Diagnostic and Interventional Imaging (2018) 99, 709—716

ORIGINAL ARTICLE / *Obstetric imaging*

Artificial intelligence assistance for fetal head biometry: Assessment of automated measurement software

G. Ambroise Grandjean[a,*], G. Hossu[b], C. Bertholdt[c], P. Noble[c], O. Morel[b], G. Grangé[c]

[a] *Inserm, IADI, université de Lorraine, 54500 Vandœuvre-lès-Nancy, France*
[b] *Inserm, CIC, université de Lorraine, CHRU de Nancy, 54000 Nancy, France*
[c] *Department of obstetrics, Port-Royal, hôpital Cochin, Assistance Publique—Hôpitaux de Paris, 75014 Paris, France*

KEYWORDS
Fetal biometry;
Artificial intelligence (AI);
Automatic measurement software;
Agreement study;
Reproducibility

Abstract
Purpose: To evaluate the feasibility and reproducibility of artificial intelligence software (Smartplanes®) to automatically identify the transthalamic plane from 3D ultrasound volumes and to measure the biparietal diameter (BPD) and head circumference (HC) in fetus.
Material and methods: Thirty fetuses were evaluated at 17—30 weeks' gestation. For each fetus two three-dimensional (3D) volumes of the fetal head along with one conventional two-dimensional (2D) image of the transthalamic plane were prospectively acquired. The Smartplanes® software identified the transthalamic plane from the 3D volumes and performed BPD and HC measurements automatically (3D auto). Two experienced sonographers also measured BPD and HC from 2D images and from the 3D volumes. Measurements were compared using Bland-Altman plots. Interclass correlation coefficient (ICC) was used to evaluate intra- and interobserver reproducibility.
Results: For each series of measurements, intra- and interobserver reproducibility rates were high with ICC values > 0.98. The 95% confidence intervals between the BPD measurements were 2 mm (3D versus 2D) and 4 mm (3D auto versus 2D) and the HC measurements were 7.5 mm (3D versus 2D) and 11 mm (3D auto versus 2D).
Conclusion: Fetal head measurements obtained automatically by Smartplanes® software from 3D volumes show good agreement with those obtained by two experienced sonographers from conventional 2D images and 3D volumes. The reproducibility of these measurements is similar to that observed by experienced sonographers.

* Corresponding author.
E-mail address: g.ambroise@chru-nancy.fr (G. Ambroise Grandjean).

https://doi.org/10.1016/j.diii.2018.08.001
2211-5684/

Ultrasound machines have now integrated software to facilitate and optimize fetal measurements. However, this software is still semi-automatic and the acquisition of the correct ultrasound plane still relies on the operator. Image pattern recognition and machine learning algorithms enable this software to identify the optimal placement of the calipers [1,2].

Three-dimensional (3D) ultrasound also has a role in obtaining planes for fetal biometric measurements. The triplane function, using a sweep of the ultrasound beam, allows a volume to be obtained from which the correct measurement planes can be extracted and calipers placed at a later moment [3].

The development of artificial intelligence software has allowed the combination of 3D volume analysis and image recognition to extract the correct measurement planes from a 3D volume for further optimal caliper placement. This technique of measuring is still at an experimental stage and has currently only been shown to perform optimally in the second trimester. A high degree of agreement between the measurements obtained via the software and those obtained conventionally from two-dimensional (2D) images was recently demonstrated [4]. This study also demonstrated that subsequent measurements were identical when the software was applied repeatedly to the same fetal volume [4]. However, a certain amount of variability was observed when the software was applied to two distinct 3D volumes that were acquired at the same ultrasound examination [4]. This variability is in line with inherent changes that occur during the acquisition of a volume such as the angle of the transducer during a sweep or due to maternal or fetal movements. The analysis of this variability shows a reproducibility which is slightly superior [4] to that seen in comparing 2D reproducibility.

The purpose of this study was to evaluate the feasibility and reproducibility of using the artificial intelligence software, Smartplanes® to automatically identify the transthalamic plane from 3D ultrasound volumes and to measure the biparietal diameter (BPD) and head circumference (HC) in fetus.

Materials and Methods

Inclusion criteria

This was a prospective cross-sectional pilot study carried out at a multidisciplinary prenatal diagnostic center. All women undergoing ultrasound examination with sonographer A, an experienced sonographer, for a period of one month, were included. The inclusion criteria were: maternal age ≥ 18 years old, gestational age between 16 and 30 weeks and a singleton pregnancy. Exclusion criteria were: suspected or known malformation of the head, the fetal head inaccessible on ultrasound, refusal to participate in the study. High BMI hampering the performance of ultrasound examination, only women with a body mass index (BMI) < 25 kg/m^2 were included.

