SCIENCEFORECAST
Connecting Research to Drive Scientific Innovation

Case Series
Published: 11 Aug, 2021

SF Journal of Gynecology and Obstetrics

Comparison of Follicular Assessment Using a Manual Versus Automated 2D Ultrasound Method

*Roopa M**

Consultant Obstetrician and Gynaecologist, Shanthi Nagar, Bangalore, Karnataka, India

Abstract

Background: High intra- and inter-observer variability in the follicular assessment using 2D ultrasound (US) is still a concern. To solve this issue, a software solution, which automatically provides follicles' count and their diameters using 2D US images obtained by a manual sweep of an ovary. The primary objective of this study was to compare result of the automated solution with manual 2D US-based assessment.

Methods: In the first phase, multiple follicular US sweeps were collected from 54 subjects; these sweeps were used. In the second phase, data from 10 subjects were collected for validation of the developed solution. During each phase, for follicles>=5mm their count and diameters were recorded using 2D US.

Results: For the total follicle count, an excellent correlation (0.787) was observed between the solution and manual assessment. The 95 % limits of agreement between the two methods was in the range of 4.232 to -4.258. The two methods had an excellent correlation (0.817) for measurement of mean follicular diameter as well. However, the solution had tendency to underestimate mean diameter by an average of 1.725 mm (+2.16 mm). The limits of agreement between the two methods for mean diameter measurement were from 2.508 mm to -5.960 mm.

Conclusion: This study validates feasibility our solution for automatic assessment of follicle count and diameter with accuracy comparable to the 2D US-based manual assessment. We further observed that the solution's performance is better than known intra- and inter-observer variability of the manual assessment. We recommend further validation of the solution to confirm these initial results and potential time gain with automated assessment.

Keywords: Ultrasonography; Ovarian follicles; Automation; Computer-assisted image processing; Infertility; Assisted reproductive techniques

OPEN ACCESS

*Correspondence:
Roopa M, Consultant Obstetrician and Gynaecologist, Shanthi Nagar, Bangalore, Karnataka, India.
E-mail: dr.nageshmruthyunjaya@gmail.com

Received Date: 25 Jun 2021
Accepted Date: 06 Aug 2021
Published Date: 11 Aug 2021

Citation: Roopa M. Comparison of Follicular Assessment Using a Manual Versus Automated 2D Ultrasound Method. SF J Gynecol Obstet. 2021; 4(1): 1013.

Copyright © 2021 Roopa M. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Infertility is estimated to touch around 16% of couples at some point of their lives. The age-standardized prevalence rate of female infertility has shown a significant increase from 1366.85 per 100,000 in 1990 to 1571.35 per 100,000 in 2017,which comes to 0.37% increase per year. In the similar time period the absolute number of couples affected by infertility have grown up from 42.0 million in 1990 to 118 million in 2017, due to population growth [1,2]. Infertility treatment generally involves an ovarian stimulation, where under the influences of drugs, multiple follicles are recruited simultaneously; however, these follicles grow at different rates. Therefore, Assisted Reproductive Technology (ART) based methods require a regular and careful monitoring of follicles. The total number of ovarian follicles (antral follicle count) and their dimensions are two important parameters, which are closely monitored during ovarian stimulation procedures.

Ultrasound (US) imaging is the most preferred method for the monitoring of ovulation stimulation. Serial US scans are done during the course of ovulation induction to track ovarian follicle growth. Conventionally, this is done using two-dimensional (2D) US, where a clinician manually counts and measures follicles' dimensions. However, there is a lack of consensus on standard protocols for measurement of follicular diameter [3-6], this along with subjectivity in assessment is responsible for high intra- and inter-observer variability observed in 2D US-based follicular assessment. This has led to development of three-dimensional (3D) US-based software solutions, which provide an automated assessment of different follicular parameters.

The 3D US-based software solutions have shown to significantly reduce intra- and inter-observer variability in follicular assessment, along with a significant reduction in time required for assessment

[7]. Although useful, US devices with 3D trans-vaginal probe and automated software are either not available or have high cost, which makes such solutions unfeasible for the resource constrained countries; unfortunately, these are countries, which have the majority of the infertile couples. Moreover, no significant difference has been observed in the success rate of assisted reproduction treatment when a 3D method was used instead of 2D method [8]. Hence, a manual 2D US-based assessment of ovarian follicles still remains the method of choice worldwide. This makes it important to have intuitive solutions to help clinicians perform a better follicular assessment using conventional 2D method and hardware.

