Vol. 5 No. 2 (2026)

Pakistan Journal of Medical & Cardiological Review

https://pakjmcr.com/index.php/1/about

Online ISSN 3007-2387

Print ISSN 3007-2379

Inter Observer Agreement of Amniotic Fluid Index Measurement Across Maternal BMI Groups at 36-40 Weeks

Muhammad Azeem
Faculty of Allied Health Sciences, The Superior University Lahore, Pakistan
Email: muhammad.azeem@superior.edu.pk
Maryam Pervaiz
Department of Medical Imaging Technology, Faculty of Allied Health Sciences, The Superior University Lahore, Pakistan Email: maryampervaiz996@gmail.com
Amna Asif
Department of Medical Imaging Technology, Faculty of Allied Health Sciences, The Superior University Lahore, Pakistan Email: hanifasif1519@gmail.com
Mugheesa Jamil
Department of Medical Imaging Technology, Faculty of Allied Health Sciences, The Superior University Lahore, Pakistan Email: mugheesahamil1@gmail.com
Kiran Shakeel
Institute of Health Sciences, Khwaja Fareed University of Engineering & Information Technology Email: kiran.shakeel@kfueit.edu.pk

Author Details

Corresponding E-mail & Author*:
Amna Asif
Email:hanifasif1519@gmail.com
MaryamPervaiz
Email:maryampervaiz996@gmail.com

Keywords: Amniotic Fluid Index, Inter-Observer Agreement, Intra-Class Correlation Coefficient, Maternal Bmi, Ultrasonography, Obstetric Ultrasound, Pakistan

Received on 13 May 2026
Accepted on 11 Jun 2026
Published on 19 Jun 2026

Abstract

One of the useful parameters in obstetric sonography is the amniotic fluid index (AFI). Indirectly evaluates fetal well-being, uteroplacental function and amniotic fluid regulation. AFI measurements are, however, dependent on the technique used by the operator and are subject to variation depending on the placement of the probe, scanning angle, the pressure of the transducer on the skin and the ability to detect fluid pockets. Further, maternal body mass index (BMI) can also be a problem as increased adipose tissue results in loss of strength of ultrasound beam and decreases spatial resolution, which makes it more difficult to view amniotic fluid pockets. This can cause an increase in the variation of opinions between observers. This is particularly important for clinical decision making during late pregnancy (36-40 weeks), but information on inter-observer agreement by BMI category in low and middle income countries such as Pakistan is scarce. The study was cross sectional observational study conducted for 4 months in Asfand Clinic, Rahim Yar Khan, Pakistan. A consecutive non-probability sample of 110 singleton pregnancies between 36-40 weeks of gestation were recruited. Sample size was determined with a standard formula (Z = 1.96, P = 0.20, d = 0.075) to yield n = 110 at 5% significance and 80% power levels and using a standard formula to calculate the Cronbach Reliability Coefficient, which was 78% for AFI. The four quadrant AFI was done by two independent expert sonographers (Observer 1 and Observer 2) with GE S6 and Aplio 300 ultrasound machines using a 3-5 MHz convex transducer and these observers were not aware of the results of each other's

measurements. Participants were categorized by BMI: Normal (18.5-24.9 kg/m², n = 10), Overweight (25.0–29.9 kg/m², n = 92), and Obese (≥30.0 kg/m², n = 8). The Pearson r, ICC (two-way mixed absolute agreement), Cronbach's alpha, Bland-Altman analysis and One-way ANOVA were used for statistical analysis and were carried out in IBM SPSS Statistics v25. Pearson r value was 0.916 ($p < 0.001$); ICC was 0.915 (95% CI: 0.877–0.940); Cronbach's alpha value was 0.955, indicating good overall inter-observer agreement. Bland-Altman analysis showed that the systematic bias was negligible with mean bias of -0.013 cm ($p = 0.900$), and limits of agreement [-2.10, +2.07] cm, with 91.8% measurements lying within these limits. High values of ICC for the subgroups were also found for Normal (ICC = 0.905), Overweight (ICC = 0.897) and Obese (ICC = 0.996). The absolute differences in measurements between BMI groups were not significant ($F = 1.117$, df = 2,107; $p = 0.331$).

Introduction

Background and Physiological Significance of Amniotic Fluid

The physiological significance and background of amniotic fluid. The cavity of the uterus, which is lined by the amniochorion, is filled with amniotic fluid (AF) which has a vital role in normal fetal development during pregnancy. It serves three main purposes: first, it provides physical protection by regulating temperature, cushioning against mechanical trauma, and absorbing shocks. Second, it supports functional development by allowing fetal movements, swallowing, breathing practice, and musculoskeletal growth. Lastly, it maintains homeostasis through antibacterial properties, preventing premature uterine contractions and preserving cervical integrity. The primary contributors to amniotic fluid volume are fetal urine production which rises to approximately 100 mL/day near term and lung liquid secretion at approximately 8 mL/kg/h. Reabsorption occurs principally through fetal swallowing and intramembranous absorption. Disruption of any of these mechanisms for example; due to renal agenesis, obstructive uropathy, esophageal atresia or placental dysfunction results in pathological alteration of amniotic fluid volume, which may represent an early clinical marker of fetal compromise [1,2].

