D Ratio	Liver	0.6	0.7
	GS	3.3	0.9
	Br	0.9	0.7

GS: Gestational sac; Br: Breast lesion

greater the intraobserver reliability. Except for the gestational sac measured by the medical students, all measurements showed higher degree of intraobserver reliability with 2D than with 3D (Table 3).

Finally, no association of age, gender or level of self-assessed computer skills and performance of the volume measurement were found.

DISCUSSION

The measurements obtained by the experts using VOCAL™ (or 3D) method were used as the gold standard for all comparisons as VOCAL™ is currently considered a highly valid and reliable method of measuring volumes.⁸ A recent study found excellent intraobserver and interobserver reliability for fetal cardiac volume assessed by extended imaging VOCAL™.⁹ Within the expert groups, the interobserver variations by VOCAL™ method are much smaller than those of the 2D method (the coefficient of variation for each of the structure measured with VOCAL™ is 16, 2 and 1% of the corresponding coefficient by the 2D method). A markedly high level of agreement was seen within the experts using VOCAL™. Therefore, only the averages of the experts' 3D measurements were used as the gold standard for assessing novices' validity.

VOCAL™ allowed both the medical students and the sonographers to produce more accurate volume measurements than they would with the 2D method. While VOCAL™ increased both the experts' inter- and intraobserver reliability, it only increased the novices' interobserver reliability and failed to improve the intraobserver reliability.

The experts have had extensive practice and experience maneuvering the regular mouse to follow the contours of a variety of sonographic images. Manipulating the mouse is a fine movement that requires a dexterity that is difficult to achieve after a few minutes of practice. Therefore, to improve the speed and precision of VOCAL™ margin-tracing, our novice participants were given the Bamboo Pen instead of the conventional mouse. Our volunteers who have never done a VOCAL™ measurement or used a Bamboo Pen could complete a structure with 12 tracings (15° rotation) in less than 5 minutes; which was a decent speed compared to a non-novice sonographers in another study who took about 1 minute to perform six tracings (30° rotation).⁸ Our study, did not evaluate or compare the utility of the Bamboo Pen versus the conventional mouse and how it could affect the performance of the novices; consequently, the results we collected from the students and sonographers cannot be extrapolated to situations where a regular mouse was used. However, the Bamboo Pen merely served as a mouse; therefore, it should not affect the novice's ability to visualize and distinguish the border of various sonographic structures.

The use of the Bamboo Pen may create the issue of inconsistency since, the experts did not use the Pen to do their VOCAL™ measurements. We feel that the aim of this study is to focus on the novices and their measurements; the experts here only served as the current clinical gold standards against which comparisons were made. We did not want to modify this current standard with introduction of a new technology. A potential future research project is to determine whether the Bamboo Pen would increase the experts' speed and accuracy.

Currently, there exists a hypothesis that 3D ultrasound is superior to 2D ultrasound following the advent of 3D ultrasound and a surge of research evaluating its accuracy and reliability. Riccabona et al analyzed the accuracy of 2D and 3D ultrasound in distance and volume measurements in a series of *in vitro* experiments comparing the ultrasound assessment with real objects' volumes.^{10,11} They showed that 3D ultrasound produced more accurate volume measurements compared with 2D method, especially for irregular-shaped objects. Since, 2D ultrasound utilizes the ellipsoid formula (length × width × depth × 0.524) to obtain

approximation of the volume rather than measuring the actual volume itself, estimation of irregular-shaped objects would show more deviation from the true measurement. The novices' data demonstrated that 3D provided more accurate volume measurements than 2D in regular shaped structures, such as the liver and the gestational sac (Table 1). However, accuracy of measurements of the breast cancer lesion, a highly irregular structure, was generally low in both novice groups.

While VOCAL™ consistently reduced interobserver variation for volumes of all three structures in both student and sonographer groups, it failed to improve intraobserver variation in 3D. Little variability in 2D is seen potentially because all three attempts of measuring the distances were done consecutively on the same still images. Most operators repeated the same caliber placements resulting in similar distances. A greater degree of intraobserver variation might have been observed had the participants were allowed to measure the distances on different still images.