The primary objective of this study is to compare automated solution with manual 2D US-based assessment for measurement of follicle count and diameter on the follicles larger than 5 mm in diameter.

Materials and Methods

This prospective observational study was split in two phases. In the first phase, 60 subjects were recruited. The inclusion criteria were women aged 18 years or above who had been advised for infertility related pelvic ultrasound scan (infertility screening or assisted reproduction treatment). All the subjects were treated by the established protocols. For a given participant, after ultrasound-based assessments for follicular monitoring, one to five 2D US sweep recordings of both the ovaries were obtained and stored in a digital format. For the study, US scan from the sixth day (post-stimulation) onwards were used. The multiple follicular assessment US sweeps obtained from these subjects were used for automatic assessment of follicles' number and their sizes. In the second phase, 10 subjects were recruited from the same. The 2D US sweep data from these subjects were used for a blind validation of the developed automated solution.

Manual 2D US-based follicular assessment

For all the participant, for each ovary, total number of ovarian follicles and their sizes were provided by the clinicians/sonologists using conventional 2D US-based method. For this assessment, only follicles larger than 5 mm were considered. For each follicle, the plane where it looks the biggest and roundest is searched for. The two biggest diameters were then measured using manual calipers. The mean of these two diameters is computed and recorded. Philips ClearVue 550 system with C9-4v probe was used for this assessment. The manual 2D US-based assessment was considered as a ground truth for algorithm development and subsequent comparison. All scans were assessed for image quality before inclusion.

Data acquisition for algorithm by 2D US sweep

A sweep of each ovary was performed in two systematic ways: 1) from the lateral end to the medial end of the ovary (LM sweep) or indifferently the opposite (medially to laterally) or 2) from the anterior side of the ovary to the posterior side of the ovary, or the opposite (AP or PA sweep). All the sweeps included a margin safety, i.e., a few images going beyond the ovary at the beginning and at the end of the sweep to be sure that the set of images contained the whole ovary. For each ovary, 4 to 5 sweeps were collected at each US scan, including 2 to 3 using the LM sweep and 2 to 3 the AP sweep. The series of images obtained were recorded using the cine loop mode of a Philips ClearVue (650 and 850) system using a C9-4v trans-vaginal probe (9-4 MHz). The cine loop time for each sweep was fixed at 10 seconds.

Assessment of stored 2D US sweeps by independent experts

Two independent experts who were blind to result of manual 2D US-based assessment and clinical history were asked to review the entire recoded US sweep and assess follicle number and diameters. An annotation tool was used for this purpose, which allowed experts to review each US sweep frame by frame for assessment.

Algorithm for automated follicular assessment

The proposed algorithm is based on region-growing approach for image segmentation. The algorithm segments ovarian follicles based on their unique geometrical and statistical properties. The acquired 2D US sweep images are first subjected to pre-processing, which comprises of two steps: contrast enhancement and de-noising. The contrast enhancement is used to increase the contrast between follicular and non-follicular regions. This is followed by image normalization to enhance the contrast between the follicular regions and to highlight boundaries between them. However, the contrast enhancement also amplifies noise and hence the intensity normalization step is followed by a de-noising procedure to mitigate the effects of noise.

The pre-processed images are then subjected to iterative region growing method. Region growing routines are a class of seed based image segmentation algorithms where pixels in a neighborhood are successively added to the current segment till a specific image intensity convergence criterion is met. The iterations involve computation of the shape of the follicle by imposing a constraint on the shape of the regions and the stability of the shape over iterations. A shape constraint is used on the segmented regions to prevent over-segmentation of the follicles. The plot of the shape parameter over the iterations is analyzed to identify the optimal point to stop region growing. For the follicle segmented using this approach, the major and minor axis lengths are computed using the best fitted ellipse method. The average of measurement of these two axes is then considered as mean diameter of the follicle. Based on all the follicles identified by this approach the total follicle count is computed. Similarly, for the entire identified follicles their mean diameter is also provided.