Ultrasound measurement of amniotic fluid volume is now a routine component of antenatal surveillance. Two principal sonographic methods are employed; the Amniotic Fluid Index (AFI), a semi-quantitative four-quadrant summation technique and the Single Deepest Pocket (SDP) measurement. The AFI technique first described by Phelan et al. in 1987 divides the maternal abdomen into four quadrants using the linea nigra as the vertical axis and the umbilicus as the horizontal axis. The deepest vertical fluid pocket in each quadrant, free of fetal parts and umbilical cord is measured perpendicularly to the floor and the four values are summed to yield the AFI. This method is widely favoured for its comprehensive, quadrant-specific fluid assessment and its established integration into biophysical surveillance protocols [2, 3].

Clinical Role of AFI in Fetal Surveillance

The 2nd of the 5-point biophysical profile or modified BPP is the AFI, along with the non-stress test (NST). AFI values are normally between 5 and 25 cm with peak values occurring in the early third trimester that then slowly wane. AFI less than 5 cm (oligohydramnios) is correlated with fetal growth restriction, placental insufficiency, umbilical cord compression and higher incidence of operative delivery. AFI over 25 cm (polyhydramnios) is associated with preterm labor, gastrointestinal and neuromuscular structural anomalies in the fetus and hyperglycemia in the mother. Therefore, it is essential that AFI is correctly measured and reproducible and that it is measured in the last week before delivery (36-40 weeks) when the decision about timing and mode of delivery is made and that it is measured correctly. AFI is a semi-

quantitative instrument that has limitations but is clinically useful. It has been reported to be associated with increased incidence of intervention for suspected [4,5]. Oligohydramnios but no improved perinatal outcomes when compared to SDP. Therefore, there is some evidence-based guidelines that recommend SDP for low-risk pregnancies. However, AFI is still the main method used in clinical practice worldwide because it is perceived as being comprehensive and familiar [5, 6].

Inter Observer Variability and Its Determinants

There is an established inter-observer variation in AFI measurement with several potential sources of this variation including different selection of deepest fluid pocket, different angulations and pressures of the probe, different interpretation of the borderline fluid pockets, and different demarcation of quadrants. The small differences in transducer position up to 2–3 degrees of obliquity can cause clinically significant variations, especially when fluid volume is at the borderline [7,8]. Standardization of the scanning technique is therefore necessary to reduce the variabilities in the measurements and to improve the diagnostic consistency.

Consistency of AFI measurement has been shown to be modulated by operator training and clinical experience. Skilled sonographers also show greater pocket identification skills and are more adept in adjusting their scanning technique to accommodate less than ideal circumstances, such as unfavourable fetal position, oligohydramnios and poor acoustic windows which can be a challenge for less experienced sonographers [9]. Structured training programmes and simulation-based competency assessment have been suggested to minimise inter-operator variation, especially in units with multiple operators and training areas [10].

Influence of Maternal BMI on Ultrasound Image Quality

Maternal obesity (BMI ≥30 kg/m2) is an emerging public health problem worldwide, especially in South Asian countries where inactivity, imbalance of calories in and out, and genetic factors are responsible for high prevalence of women of reproductive age (WRA) in Pakistan. Obesity has a well-defined negative effect on obstetric ultrasound quality: the thickness of subcutaneous and visceral fat reduces penetration depth of the ultrasound beam, decreases image resolution and causes greater acoustic shadowing artefacts. These effects can sometimes cause less accurate observations of amniotic fluid pockets and lead to a higher intra-observer and inter-observer discrepancy [11, 12].

Obesity during pregnancy has obstetric consequences in addition to the image quality. Women with obesity have increased risks of gestational diabetes mellitus, pre-eclampsia, venous thromboembolism, wound infection, induced labour and caesarean section. There's also a higher risk of adverse outcomes in the baby, such as stillbirth and preterm birth, macrosomia and intrauterine growth restriction. The accuracy of the ultrasonographic assessment, especially the reliable measurement of AFI is of even greater clinical value and technically challenging in the Obstetric population of the obese women [12, 13].

Rationale and Objectives

Although there are several documents describing the variability in AFI measurement and its factors, there is not enough evidence regarding inter-observer agreement in particular among late pregnancy (36–40 weeks) women with variable BMI, in the Pakistani clinical settings, where BMI distribution within the obstetric population is likely to be different as compared to the higher income countries. To fill this evidence gap, this study aimed to precisely quantify the inter-observer agreement within BMI stratified subgroups and assess if there was a significant difference between measurement consistency within each BMI group when using a standard scanning protocol, when scanning by a trained physiotherapist.