The biometric images and the 3D ultrasound volumes were recorded anonymously. The observational study using anonymized data did not require Institutional Review Board approval.

Imaging protocol

All examinations were performed using a Resona 7® (Mindray) ultrasound unit equipped with a D8-4U probe. This ultrasound machine includes a software program, Smartplanes®, which enables the automated identification of the correct scanning planes within the head volume and the automatic positioning of the calipers and ellipse for measuring the biparietal diameter (BPD) and head circumference (HC). The software has a large stored dataset of 5000 images corresponding to the desired ultrasound planes for biometric measurements. It has sample images demonstrating correct caliper placement and others where the caliper placement is erroneous. Using this data-bank, the artificial intelligence program is able to select the transthalamic plane and make the biometric measurements while taking into account the variability of the anatomical structures and the fluctuations in the characteristics of the image (Fig. 1).

For each fetus, operator A (G. G.), who was a gynecologist with strong experience in obstetric ultrasound, recorded a 2D image of the transthalamic plane and also acquired two 3D volumes of the head using the Smartplanes® software.

The 3D volume was acquired from a transverse plan containing the septum cavum pellucidum anteriorly and the cerebellum posteriorly using a 5-MHz probe with harmonics. After each 3D volume was obtained, the Smartplanes® software identified the transthalamic plane and positioned the calipers between the outer border of the proximal parietal bone and the inner border of the distal parietal bone (outer to inner) in order to measure the BPD. An ellipse was positioned around the outer border of the skull to measure HC. By obtaining these measurements, the program was able to register that the automatic analysis was successful and this allowed the volume to be stored. When measurements were not obtained automatically, the volume was not stored. The failure was recorded, and a further volume acquisition was performed in order to obtain measurements and store the volume.

After all images had been collected, measurements were performed manually by operator A and by operator B (G. A.) who was a midwife specialist in 2D and 3D ultrasound. 2D images and 3D volumes were randomly presented to the two operators to minimize recall bias. Each operator was blinded to the measurements made by the other operator. BPD and HC measurements were performed twice by each operator on each 2D transthalamic image (2D measurements). They also examined the 3D volumes to obtain the transthalamic plane and then measured BPD and HC (3D measurements) (Supplementary data). Automated measurements obtained by the Smartplanes® software were also recorded (3D auto measurements). In order to measure the BPD, the operators applied their usual practice of positioning the calipers on the outer borders of the parietal bones (outer to outer). This was different from Smartplanes® measurement of ''outer to inner'' and therefore a systematic difference between the measurements of slightly less than 2 mm was expected. For HC measurement the method used by the operators and the software was identical. The study flow chart is shown on Fig. 2.