Statistical analysis

Manual 2D US-based assessment done by a clinician was considered as a ground truth for all comparisons. The distribution of the data was first analyzed using Shapiro-Wilk test. Based on distribution, Student's t-test or Wilcoxon's signed-rank test was used to compare total number of follicles detected by the manual assessment and the automated solution in each US sweep. Only monitoring ultrasound scans were considered (basal scans excluded), in this method only follicles larger or equal to 5 mm in mean diameter were considered; whereas the automated solution was able to detect follicles smaller than 5 mm in diameter; therefore, for comparison of count, only those follicles, which were estimated to be larger than 5 mm in diameter by algorithm was used. The correlation between the two methods was determined using Pearson's or Spearman's rank correlation coefficient. The mean follicular diameter determined by the two methods was also compared similarly. The limits of agreement between the two method for follicle counts and diameters was assessed by the Bland-Altman method. The Bland-Altman method is considered as a gold standard for method comparison studies, and has been extensively used to compare different methods of follicular assessment [8]. The algorithm results were also compared with two independent experts' assessment using the same methodology. For all

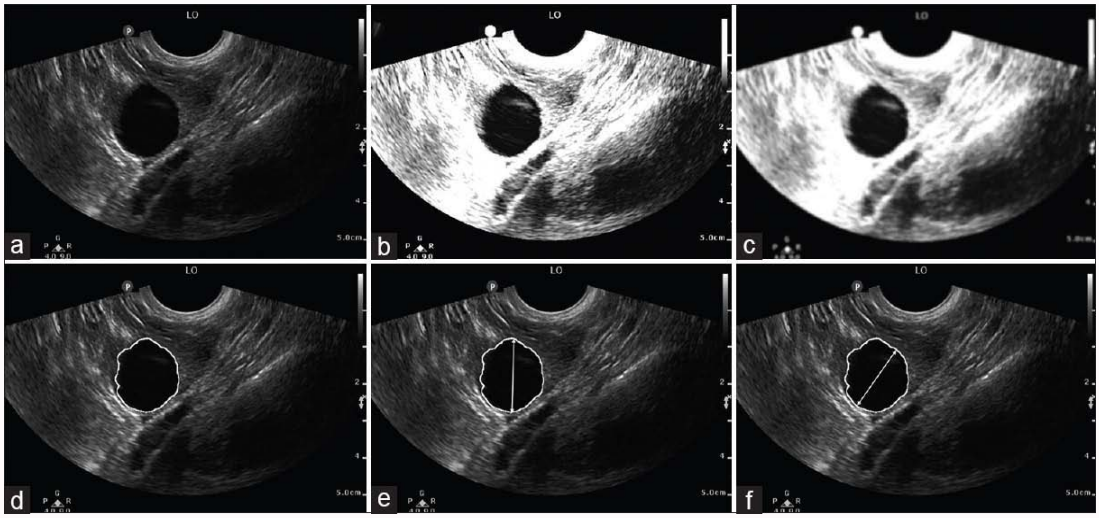

Figure 1: Image processing steps in the automated follicular assessment; (a) Original image; (b) Contrast-enhanced image; (c) De-noised image; (d) Outline of a segmented follicle; (e and f) Measurement of different follicular axis.

comparisons, a *P*- value<0.05 was considered to denote a statistically significant difference. All statistical analyses were performed in R (version 3.6.2) and MATLAB®.

Results

In total, 60 subjects were recruited in the phase one. Out of this, six subjects' data were rejected due bad image quality (incomplete sweeps of the ovaries or poor image quality). The data from remaining 54 subjects were used for algorithm development. Out of these, 29 participants were undergoing IUI (intra-uterine insemination) and 25 were undergoing IVF/ICSI (*in vitro* fertilization/fecundation with or without intracytoplasmic sperm injection). Total ten subjects were enrolled in the second phase. Five participants from this group were taking IUI treatment whereas the remaining five were undergoing IVF/ICSI treatment.

In the phase two, total 86 2D US scans were performed on participants for follicular monitoring over the course of their menstrual cycles. The clinicians recorded a total of 1484 follicles in these manual scans with a mean follicular diameter in the range from 5 mm to 20 mm. Total 251, US sweep recordings were obtained from theses assessments; out these, 17 recordings were discarded due to bad image quality; remaining 234 records (1411 follicles) were used for final analysis.