The major aim was to test the agreement between 4 quadrant AFI measurements at 36-40 weeks gestation among groups of mothers with different BMIs. The secondary goal was to assess the effect of the maternal BMI on AFI measurement accuracy, reliability and precision between two independent expert sonographers. The study hypothesis was that inter-observer agreement would be significantly less in women with higher BMI than in women of normal BMI.

Literature Review

AFI measurement has been tested to be reproducible in several clinical settings. Sande et al. [14] performed a prospective study and found that the inter-observer variability was satisfactory for both AFI and SDP in all trimesters of gestation. They found that the directionality of the probe and the correct identification of the deepest vertical pocket were important technical factors to ensure measurement consistency. Importantly, BMI was not identified as significantly affecting the agreement when standardised guidelines were used. A detailed study of inter-observer variation of fetal biometry and amniotic fluid measurements by maternal BMI in late pregnancy was conducted by Martins et al. [15].

The authors' findings showed that with the four quadrant method applied constantly, the BMI had no negative effects on the image quality but that increased BMI did have a negative effect on soft-tissue attenuation. Their point was that experience of the sonographer is the foremost safeguard against image problems arising from BMI. Bhide et al. [16] evaluated the agreement between AFI and SDP by standardizing the measurement procedures, across various BMI categories. They noted good inter-observer agreement despite BMI being higher in the image quality of ultrasound images, concluding that standardized methodology did not compromise the reliability of AFI regardless of body habitus.

This evidence formed the basis for the continued use of AFI in clinical practice in a variety of patients. Odeh et al. [17] evaluated the reproducibility of AFI and SDP measurements in groups according to BMI and found excellent agreement between observers in all BMI groups. Critically, they conclude that this variation in BMI did not affect the clinical classification of amniotic fluid status (normal vs. abnormal), except that they suggest caution when considering borderline AFI in obese patients because of the potential for variability affecting management.

Coomarasamy et al. [18] in this study specifically looked at the contribution of operator experience on the reproducibility of AFI measurement. Their study showed that although the higher BMI sonographers' measurements were more consistent and variable between sonographers, the lower BMI sonographers' measurements were not. This important discovery made operator skill more important in contributing to inter-observer variability than patient-related factors like BMI. Khalil et al. [19] described the mechanisms which contribute to the inaccuracy of the measurement of AFI in the context of maternal obesity. It was estimated that the main cause of ultrasound beam attenuation was due to increased adipose tissue mass, and that deep fluid pocket visualization was difficult. The authors recommended some adaptive methods, including modification of the placement of the probe, adjustment of the probe angle, and gain optimization, which are useful for the sonographers to overcome the challenges of imaging with BMI and to maintain the accuracy of measurement. Hebbar et al. [20] have reviewed the use of AFI for routine obstetric monitoring and highlighted the technical factors that might affect the variability like fetal position, operator's technique and maternal adiposity. They recommended that AFI be used in conjunction with other methods of surveillance such as the non-stress test and biophysical profile because they felt this combination offers greater diagnostic confidence, and that the use of AFI alone has limitations in certain groups of technically challenging patients.

Singh et al. [24] found that inaccuracy of measuring AFI was not the most important source of variation when evaluating the AFI, but rather the choice of the amniotic fluid pocket. The problem was more prominent in subjects with higher BMI, and more discrimination being required to differentiate fluid pockets from neighbouring structures like fetal limbs, umbilical cord loops. In spite of this, the variation was comparable to clinically acceptable limits and the four quadrant AFI method is still clinically useful when the sonographer is trained to identify the pockets in the obese population.

Amrutha and colleagues [25] explored how maternal BMI affects AFI estimation and concluded that obese women were more likely to experience a low AFI value because of poor visualization of the deep shadows of the fluid which is caused by the increased thickness of the abdominal wall. The study, though, also showed that with patience in optimizing the acquisition of images by varying both the contact pressure of the probe and the scanning angle, fluid pocket visualization could be significantly improved and underestimation of the measurements avoided.

In obese women, Gupta et al. [26] also showed that it was often necessary to repeat repositioning attempts and take longer scans to obtain good images. In spite of these difficulties, affordable AFIs could be achieved and the authors stated that technical perseverance and not the exclusion of patients or diagnostic abandonment was indeed the right clinical approach in the face of imaging difficulties in this patient group. When specifically examining the relationship between maternal BMI and ultrasound window quality, Kendall et al. [27] determined that there was a consistent relationship between the two, with higher BMI consistently causing a decrease in window size and the margin of the amniotic fluid pockets was less clear. They recommended system level technical interventions such as reducing the ultrasound frequency downward, increasing the gain to make the fluid pocket brighter and using harmonic imaging to help reduce the image quality degradation caused by BMI. They reported that these changes should be codified into standard obstetric ultrasound procedures for women who are overweight.