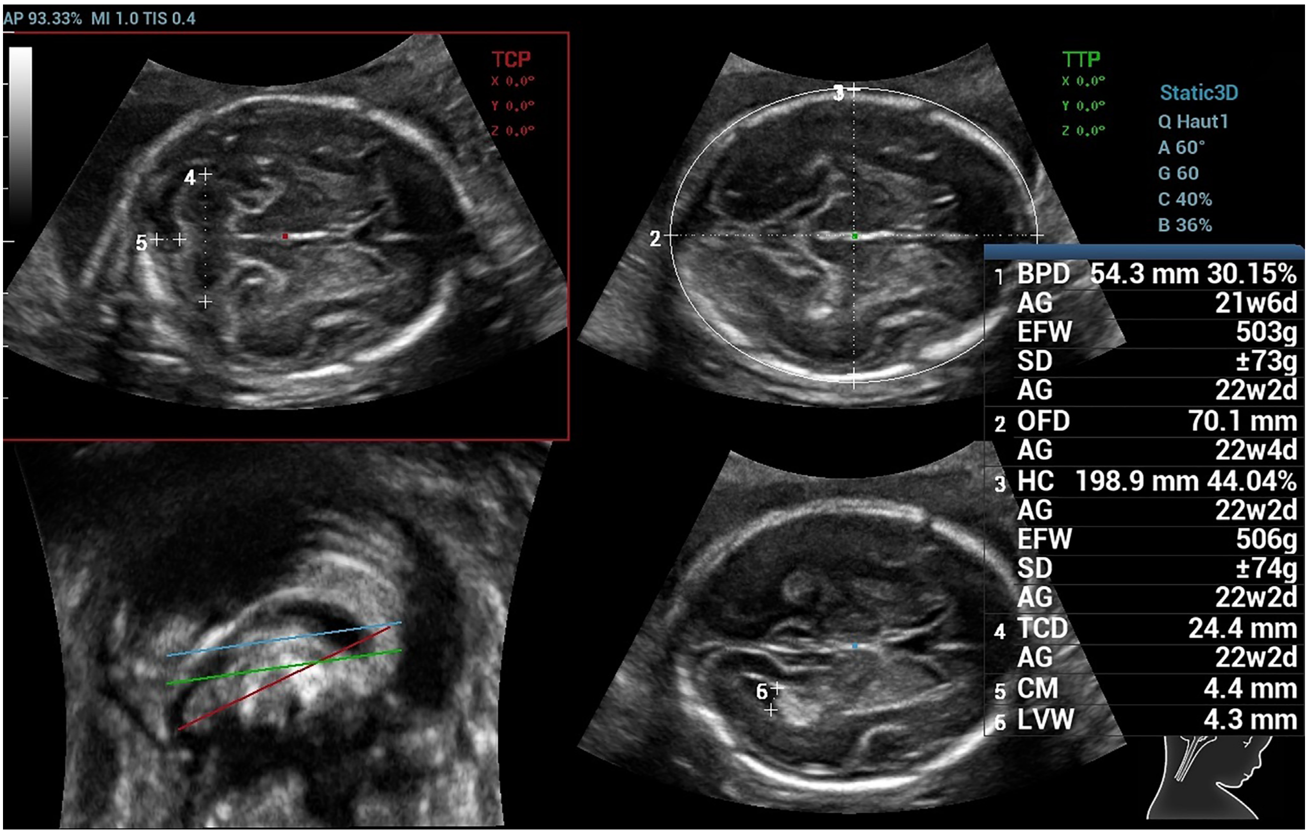

Figure 1. Results obtained after analysis of a head volume using SmartPlanes® software. Transcerebellar plane, transthalamic plane, median sagittal plane, and transventricular plane.

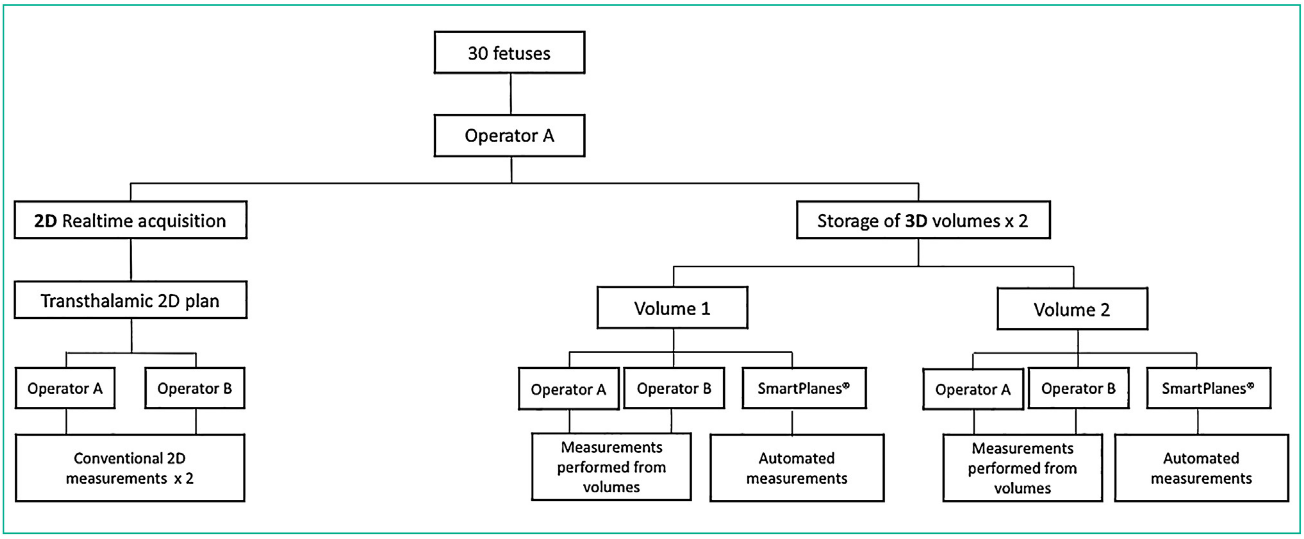

Figure 2. Study flow chart diagram.

Operator A performed image acquisitions in order to meet the quality criteria of the Intergrowth 21st study [5]. The two experienced operators then independently did quality control when measurements were carried out. Only images that met the intergrowth quality criteria were used for this study.