Comparison between algorithm and manual 2D US-based assessment

Manual assessment of each ovary was performed by a clinician in real time using 2D US scan; only follicles larger than 5 mm were considered. Each ovary was recorded to have an average 6.35 (+3.21) follicles (median=6; interquartile range=4 to 9). The average number of follicles (larger than 5 mm) detected by the algorithm was 6.33 (+3.80) follicles (median=6; interquartile range=4 to 8).

The mean follicular diameter was 10.74 (+3.64) mm (median=10.5 mm; interquartile range=8 to 13 mm). The mean follicular diameter estimated by the algorithm was 9.01 (+3.44) mm (median=8.31 mm; interquartile range=6.42 to 11.11 mm). Both follicle counts and diameters were not normally distributed.

No significant difference was observed in total follicle count between the algorithm and manual 2D assessment by Wilcoxon's signed-rank test. The two methods had an excellent correlation, with Spearman's rank correlation coefficient of 0.787. The 95 % limits of agreement between the two methods were 4.232 for the upper limit and -4.258 for the lower limit.

The two methods were found to have a statistically significant difference in measurement of mean follicular diameter, with the algorithm underestimating mean diameter by an average of -1.725 mm (+2.16 mm). However, the two methods had an excellent correlation for mean follicular diameter measurement (Spearman's coefficient=0.817). The Bland-Altman plots for the limits of agreement with 95% confidence intervals for the two methods are presented in Figure 1. The upper limit of agreement between the two methods was 2.508 mm, whereas the lower limit of agreement was -5.960 mm. All comparison-related results are summarized in Table 1 and Table 2, Figure 2.

Comparison between the algorithm and two independent experts

The two independent experts performed follicular assessments on the recorded 2D US sweeps. No significant difference was observed in total follicle count between the algorithm and Expert-1's assessment with an excellent correlation coefficient of 0.784. The algorithm also had an excellent correlation with the Expert-2 (0.765), but there was a significant difference (*P*=0.00014) in total follicular count with the Expert-2 being able to detect on an average 0.6 more follicles than the algorithm. In fact, it was further observed that the Expert-2 was able to detect significantly more follicles than manual 2D assessment (*P*<0.001) as well as the Expert-1 (*P*<0.001). The limits of agreements of the algorithm with two experts for follicular count are summarized in Table 1.

The algorithm had an excellent correlation with both the experts for the mean follicular diameter (Table 2). However, there was a significant difference in mean diameter measurement, with the algorithm having a tendency to underestimate the mean diameter in comparison to the experts; the difference was more prominent for the Expert-2 with mean difference of -2.92 mm (+2.22). It was observed that Expert-2 had a statistically significant difference in mean diameter measurement in comparison to 2D manual assessment and Expert-1, with Expert-2 having a general trend to overestimate the

Table 1: Comparison of algorithm'sresult for total follicular count with other methods.

Method to compare	Mean difference (SD)	Spearman's coefficient	Upper LA	Lower LA	Range between LA
Algorithm *Vs* 2D US-based assessment	-0.012 (±2.16)	0.787	4.232	-4.258	8.490
Algorithm *Vs* Expert-1 (US sweep assessment)	-0.196 (± 2.16)	0.784	4.036	-4.429	8.466
Algorithm *Vs* Expert-2 (US sweep assessment)	-0.632 (± 2.4)	0.765	4.106	-5.371	9.478

2D=Two dimensional; LA=95% limits of agreement by the Bland-Altman method; SD=Standard Deviation; US=Ultrasound.

Table 2: Comparison of mean follicular diameter estimated by algorithm with other methods.

Method to compare	Mean difference (SD)	Spearman's coefficient	Upper LA	Lower LA	Range between LA
Algorithm *Vs* 2D US-based assessment	-1.725 mm (± 2.16)	0.817	2.508	-5.960	8.468
Algorithm *Vs* Expert-1 (US sweep assessment)	-1.174 mm (± 2.05)	0.836	2.841	-5.190	8.032
Algorithm *Vs* Expert-2 (US sweep assessment)	-2.92 mm (± 2.22)	0.828	1.435	-7.275	8.711

2D=Two dimensional; LA=95% limits of agreement by the Bland-Altman method; SD=Standard Deviation; US=Ultrasound.