Students vs Sonographers using VOCAL™

Novices have been able to perform certain tasks, such as operating ultrasound guidance for embryo transfer along with an obstetrics and gynecology physician or producing images of the abdominal aorta and kidney in the correct plane with remote assistance.^{12,13} However, experience is still immensely important in sophisticated sonographic tasks. For instance, trainees in obstetrics and gynecology with at least 6 months of gynecologic ultrasound performed significantly poorer than experts in recognizing patterns of malignant adnexal masses in static 2D ultrasound images.¹⁴ Similarly, physicians with experience in abdominal ultrasound but not familiar with contrast-enhanced ultrasound were less able to differentiate between benign and malignant liver tumors by interpreting tumor enhancement pattern following contrast injection.¹⁵

Unlike the liver and gestational sac, the breast cancer lesion displays tissue retraction which causes the margin to be jagged and indistinct (Fig. 3). This posed a big challenge for our novices. With no training or instruction on how to differentiate between healthy and cancerous region, the students were expected to perform poorly. The sonographers with undoubtedly more exposure and experience in recognizing cancerous lesions, however, did not perform better than the students (Figs 2A to C). The sonographers displayed an interobserver variation that is more than 3 times larger than that of the student group (0.490 vs 0.136). Similarly, the sonographers showed less interobserver reliability in measuring the other two structures. In contrast to the students who do not have any ultrasound knowledge,

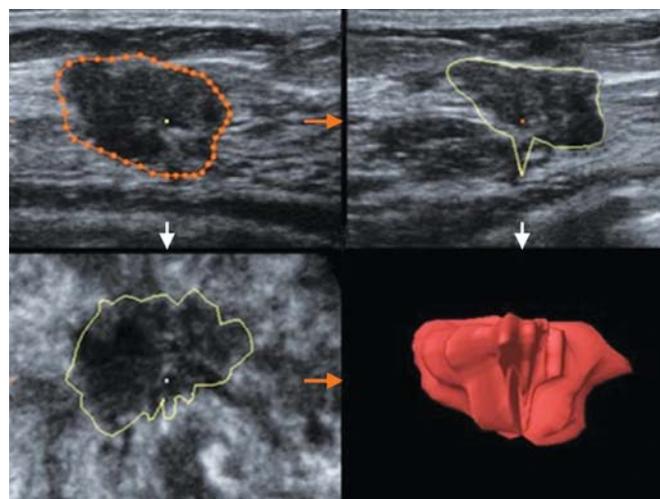


Fig. 3: The breast lesion in transverse, longitudinal and coronal planes, with its constructed 3D image

the wide differences in backgrounds and practices of the sonographers may be responsible for their large SDs (Table 2). While measuring the distances by 2D method, many sonographers failed to select the longest distance and/or neglected to ensure that the length was perpendicular to the width. The students, on the other hand, were exposed to only one set of instructions which they all followed faithfully. A high degree of variation among the sonographers was observed during VOCAL™ tracing of the breast cancer lesion, which signified not all sonographers know to recognize the correct border of healthy/cancerous tissue. Most students though primarily were guessing the border of the breast lesion, they all seemed to follow the same visual cues in tracing the lesion, which may explain why they were more reliable within their group. Motivation is another possible explanation for the increased interobserver variation in the sonographers. They were observed to be less motivated than the students in attempting to obtain the most precise tracing.

This study does not recommend that novices, such as students and sonographers should be allowed to make offline 3D volume measurements without supervision. The stored 3D data would permit reevaluation of the novice's measurements by a 3D ultrasound expert to maintain quality assurance. Our initial data suggested that with some additional and focused training, novices have the potential to become competent in measuring a limited number of simple structures using VOCAL™.

Although we had a good sample size of novices, a limitation of the study is the small number of structures being evaluated. The use of the Bamboo Pen and the regular mouse was not being evaluated in this study, and therefore, may be responsible for some of the success in the novices.

CONCLUSION

The VOCAL™ method performed by novices (medical students and sonographers) allowed accurate and reliable volume measurements of simple, regular-shaped structure, such as the gestational sac. Measurements of complex structures with irregular borders due to tissue retraction, such as breast cancer resulted in high interobserver variation for both groups of novices.

ACKNOWLEDGMENTS

The authors are grateful to Pooh R, MD; Kurjak A, MD, PhD and Milan Stanojevic, MD, PhD, the experts in 3D ultrasound who performed volume measurements.