The effectiveness of AFI compared to SDP is currently under discussion in high BMI populations. SDP is a simpler technique, may be less subject to inter-observer variability as it only requires one pocket of amniotic fluid to be identified rather than four. Several randomized studies have shown that SDP can decrease the need for labour induction in the presence of suspected oligohydramnios, while maintaining good perinatal results, and some international societies have suggested considering SDP in low-risk pregnancies. AFI is, however, still widely used in many centres due to its extent and ability to evaluate fluid distribution quadrant by quadrant and its better fit to the existing biophysical surveillance protocols [5, 6]. New obstetric ultrasound technology will provide the potential to minimize the use of observer measurement in amniotic fluid. Three-Dimensional ultrasound and automated volume calculation algorithms have been investigated to provide an objective and reliable means of calculating AFV that is less prone to the subjective interpretation inherent in the 2D AFI measurement. Preliminary studies have reported good results, and automated three-dimensional measurements have been shown to have less inter-observer variability than conventional two-dimensional AFI measurements.

The routine use of three-dimensional amniotic fluid assessment, however, is limited by the equipment's cost, the lack of availability especially in low resource areas, longer acquisition time and the requirement for specialized operator training. Thus, in most obstetric units around the world including in Pakistan, two dimensional AFI is still the standard of care [10, 15]. When considered as a group, published literature clearly conveys a negative correlation between maternal BMI and ultrasound image quality, and an added technical challenge for AFI measurement. When standardized four quadrant techniques are strictly adhered to by highly trained and experienced sonographers, however, generally acceptable inter-observer agreement can be

obtained between BMI groups. Based on this body of evidence, the most important finding is that protocol compliance plus sonographer training and technical adaptation have a greater effect on reliability of measurements than does maternal body habitus alone. The current study adds to the international evidence base of primary data collected in a clinical setting within Pakistan where overweight and obesity in pregnancy are becoming more common.

Materials and Methods

Study Design

Study Design The study involved a cross sectional observational design with the aim of evaluating inter-observer agreement in the measurement of AFI in different BMI groups of mothers. The choice of cross-sectional methodology was justified because of its feasibility, cost-effectiveness and appropriateness in the context of defining a clinical population and measuring agreement statistics at a specific point in time. The study received institutional approval and informed consent was obtained from all participants before they joined the study.

Study Setting and Duration

The data were collected from the Asfand Clinic at Rahim Yar Khan, Pakistan which had modern diagnostic ultrasound facilities and a large number of people attending clinic for antenatal check-ups. The study period lasted four months after the approval of the synopsis, which was long enough for a continuous recruitment of patients, a standardisation of data collection and statistical analysis.

Sample Size Calculation

Determination of sample size was done based on the standard proportional formula as follows: $n = (Z^2 \times P \times (1 - P)) / d^2$ With $Z = 1.96$ (95% confidence level), $P = 0.20$ (estimated proportion based on prior literature) and $d = 0.075$ (margin of error). The sample size was determined based on Cronbach Reliability Coefficient (CRC) of 78% for AFI measurements reported by Pourhoseingholi et al. [21] and a target of 80% statistical power at 5% significance level yielding a minimum sample of 110 participants.

Sampling Methods and Recruitment.

Consecutive sampling technique was used as a non-probability sampling method. All the patients attending the antenatal ultrasound clinic of Asfand Clinic during the study period who fulfilled the inclusion criteria were screened for eligibility and then consecutively selected for inclusion in the study until the required number of samples were obtained we used the convenient sampling technique.

Eligibility Criteria

Inclusion Criteria

Pregnant women with singleton gestations at 36-40 weeks confirmed by reliable dating

All maternal BMI categories (normal, overweight, obese) included to enable BMI stratified analysis

Ability and willingness to provide written informed consent

Capacity to maintain supine positioning for the duration of the ultrasound scan

Exclusion Criteria

Gestational diabetes mellitus (GDM) related to polyhydramnios and affecting AFI values

Confirmed fetal anomalies - structural anomalies that affect fetal swallowing or renal function change amniotic fluid dynamics

Placental dysfunction with intrauterine growth restriction (IUGR) with potential for bias in AFI values if associated with oligohydramnios
Multiple gestations in multifetal pregnancies confuses 4 quadrant AFI interpretation due to altered fluid distribution
Ruptured membranes renders AFI measurement useless..

Ultrasound Equipment
Two calibrated machines (i) GE S6, (ii) Toshiba Aplio 300 with a 3–5 MHz convex array transducer) were used for all scans. These systems are well known and widely used in obstetric ultrasonography and are well suited for the AFI measurements in clinical routine with good image quality. To ensure consistency of measurement, machines were serviced and calibrated before commencement of data collection.