Statistical analysis

The concordance between the different measurements (2D, 3D and 3D auto) was evaluated by comparing the different series of measurements to the average of the measurements obtained by operator A in the original 2D transthalamic

Table 1 Biparietal diameter and head circumference values obtained by two independent operators and using SmartPlanes® technology.

	Biparietal diameter		Head circumference	
	2D	3D	2D	3D
Operator A	59.8 ± 10.0	59.8 ± 10.0	209.3 ± 36.5	208.9 ± 36.0
Operator B	60.0 ± 10.1	59.5 ± 9.9	211.3 ± 36.5	206.5 ± 38.8
SmartPlanes®	—	58.2 ± 9.7	—	212.9 ± 36.1

2D indicates two-dimensional. 3D indicates three-dimensional. Results are expressed in mm as mean ± standard deviation.

image. Intraoperator reproducibility was evaluated by comparing values obtained with repeated measurements of the same parameters by each operator and the Smartplanes® software. The interoperator reproducibility was assessed by comparing the average of the two 3D measures obtained by operator B and the average of the 3D auto measures obtained by the software against the average of the 3D measurements obtained by operator A. The minimal sample size to assess reproducibility for these measurements was estimated at 30 fetuses. Variability between the series of measurements was assessed using interclass correlation coefficient (ICC) with its 95% confidence interval (CI) and Bland-Altman plots. Significance was estimated using F test. A threshold < 5% was considered significant. The statistical analysis was performed with R software version 3.3.1.

Results

A total of 30 patients were included in the study. The mean gestational age of the fetuses was 23 weeks ± 3.2 (SD) (range: 17—29 weeks gestation).

In 2/60 acquisitions (3.3%), the initially obtained 3D volume did not permit the Smartplanes® software to obtain the BPD and HC measurements. However, a second acquisition carried out immediately afterwards enabled all of the planned measurements to be made. The total duration for both volume acquisition and automated analysis was always less than 10 s. Table 1 presents the means and standard deviations of the measurements made the two operators and the software.

The 95% confidence intervals (95% CI) of the difference observed between 2D measurements and 3D measurements of BPD and HC were 2 and 7.5 mm, respectively. These intervals were 4 and 11 mm, respectively for the differences between measurements obtained by the software (3D auto) and the conventional 2D measurements (Fig. 3).

Regarding the intraobserver reproducibility, the ICCs were greater than 0.99 for comparison of the measurements 3D and 3D auto obtained from the two separate volumes acquired during the same ultrasound examination by the two operators and the software (Table 2). Regarding BPD, the 95% CI for the difference observed between the repeated measurements was of 1 mm for the two operators (3D measures) and 2.5 mm for the software (3D auto measures). Regarding HC, the 95% CI for the difference observed was 5 and 7.5 mm, respectively for the 3D measurements of operators A and B and 10 mm for the 3D auto software (Fig. 4).

Regarding interobserver reproducibility, ICCs were > 0.90 for comparisons of 3D measurements by operator A with 3D measurements by operator B and 3D auto measurements (Table 3). The 95% CI for BPD measurements between operator A compared with operator B and the 3D auto software were 1.5 and 3 mm, respectively. For HC measurements the intervals were 6 and 9 mm respectively (Fig. 5).

Discussion

The reproducibility of a set of measurements has an impact on its accuracy and therefore its predictive value. This is particularly important in the second and third trimesters when fetal measurements are being used to predict growth and may result in important decisions regarding timing of delivery. The accuracy of these measurements is affected by intraoperator and interoperator variability, especially in a busy hospital setting where measurements to monitor growth may be performed by different operators [6].

Smartplanes® software makes it possible an automated search of transthalamic plane from a 3D volume and then automatic measurement of BDP and HC values. This has the potential advantage of standardizing the measurement technique and therefore minimizing inter operator variability.

This study shows that measurements obtained using Smartplanes® from a 3D volume agree closely with those obtained by two experienced sonographers from 2D images and 3D volumes. The observed variability is lower than previously reported for comparisons between 2D measurement and those measured manually from 3D volumes [3]. Regarding all comparisons made in our study, the reproducibility was greater or equal to that reported in a larger scale 2D study [7]. The intraoperator reproducibility of the software is very slightly lower than that observed by the two expert sonographers, however, it is still very high.