REFERENCES

1. Alcázar JL, Jurado M. Three-dimensional ultrasound for assessing women with gynecological cancer: A systematic review. *Gynecol Oncol* 2011;120:340-46.
2. Huster KM, Haas K, Schoenborn J, et al. Reproducibility of placental volume and vasculature indices obtained by 3-dimensional power Doppler sonography. *J Ultrasound Med* 2010;29:911-16.
3. Chiu B, Egger M, Spence JD, et al. Quantification of carotid vessel wall and plaque thickness change using 3D ultrasound images. *Med Phys* 2008;35:3691-3710.
4. De Jong-Pleij EA, Ribbert LS, Tromp E, et al. Three-dimensional multiplanar ultrasound is a valuable tool in the study of the fetal profile in the second trimester of pregnancy. *Ultrasound Obstet Gynecol* 2010;35:195-200.
5. Yaman C, Ebner T, Jesacher K, et al. Reproducibility of three-dimensional ultrasound endometrial volume measurements in patients with postmenopausal bleeding. *Ultrasound Obstet Gynecol* 2002;19:282-86.
6. Moawad NS, Gibbons H, Liu J, et al. Comparison of 3- and 2-dimensional sonographic techniques for counting ovarian follicles. *J Ultrasound Med* 2009;28:1281-88.
7. Abele H, Hoopmann M, Wright D, et al. Intra- and interoperator reliability of manual and semi-automated measurement of fetal nuchal translucency by sonographers with different levels of experience. *Ultrasound Obstet Gynecol* 2010;36:417-22.
8. Kusanovic JP, Nien JK, Gonçalves LF, et al. The use of inversion mode and 3D manual segmentation in volume measurement of fetal fluid-filled structures: Comparison with virtual organ computer-aided analysis (VOCAL). *Ultrasound Obstet Gynecol* 2008;31:177-86.
9. Barreto EQ, Milani HF, Araujo Jr E, et al. Reproducibility of fetal heart volume by 3D-sonography using the XI VOCAL method. *Cardiovasc Ultrasound* [serial online]. 2010 May;8:17. Available from cardiovascularultrasound.com. Accessed January 10, 2011.
10. Riccabona M, Nelson TR, Pretorius DH, et al. Distance and volume measurement using three-dimensional ultrasonography. *J Ultrasound Med* 1995;14:881-86.
11. Riccabona M, Nelson TR, Pretorius DH. Three-dimensional ultrasound: accuracy of distance and volume measurements. *Ultrasound Obstet Gynecol* 1996;7:429-34.
12. Harris ID, Styer AK, Petrozza JC. Ultrasonographer experience does not impact outcomes following ultrasound-guided embryo transfer. *Fertil Steril* 2009;92:918-22.
13. Sheehan FH, Ricci MA, Murtagh C, et al. Expert visual guidance of ultrasound for telemedicine. *J Telemed Telecare* 2010;16:77-82.
14. Van Holsbeke C, Daemen A, Yazbek J, et al. Ultrasound methods to distinguish between malignant and benign adnexal masses in the hands of examiners with different levels of experience. *Ultrasound Obstet Gynecol* 2009;34:454-61.
15. Quaiá E, Alaimo V, Baratella E, et al. Effect of observer experience in the differentiation between benign and malignant liver tumors after ultrasound contrast agent injection. *J Ultrasound Med* 2010;29:25-36.

ABOUT THE AUTHORS

Tho Quynh Nguyen

Student Physician, Paul L Foster School of Medicine, Texas Tech University Health Sciences Center, El Paso, Texas, USA

Melanie Flores

Radiology Resident, Department of Radiology, Paul L Foster School of Medicine, Texas Tech University Health Sciences Center, El Paso Texas, USA

Melchor Ortiz

Division of Biostatistics and Epidemiology, Texas Tech University Health Sciences Center, El Paso, Texas, USA

Soyoung Lee

Division of Biostatistics and Epidemiology, Texas Tech University Health Sciences Center, El Paso, Texas, USA

Zuber D Mulla

Associate Professor, Director of Epidemiologic Research, Department of Obstetrics and Gynecology, Texas Tech University Health Sciences Center, El Paso, Texas, USA

Sanja Kupescic Plavsic (Corresponding Author)

Professor, Department of Obstetrics and Gynecology, Department of Radiology, Texas Tech University Health Sciences Center, 4800 Alberta Avenue, El Paso, Texas 79905, USA, e-mail: sanja.kupescic@ttuhsc.edu