Scanning Protocol
Every scan was done in the supine position, with slight left lateral tilt as necessary to make the scan more comfortable for the mother. Sufficient ultrasound gel was used for the correct acoustic coupling. A protocol was standardized to reduce the variability in technique. All fetal biometric parameters biparietal diameter (BPD), femur length (FL) and abdominal circumference (AC) were measured in addition to AFI.
For AFI measurement, the transducer was held perpendicular to the maternal abdomen and the floor in each quadrant. The maximum vertical fluid pocket, free of fetal parts and umbilical cord loops, was identified and measured in each of the four uterine quadrants. The four pocket measurements were summed to yield the total AFI. Both observers performed measurements independently and consecutively, with no communication between them during the scanning session. Findings were recorded on a standardized predesigned proforma.

BMI Classification
Maternal weight and height were measured at the time of the ultrasound appointment using calibrated equipment. BMI was computed as weight (kg) divided by height squared (m^2). Participants were classified per WHO BMI categories; Normal weight (18.5–24.9 kg/m^2), Overweight (25.0–29.9 kg/m^2) and Obese (≥30.0 kg/m^2).

Data Collection Procedure
Following enrolment and consent, each participant underwent a standardized obstetric ultrasound by Observer 1, during which fetal biometric measurements and four-quadrant AFI were recorded. Observer 2 then independently performed AFI measurements on the same patient without knowledge of Observer 1's values. Both sets of measurements were entered into the predesigned proforma for each participant. Data entry was conducted in a chain system to minimize transcription error, with completed proformas checked for completeness and accuracy before entry into SPSS.

Statistical Analysis
All statistical analyses were performed using IBM SPSS Statistics Version 25. Descriptive statistics (mean ± SD, median, minimum and maximum) were computed for all continuous variables. The normality of measurement distributions was assessed visually and via descriptive statistics. The following tests were applied;
Pearson correlation coefficient (r): to quantify the linear association between Observer 1 and Observer 2 AFI values
Intraclass Correlation Coefficient (ICC): two-way mixed model, absolute agreement, with 95% confidence intervals. ICC ≥ 0.90 was classified as excellent, 0.75–0.89 as good, and < 0.75 as poor-to-moderate agreement

Cronbach's reliability coefficient (alpha, α): to assess internal consistency of repeated measurements

Paired-samples t-test and Bland-Altman analysis: to evaluate systematic bias between observers. Limits of agreement (LoA) were calculated as Bias ± 1.96 × SD of differences

One-way ANOVA: to test for significant differences in absolute AFI discrepancy (|Observer 1 − Observer 2|) across the three BMI groups

Statistical significance was set at $p \leq 0.05$ for all tests. Data are presented in tabular and graphical form.

Results

Demographic and Clinical Characteristics

A total of 110 pregnant women at 36–40 weeks gestation were enrolled. The demographic and clinical characteristics of the study sample are presented in Table 1. The majority of participants were overweight (n = 92, 83.6%), reflecting the BMI distribution of the obstetric population at the study site. Normal weight women accounted for 9.1% (n = 10) and obese women for 7.3% (n = 8). The mean ± SD maternal BMI across the full sample was 26.74 ± 2.25 kg/m². Mean AFI values measured by both observers were similar within each BMI group, ranging from approximately 12.24 to 13.93 cm.

Table 1. Demographic Characteristics and Mean AFI by BMI Group (n = 110)

Variable	Normal (n=10)	Overweight (n=92)	Obese (n=8)	Total (n=110)
BMI mean ± SD (kg/m²)	22.7 ± 1.91	26.8 ± 1.44	31.1 ± 1.26	26.74 ± 2.25
Gestational Age (wks+days)	37–39	36–40	37–40	36–40
AFI Observer 1 mean ± SD (cm)	12.48 ± 2.14	12.43 ± 2.34	13.88 ± 4.06	12.54 ± 2.48
AFI Observer 2 mean ± SD (cm)	12.24 ± 2.41	12.47 ± 2.54	13.93 ± 4.07	12.56 ± 2.66
% of Total Sample	9.1%	83.6%	7.3%	100%

Overall Inter-Observer Agreement

The overall inter-observer agreement statistics for all the 110 participants are shown in Table 2. There was a very strong positive linear relationship between the AFI values of Observer 1 and Observer 2, with a Pearson correlation of $r = 0.916$ ($p < 0.001$), and the variance in AFI values between the two observers accounted for 84% of the total variance ($R^2 = 0.840$). The ICC (two-way mixed, absolute agreement model) was 0.915 with a relatively narrow 95% confidence interval of [0.877 to 0.940] and well above the threshold of the excellent level of 0.90. The internal consistency (Cronbach's alpha) of the repeated measurements by the two observers was a remarkable 0.955.