Measurements of the BPD yielded a systematic difference of about 1.5 mm between those taken by the sonographers and the software. This can be explained by the discordance in caliper placement between the sonographers (outer to outer) and the software (outer to inner). Any changes in the usual measurement techniques of a sonographer could theoretically have an impact on the reproducibility of those measurements. It is because of this that the operators applied their usual technique of BDP measurement. While there is an international agreement regarding placement of the ellipse for HC measurement, recommendations

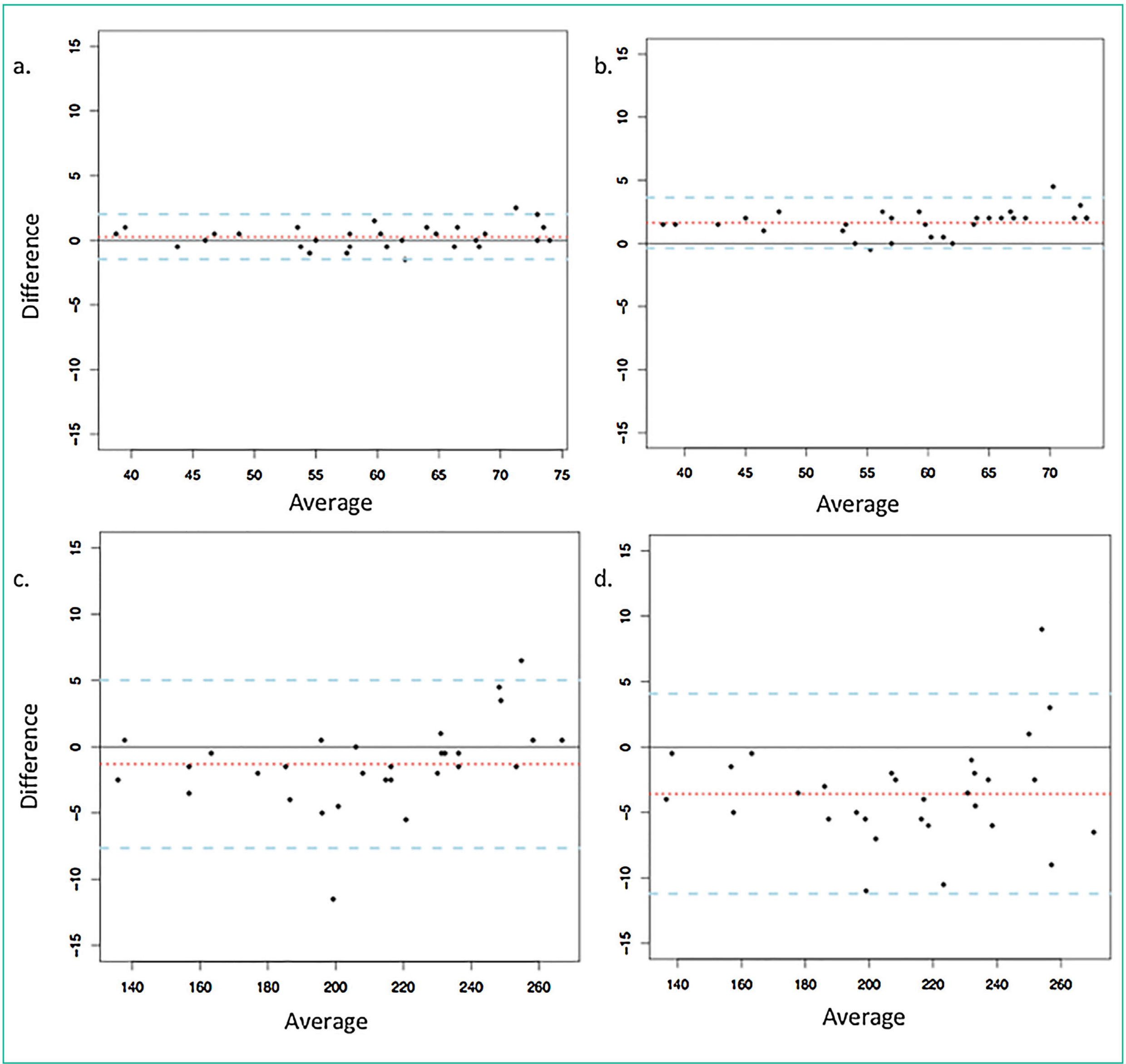

Figure 3. Graphs show agreement between 3D measurements (mean operator B) and 2D conventional measurements for BPD (a) and HC (c) and between automated measurements (mean Smartplanes®) and 2D conventional measurements for BDP (b) and HC (d). Difference and average expressed in millimeters.

Table 2 Intraobserver variability for fetal head measurements obtained by two independent operators and using SmartPlanes® technology.