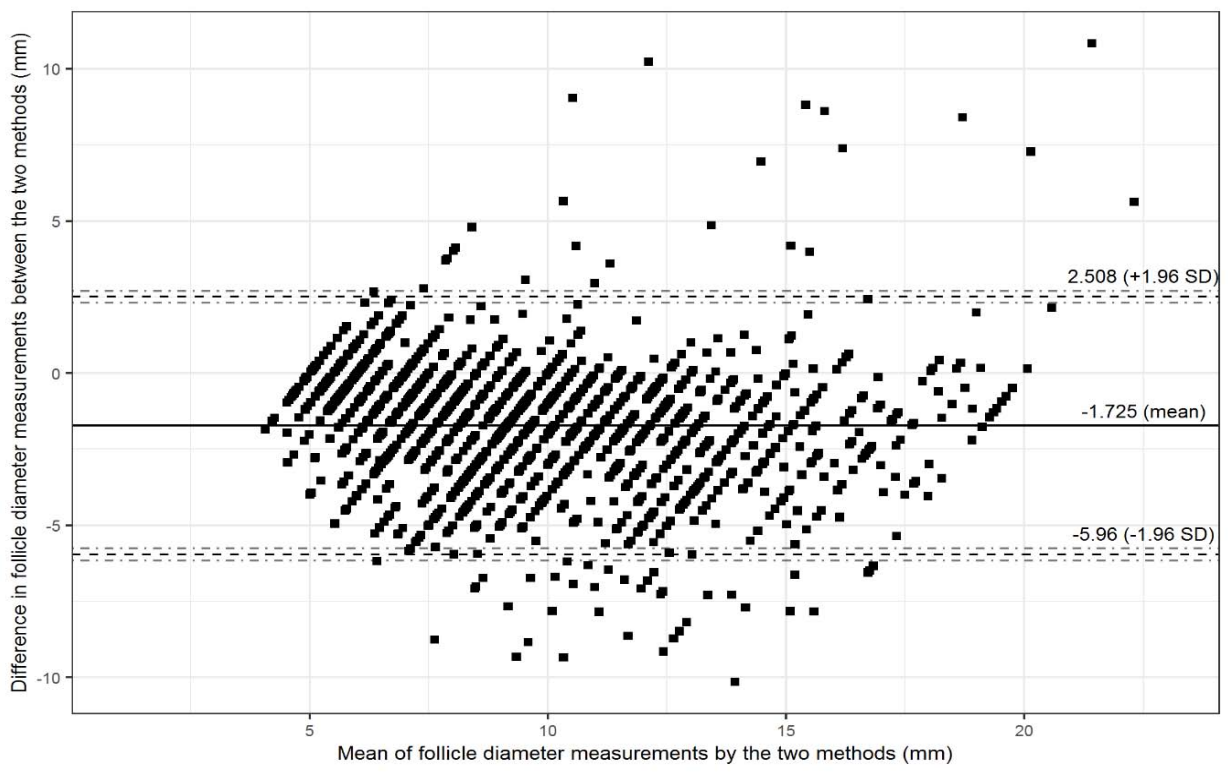

Figure 2: Bland-Altman plot of the limits of agreement between automated solution and 2D US-based manual assessment for measurement of follicle diameter.

mean follicular diameter.

Discussion

Infertility is a significant problem worldwide and factors such as delayed conception, pollution, environment and lifestyle changes are further like to make it complicated. The modern ART relies heavily on ultrasound-based monitoring for infertility treatment. The 2D US-based manual assessment is the most preferred method for follicular monitoring worldwide; although it is known to have high intra and inter-observer variability. We have developed a novel software solution to make conventional 2D US-based follicular assessment objective and fast. The purpose of this study was to present the validation results of our solution on a blind data set. We observed that it was feasible to use our software solution for automatic assessment of follicle count and measurement with accuracy comparable to the real time 2D US-based manual assessment [9].

For the total follicle count, an excellent correlation was observed between our software solution and 2D US-based manual assessment. Although not statistically significant, the algorithm had a tendency to underestimate total follicle count (-0.012) in comparison to the 2D manual assessment. During processing, the algorithm applies certain shape constraints to detect follicles. The same trend had been observed with 3D US-based automated solutions as well [10-12]. The limits of agreements observed with our algorithm are within the intra- and inter-observer limits of agreements reported for total follicle count by manual 2D US-based method in literature [10,13]. This provides an indication that our software solution is a reliable alternative to conventional 2D method with a better accuracy.