Table 2. Inter-Observer Agreement Statistics-Overall and by BMI Group

Statistical	Overall	Normal	Overweight	Obese	Interpretation

Measure	(N=110)	(n=10)	(n=92)	(n=8)	
Pearson r	0.916	0.908	0.899	0.996	Very strong
Pearson p-value	< 0.001	< 0.001	< 0.001	< 0.001	Highly significant
ICC (absolute agreement)	0.915	0.905	0.897	0.996	Excellent
ICC 95% CI	[0.877, 0.940]	—	—	—	Narrow and precise
Cronbach's Alpha (α)	0.955	—	—	—	Excellent reliability
Mean Bias (cm)	−0.013	0.240	−0.037	−0.050	No systematic bias
SD of Differences (cm)	1.065	1.012	1.112	0.385	Acceptable variation
95% Limits of Agreement	[−2.10, +2.07]	—	—	—	Within clinical tolerance
% Within LoA	91.8%	—	—	—	Acceptable concordance

Bland-Altman Analysis

Bland-Altman analysis was performed to evaluate systematic bias and the clinical acceptability of measurement differences. The mean bias was -0.013 cm (paired t-test: t = -0.125, df = 109, p = 0.900), confirming the absence of statistically significant systematic error between the two observers. The standard deviation of the inter-observer differences was 1.065 cm. The 95% limits of agreement (LoA) ranged from -2.10 cm to +2.07 cm a total width of approximately 4.17 cm. A total of 91.8% of measurement pairs fell within these limits of agreement, confirming clinically acceptable concordance. Given that the normal AFI range spans 20 cm (5–25 cm), a maximum expected discrepancy of approximately 2 cm is unlikely to alter clinical classification or management decisions in the majority of clinical scenarios.

ICC and Pearson r by Maternal BMI Group

Subgroup analysis by BMI category confirmed excellent agreement in all three groups. ICC values were; Normal weight = 0.905, Overweight = 0.897 and Obese = 0.996 all substantially exceeding the 0.75 threshold for good agreement and the 0.90 threshold for excellent agreement in two of the three groups. Pearson r values showed a similar pattern; Normal = 0.908, Overweight = 0.899, Obese = 0.996. The obese subgroup demonstrated the highest agreement statistics of all groups, which was an unexpected finding given the anticipated challenges of ultrasound imaging in women with elevated BMI. This observation may reflect heightened operator attentiveness and technical adaptation during imaging sessions with obese participants, or it may represent a statistical artefact of the small obese group sample size (n = 8).

ANOVA Effect of BMI Group on Inter-Observer Variability

One-way ANOVA was conducted to test whether the absolute inter-observer AFI difference (Observer 1: Observer 2) differed significantly across BMI categories. The mean absolute differences were; Normal = 0.820 cm (SD = 0.986), Overweight = 0.715 cm (SD = 1.024) and Obese = 0.300 cm (SD = 0.350). The ANOVA yielded F = 1.117 (df = 2,107, p = 0.331), indicating no statistically significant difference in inter-observer variability across BMI groups. The null hypothesis that maternal BMI does not significantly influence inter-observer AFI agreement when standardized protocols are used was therefore accepted. These findings are summarized in Table 3.

Table 3. ANOVA Results (AFI Difference) by BMI Group

BMI Group	N	Mean |Diff| (cm)	SD	95% CI	ICC
Normal (18.5–24.9 kg/m²)	10	0.820	0.986	[0.115, 1.525]	0.905
Overweight (25.0–29.9 kg/m²)	92	0.715	1.024	[0.503, 0.928]	0.897
Obese (≥30.0 kg/m²)	8	0.300	0.350	[0.007, 0.593]	0.996
ANOVA Result	—	F = 1.117	—	p = 0.331	NS

Summary of All Statistical Tests

Table 4. Comprehensive Statistical Analysis

Statistical Test	Result	p-value	H_0	Interpretation
Pearson Correlation	r = 0.916	< 0.001	Rejected	Very strong positive linear correlation between observers
ICC (Two-Way Mixed)	0.915 [0.877–0.940]	< 0.001	Rejected	Excellent absolute agreement
Cronbach's Alpha	α = 0.955	—	—	Excellent internal consistency of repeated measures
Paired T-Test	t = −0.125, df = 109	p = 0.901	Not rejected	No significant systematic observer bias
Bland-Altman	Bias = −0.013 cm LoA [−2.10, +2.07]	NS	Not rejected	Clinically acceptable concordance; 91.8% within LoA
One-Way ANOVA	F = 1.117, df = 2,107	p = 0.331	Not rejected	No significant BMI group effect on inter-observer variability

Results and Discussion

The present study does offer reliable evidence for inter-observer agreement for measuring AFI in all four maternal BMI groups for all four quadrants with one method of scanning. The overall ICC value of 0.915, Pearson r value of 0.916 and Cronbach's alpha value of 0.955 indicated good reproducibility for measuring AFI

between two independent expert sonographers. These statistics compare favourably with reported values of excellent agreement (ICC ≥ 0.90) and are similar to those with the highest reported degree of agreement in the international obstetric ultrasound literature [22].