	Operator A	Operator B	SmartPlanes®
BPD	0.998 < ICC < 0.999	0.997 < ICC < 0.998	0.994 < ICC < 0.997
HC	0.997 < ICC < 0.998	0.996 < ICC < 0.998	0.990 < ICC < 0.995

BPD indicates parietal diameter. HC indicates head circumference. ICC indicates intraclass correlation coefficient.

concerning the BPD measurements vary. This explains why the software was programed to measure the BPD differently.

The fetal position made it possible to find an acquisition plane close to the expected scanning plane for all 60 head volumes studied. The software provided correct caliper positions after one, or occasionally two, acquisition attempts. Thus, the use of Smartplane® software significantly limited constraints associated with probe manipulation and orientation of the ultrasound beam.

Acquisition of an ultrasound volume is usually a simple procedure, even for less experienced operators. However, treatment of the volume and caliper positioning to obtain

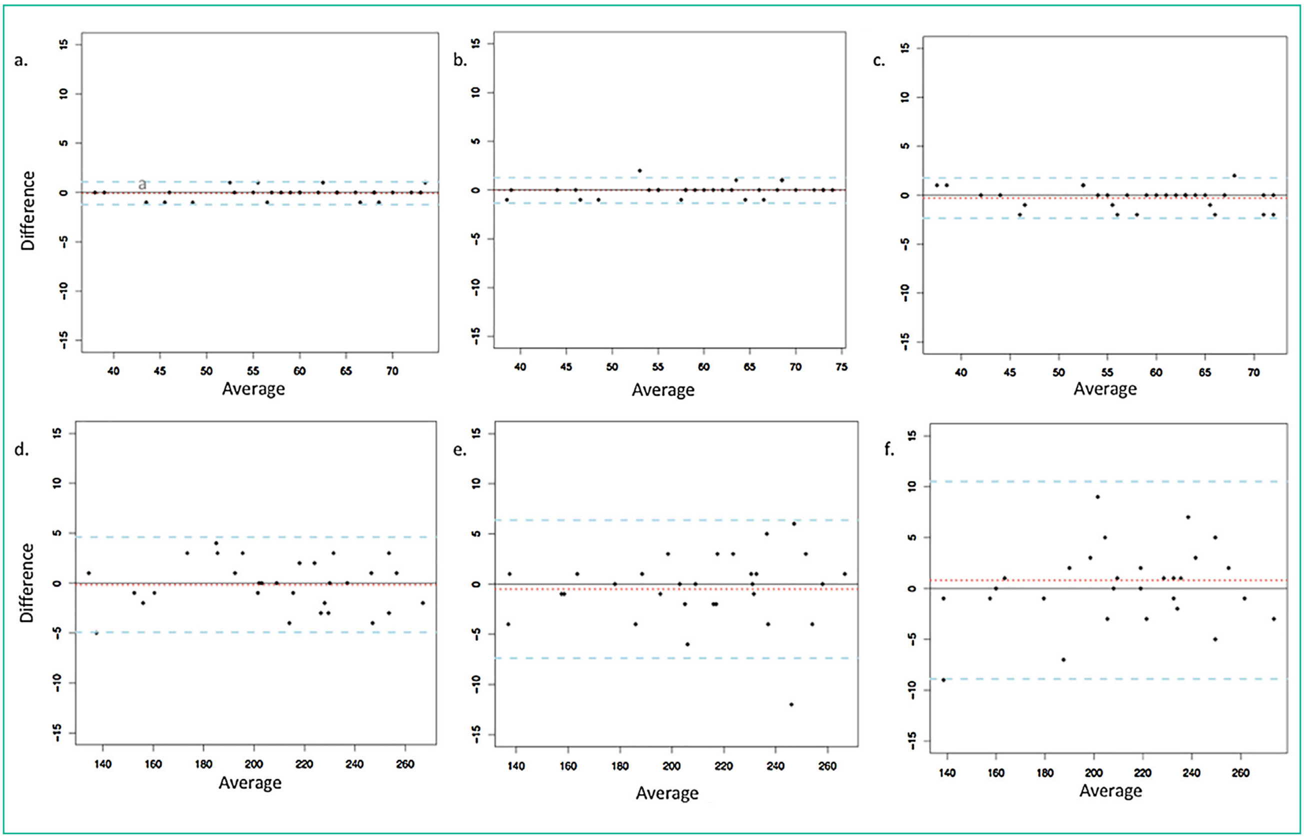

Figure 4. Graphs show intraobserver variability for the BPD (a) and HC (d) measurements of operator A, the BPD (b) and HC (e) measurements of operator B and the BPD (c) and HC (f) measurements of Smartplanes®. Differences and averages are expressed in millimeters.