Our algorithm was found to have an excellent correlation with the 2D manual assessment for measurement of follicular diameter as well. However, the algorithm had a tendency to underestimate mean

follicular diameter in comparison to the 2D manual assessment. We postulate that two principle factors could be contributing to it; the first group is related to how the algorithm works; whereas, the second factor group is related to the way follicular diameter is measured in a conventional practice. For follicular detection we have used seed-based region growing image segmentation algorithms. This algorithm detects a follicle in an iterative process staring from a small hypoechogenic region as a seed and then growing its border in outward directions. To prevent an overestimation of a follicle's size, the iterative process is restricted within the follicle's border; this may lead to an underestimation in the follicle's diameter. The other algorithmic factor is related to the heterogeneous aspects of follicles: some follicles may contain echoic regions within their boundaries and these regions can't be detected automatically by the algorithm. Another factor is related to the limitation of measuring all follicles in a single ovarian sweep where each follicle might not be visible in the right plane, i.e., where it presents its biggest mean diameter. The other principle factor is related to the way follicular diameter is measured in a conventional practice. Measurement of mean follicular diameter using conventional 2D US-based method has been associated high intra- and inter-observer variability due to lack of a consensus on standard protocols [5]. The placement of measurement calipers on US image is also an important factor in high subjectivity in assessment of follicular diameter. We observed that during measurement of diameter, clinicians have a tendency to put the calipers slightly outside of the follicular borders; this is mostly done as a safety margin so as not to miss any follicular part. This might lead to a systematic over estimation in follicle size by the manual method. We observed that 3D US-based automated software solutions also have a similar tendency to underestimate mean follicular diameter in comparisons to manual 2D US-based methods [14,15]. We further observed that the limits of agreements between our algorithm and the manual 2D method are within the inter-observer limits of agreements reported for the manual 2D method for measurement of mean follicular diameter [14]. This supports our hypothesis and demonstrates reliability of our solution for measurement of mean follicular diameter.

Apart of from reduction in intra- and inter-observer variability another important advantage of automated software solution is a significant reduction in time required for follicular assessment [10-12,14,16]. In the present study, we did not measure the time required for the manual 2D US-based assessment. It has been reported in the literature that mean time required for such assessment ranges from 56.8 seconds to 9.6 minutes with median of 314.4 seconds [8]. For our solution the cineloop time (recording time) for each sweep was fixed at 10 seconds. The time taken by our algorithm was in the range of 30 to 60 seconds (based on number of follicles) for automatic assessment of follicular count and measurement in a US sweep. Considering this, we believe that our software solution can bring a significant time saving for follicular assessment. As a future work, we would like to confirm these initial results and the potential examination time gain in an integrated system (ultrasound device with the automation software solution).

A small sample size is an important limitation of our study. We have tried to compensate the small sample size by obtaining multiple US sweeps from each participant, which provided as total 231 sweeps with more than 1431 follicles of different size. The other limitation is regarding follicle size; in the present study, we tested our algorithm on follicles larger than 5 mm in diameter. We are also exploring possibility of incorporating post-processing options to allow clinicians to manually add missed follicles and correct measurements.

To conclude, this study validates the reliability and performance of our automated solution for follicle count and measurement using 2D US sweeps. We observed that our solution's performance is better than known intra- and inter-observer variability of the manual 2D US-based assessment. We believe that this solution could be very helpful in reducing measurement variability during follicular assessment and can make conventional 2D US-based monitoring more objective and much faster. We recommend further validation of these solutions with well-designed multicenter studies.

Conflicts of Interest Statement

Celine Firtion, Ganesan Ramachandran, Sindhu P. Nellur Prakash, Sujitkumar Hiwale, Pallavi Vajinepalli, Dr Indira, and Dr. Devika Gunasheela declare that they have no conflict of interest.

Acknowledgments

We would like to acknowledge Chandan Mishra and Devdatt Kawathekar from Philips Ultrasound Business for their support in the project. We would also like to thank the staff of Gunasheela hospitals and researchers from Philips Research who were involved in the early phases of the project.