The Bland-Altman analysis confirmed the clinical acceptability of measurement discrepancies; the mean bias was -0.013 cm and the 95% limits of agreement were within the range [-2.10, +2.07] cm, which means that the differences between observers are unlikely to lead to different clinical classifications in most cases. The clinical thresholds for oligohydramnios (AFI ≤ 5 cm) and polyhydramnios (AFI ≥ 25 cm) are 20 cm apart, so the maximum expected difference in measurement from one observer to another of about 2 cm would not affect clinical management for women with AFI measurements in the middle range. However, patients with borderline AFI, especially those with AFI in the 5-7 cm range where decision to intervene may be close, even a 2 cm difference may have an impact on clinical decision making, thus emphasizing the need for standardization of measurement technique and clinical correlation.

The result of one-way ANOVA with standardized protocols ($F = 1.117$, $p = 0.331$) supports the hypothesis that the inter-observer variability in AFI measured with standardized protocols is not significantly different among women with different BMI categories. This finding directly challenges the research hypothesis and corroborates with the results reported by Martins et al. [15] and Bhide et al. [16] and Sahu et al. [13] who also found the inter-observer agreement acceptable in all the BMI groups with consistent technique. The mean absolute differences for the group Normal were Normal 0.820 cm, Overweight 0.715 cm, and Obese 0.300 cm, which showed an inverse relationship to BMI, in other words, no greater discrepancy with measurement was found as BMI increased in this cohort.

The ICC unexpectedly was very high (0.996) in the obese group and the mean absolute difference was lowest (0.300 cm) in this group, and this should be interpreted carefully. One possible reason is that sonographers were more careful and used more careful technique in obese patients thus reducing the "natural variability" they often displayed. Khalil et al. [19] mentioned that "the expert sonographer who is aware of the imaging difficulties associated with obesity will modify the probe pressure, angle, and optimize the image to deal with the limitations imposed by the adipose tissue and its attenuation. Another interpretation is statistical; for this small group of $n = 8$ obese individuals, the confidence interval for this group's ICC is large and the findings may not be reflective of the true agreement in the obese population.

Further research with a larger group of obese individuals is required to clarify this issue. Comparing the article with existing literature. The present findings are in line with the general trends of the literature in this area. While patient BMI was not the most important factor, operator experience was shown to be the most important determinant of consistency of AFI measurement by Coomarasamy et al. [18] who reported that only experienced sonographers showed a low variability in some obese women. The results of the present study support this. Both observers were sonographists and their experience together may have been a factor in the high overall agreement. The inter-observer discrepancies were also minor, with no significant differences in the AFI classification (normal vs. abnormal) between BMI groups, as found by Odeh et al. [17] which further corroborated the clinical strength of the four quadrant method.

Das et al [23] highlight the importance of protocol adherence for accurate measurements: the differences in the AFI were more likely to be related to the technique than to maternal BMI. When following consistent guidelines, Sande et al. [14] showed good reproducibility for both AFI and SDP, irrespective of BMI. This data, combined with the results of the present study, indicates that the four-quadrant AFI is a good indicator of fetal surveillance in the entire range of maternal BMI

categories experienced in obstetric practice. The population circumstances and context of Pakistan are unique. The demographic profile of the present study is in line with the obstetric population of Rahim Yar Khan as 83.6% of the women enrolled in the study were overweight. It is aligned with national epidemiological data showing the increasing prevalence of overweight and obesity in Pakistani women of reproductive age, due to urbanization, more sedentary lifestyle, shift in dietary pattern towards high calorie food intake and cultural norms that might have a negative effect on physical activity among women.

The mean maternal BMI of 26.74 ± 2.25 kg/m² indicates that the average participant was in the overweight range of BMI, highlighting the clinical relevance of understanding the effect of body habitus on obstetric ultrasound performance in this population. Specific contextual factors are present in the healthcare system of Pakistan which influence the provision of obstetric ultrasound services. Some facilities like the study site use a few sonographers who have a different level of training in standardized measurements and perform a lot of scans. The fact that a very high level of inter-observer agreement can be obtained with standardized technique with no special equipment or high-tech imaging techniques is especially useful in this setting. It states that high quality measurement of AFI is possible within the limited resources available in Pakistani obstetric practice, if resources are invested in the standardization of the protocol and constant training of sonographers.

The results of the study also have implications for the interpretation of the clinical decisions in the context of Pakistan. The high prevalence of overweight and obesity in the study population was coupled with a lack of significant difference in measurement reproducibility among the BMI categories, which means clinicians can trust the reliability of BMI-based surveillance and intervention thresholds, such as delivery timing, delivery frequency, and intervention thresholds, across BMI categories without differential concern about measurement reliability. Clinically relevant in situations where BMI might otherwise be used as a justification for lowered confidence in AFI or for dropping from standard AFI monitoring.

Methodological Strengths

The design of this study enhances the reliability of the results found in this study. A comparison of two experienced sonographers, with similar clinical experience, rather than between an experienced and novice operator, provides the agreement statistics a realistic clinical picture of routine obstetric ultrasound performed by qualified operators. The scanning protocol was standardized and provided for all techniques, such as probe type, patient positioning, and quadrant measurement rules, to restrict technique-related sources of variability and to increase the internal validity of the agreement analysis.