Table 3 Interobserver variability for fetal head measurements.

	Operator A vs. Operator B	Operator A vs. SmartPlanes®
BPD	0.998 < ICC < 0.999	0.807 < ICC < 0.997
HC	0.996 < ICC < 0.998	0.857 < ICC < 0.996

BPD indicates parietal diameter. HC indicates head circumference. ICC indicates intraclass correlation coefficient. The comparison was made on the mean of two measurements obtained from repeated volume acquisition by operator A or B or using Smartplanes® software.

the measurements require skill, time, knowledge of fetal anatomy and mastery of the volume-processing software. These results confirm that the Smartplane® automatic measurement software allows accurate, reproducible and rapid biometric measurements to be taken from 3D volumes. Furthermore, unlike the 5D technology [8], Smartplane® does not require that anatomical landmarks be placed in the volume to obtain head measurements therefore potentially overcoming the inaccuracies or the failures of measurements associated with the inexperienced operator. This, however, must be confirmed by further studies using the Smartplane® technology to compare accuracy of measurements using 3D volume acquisitions obtained by experienced and inexperienced operators.

Our results are consistent with those obtained with 5D technology [4]. That study, using 5D software on a series of 120 fetal head measurements, demonstrated that the software was successful in obtaining standard fetal head measurements 98.3% of the time with significant time saving compared to standard 2D biometry. Their intra- and interobserver reproducibility data was also similar to ours, showing results that were slightly higher than those observed in 2D.

However, these studies have both been performed on 3D volumes where the acquisition plane was optimal. It has been demonstrated, in a study comparing a series of non-automated measurements taken from 2D images and 3D volumes, that past a certain angle between the acquisition plane and the measurement plane, it is not always possible to reconstruct a 2D image to achieve the same quality as obtained by the real-time 2D image and therefore the concordance of biometric measurements is not guaranteed. Fetal position, especially in the third trimester, does not always allow for optimal acquisition of 3D volumes, therefore the Smartplane® software should also be evaluated comparing different acquisition angles.