References

1. Sun H, Gong TT, Jiang YT, Zhang S, Zhao YH, Wu QJ. Global, regional, and national prevalence and disability-adjusted life-years for infertility in 195 countries and territories, 1990–2017: results from a global burden of disease study, 2017. Aging. 2019; 11: 10952-10991.
2. Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA. National, Regional, and Global Trends in Infertility Prevalence Since 1990: A Systematic Analysis of 277 Health Surveys. PLOS Med. 2012; 9: e1001356.
3. Ata B, Seyhan A, Reinblatt SL, Shalom-Paz E, Krishnamurthy S, Tan SL. Comparison of automated and manual follicle monitoring in an unrestricted population of 100 women undergoing controlled ovarian stimulation for IVF. Hum Reprod. 2011; 26: 127-133.
4. Raine-Fenning N, Jayaprakasan K, Clewes J, Joergner I, Bonaki SD, Chamberlain S, et al. SonoAVC: a novel method of automatic volume calculation. Ultrasound Obstet Gynecol. 2008; 31: 691-696.
5. Duijkers IJM, Louwé LA, Braat DDM, Klipping C. One, two or three: how many directions are useful in transvaginal ultrasound measurement of ovarian follicles? Eur J Obstet Gynecol Reprod Biol. 2004; 117: 60-63.
6. Forman RG, Robinson J, Yudkin P, Egan D, Reynolds K, Barlow DH. What is the true follicular diameter: an assessment of the reproducibility of transvaginal ultrasound monitoring in stimulated cycles. Fertil Steril. 1991; 56: 989-992.
7. Murtinger M, Zech MH, Spitzer D, Zech NH. Outpatient Follicle Monitoring: A Plea for Standardization in Ultrasound Based Follicle Monitoring and Data Transfer. J Reprod Infertil. 2014; 15: 105-108.
8. Vandekerckhove F, Bracke V, De Sutter P. The Value of Automated Follicle Volume Measurements in IVF/ICSI. Front Surg. 2014; 1: 18.
9. Peres Fagundes PA, Chapon R, Olsen PR, Schuster AK, Mattia MMC, Cunha-Filho JS. Evaluation of three-dimensional SonoAVC ultrasound for antral follicle count in infertile women: its agreement with conventional two-dimensional ultrasound and serum levels of anti-Müllerian hormone. Reprod Biol Endocrinol. 2017; 15: 96.
10. Deb S, Jayaprakasan K, Campbell BK, Clewes JS, Johnson IR, Raine-Fenning NJ. Intraobserver and interobserver reliability of automated antral follicle counts made using three-dimensional ultrasound and SonoAVC. Ultrasound Obstet Gynecol. 2009; 33: 477-483.

11. Raine-Fenning N, Jayaprakasan K, Deb S, Clewes J, Joergner I, Dehghani Bonaki S, et al. Automated follicle tracking improves measurement reliability in patients undergoing ovarian stimulation. Reprod Biomed Online. 2009; 18: 658-663.

12. Deb S, Campbell BK, Clewes JS, Raine-Fenning NJ. Quantitative analysis of antral follicle number and size: a comparison of two-dimensional and automated three-dimensional ultrasound techniques. Ultrasound Obstet Gynecol. 2010; 35: 354-360.

13. Scheffer GJ, Broekmans FJM, Bancsi LF, Habbema JDF, Looman CWN, Velde ERT. Quantitative transvaginal two- and three-dimensional sonography of the ovaries: reproducibility of antral follicle counts. Ultrasound Obstet Gynecol. 2002; 20: 270-275.

14. Raine-Fenning N, Jayaprakasan K, Chamberlain S, Devlin L, Priddle H, Johnson I. Automated measurements of follicle diameter: a chance to standardize? Fertil Steril. 2009; 91: 1469-1472.

15. Pan P, Chen X, Li Y, Zhang Q, Zhao X, Bodombossou-Djobo MMA, et al. Comparison of manual and automated measurements of monodominant follicle diameter with different follicle size in infertile patients. PloS One. 2013; 8: e77095.

16. Deutch TD, Joergner I, Matson DO, Oehninger S, Bocca S, Hoenigmann D, et al. Automated assessment of ovarian follicles using a novel three-dimensional ultrasound software. Fertil Steril. 2009; 92: 1562-1568.