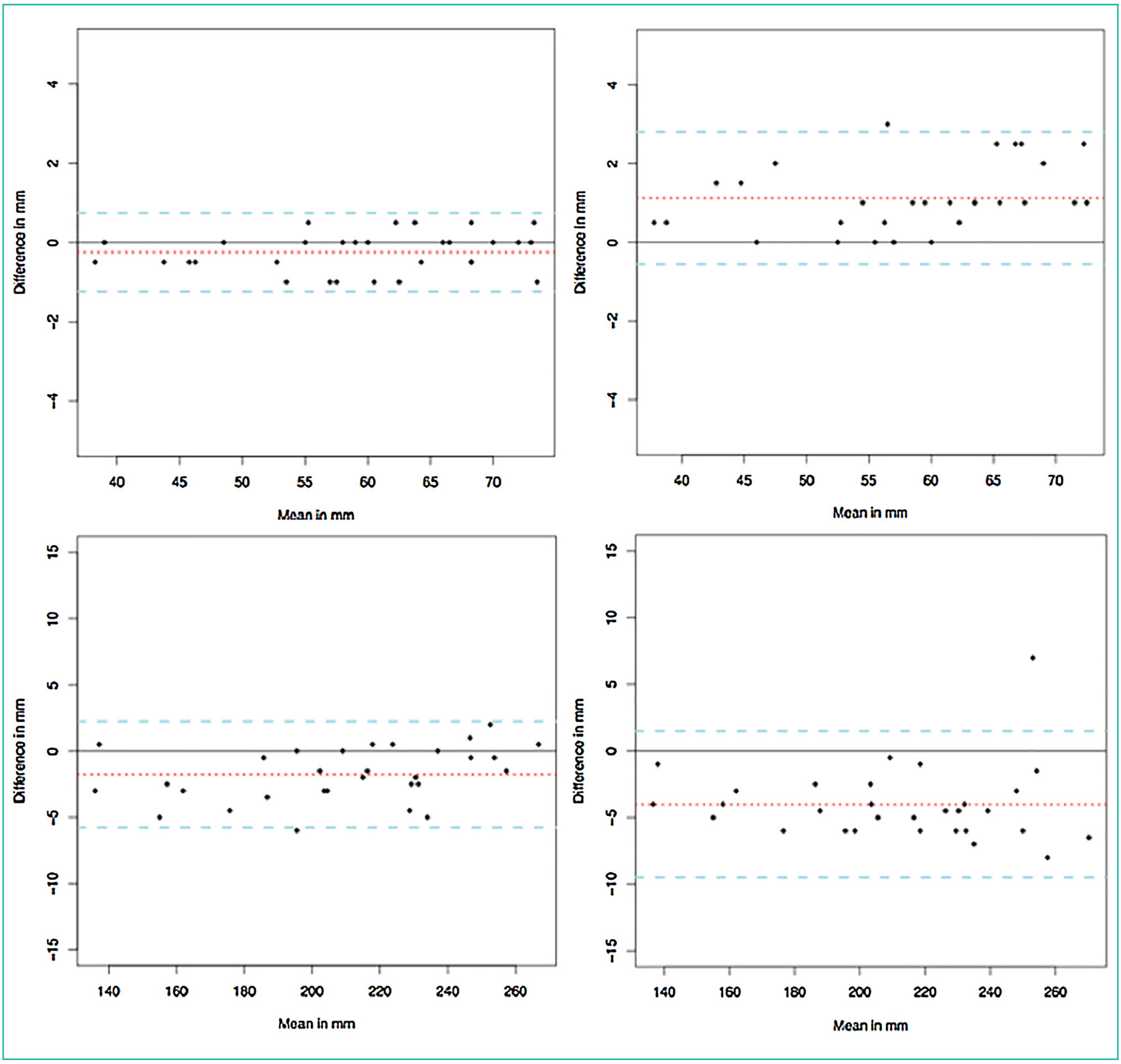

Figure 5. Graphs show interobserver variability between operator A and operator B of BPD (a) and HC (c) 3D measurements, and between operator A and Smartplanes® of BPD (b) and HC (e) 3D measurements. Difference and average are expressed in millimeters.

In 3D, the operator only has to position the probe opposite the cephalic pole. It was previously demonstrated that the time required to acquire the volumes was significantly shorter than that required for the acquisition of a 2D measurement set including femoral length, cephalic and abdominal perimeters: 45 s versus 117 s, respectively [3]. The difference between 3D and 2D might be even more important for an inexperienced sonographer.

In conclusion, advances in image recognition and algorithms have allowed the development of software capable of automatically selecting the correct plane and performing measurements of the fetal head which are similar to those obtained by experienced sonographers. More studies need to be done to confirm use of this technology in a variety of clinical settings. However, these results have opened the door to the practice of ultrasound assisted by artificial intelligence.

Appendix A. Appendix A Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.diii.2018.08.001.

Disclosure of interest

The authors declare that they have no competing interest.

References

[1] Yazdi B, Zanker P, Wanger P, Sonek J, Pintoffl K, Hoopmann M, et al. Optimal caliper placement: manual vs automated methods. Ultrasound Obstet Gynecol 2014;43:170—5.

[2] Zalud I, Good S, Carneiro G, Georgescu B, Aoki K, Green L, et al. Fetal biometry: a comparison between experienced sonographers and automated measurements. J Matern Fetal Neonatal Med 2009;22:43—50.

[3] Sarris I, Ohuma E, Ioannou C, Sande J, Altman DG, Papageorghiou AT, et al. Fetal biometry: how well can offline measurements from three-dimensional volumes substitute real-time two-dimensional measurements? Ultrasound Obstet Gynecol 2013;42:560—70.

[4] Rizzo G, Aiello E, Maria Elena Pietrolucci, Arduini D. The feasibility of using 5D CNS software in obtaining standard fetal head measurements from volumes acquired by three-dimensional ultrasonography: comparison with two-dimensional ultrasound. J Matern Fetal Neonatal Med 2016;29:2217—22.

[5] Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Ismail LC, Lambert A, et al. International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. Lancet 2014;384:869—79.

[6] Hamdaoui N, Lesieur E, Mancini J, Dabadie A, Bourderionnet V, Pico H, et al. Herman score in prenatal screening for Down syndrome: can a junior assess a senior? Diagn Interv Imaging 2017;98:155—60.

[7] Sarris I, Ioannou C, Chamberlain P, Ohuma E, Roseman F, Hoch L, et al. Intra- and interobserver variability in fetal ultrasound measurements. Ultrasound Obstet Gynecol 2012;39:266—73.

[8] Rizzo G, Pietrolucci ME, Capece G, Cimmino E, Colosi E, Ferrentino S, et al. Satisfactory rate of post-processing visualization of fetal cerebral axial, sagittal, and coronal planes from three-dimensional volumes acquired in routine second trimester ultrasound practice by sonographers of peripheral centers. J Matern Fetal Neonatal Med 2011;24:1071—6.