Use of multiple complementary statistical techniques (Pearson's r linear association, ICC absolute agreement with confidence intervals, Cronbach's alpha internal consistency, Bland-Altman analysis for systematic bias, and ANOVA between-group comparison) gives a multi-dimensional and comprehensive characterization of inter-observer agreement [22].

It is also informative for clinical decision making to include Bland-Altman limits of agreement, representing 95% of the differences between the observers that individual patients may experience, and the extent to which observer differences are likely to fall within clinically significant limits. The results of this study have a number of practical implications. They offer a sense of security that the four-quadrant AFI measurement is repeatable and clinically valid in overweight/obese pregnant women using a standard protocol. The diagnostic interpretation thresholds for clinicians and sonographists should not be changed based on the BMI of the mother alone, as long as technique is standardized.

Second, the results confirm the importance of education for sonographers and standardization of protocols as quality improvement targets in obstetric ultrasound. The availability of investment in structured training programmes, such as standardized AFI measurement guidelines, the simulation of skills and competency-based assessment, is likely to lead to a higher level of reproducibility of the measurement than technological interventions alone. Third, it is important to document maternal BMI at the time of each AFI scan as a standard practice. BMI annotation has the benefit of facilitating contextual interpretation of AFI values and can help to audit measurement variability across BMI categories within clinical quality assurance programmes, especially in the borderline cases.

Limitations

The following are limitations of this study. The size of the obese subgroup (n = 8) restricted the statistical power and generalizability of findings specifically for this subgroup. This study was done at one clinical site in Rahim Yar Khan, Pakistan, which might not be representative of other clinical settings with varying patient populations, equipment, or operator experience profiles. There was always going to be a time delay between the sequential measurements made by Observer 1 and Observer 2, and there was a risk that the amniotic fluid volume would have changed between the two observations due to one or more of the following: Fetal urination; swallowing; maternal position. This study did not include transvaginal ultrasonography because this technique may provide better acoustic penetration in obese women due to less tissue interface between the abdomen and the bladder or vagina. Finally, maternal hydration level, fetal presentation and bladder filling between measurements were not systematically controlled but are normal to routine clinical AFI measurement.

Conclusion

This cross-sectional study demonstrates that the four quadrant AFI method has good inter-observer agreement, all the way up to 36-40 weeks' gestation when measuring AFI for normal, overweight, and obese women with standardized scanning procedures. Mean bias of −0.013 cm, ICC of 0.915 (95% CI: 0.877–0.940) and Pearson r of 0.916, along with Cronbach's alpha of 0.955, all signify excellent measurement reproducibility. The inter-observer variability was not statistically significantly different among maternal BMI group (one-way ANOVA $F = 1.117$, $p = 0.331$) and the null hypothesis is accepted. The results highlight the clinical applicability and validity of the standardized four quadrant AFI measurement in all the BMI categories used in modern obstetric practice. The reliability of AFI is more dependent upon protocol standardization and the sonographer's training in measuring AFI than on patient BMI. Larger numbers of obese and/or underweight patients, studies over extended periods of time to determine intrapatient variability, and the use of more sophisticated imaging methods (e.g., 3D ultrasound, automatic volume measurement) to optimize AFO monitoring in more difficult patient populations should be included in future studies.

Clinical Recommendations

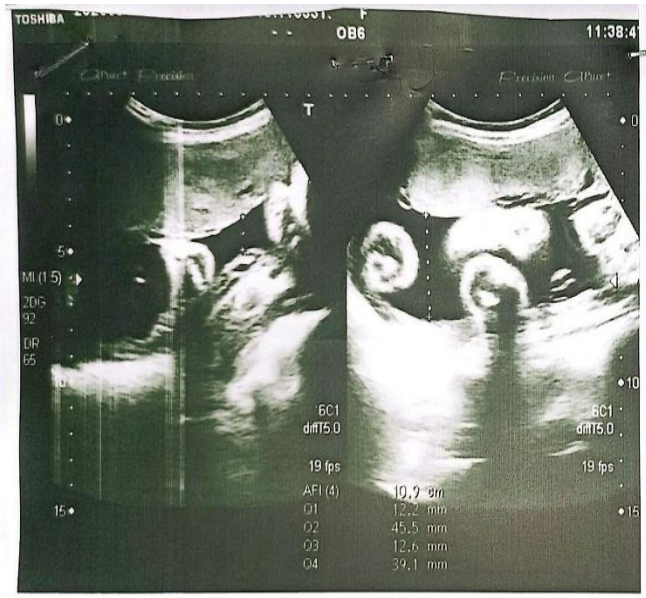
Continue to use four quadrant AFI as a standard of care for all BMI categories in routine obstetric care at term.

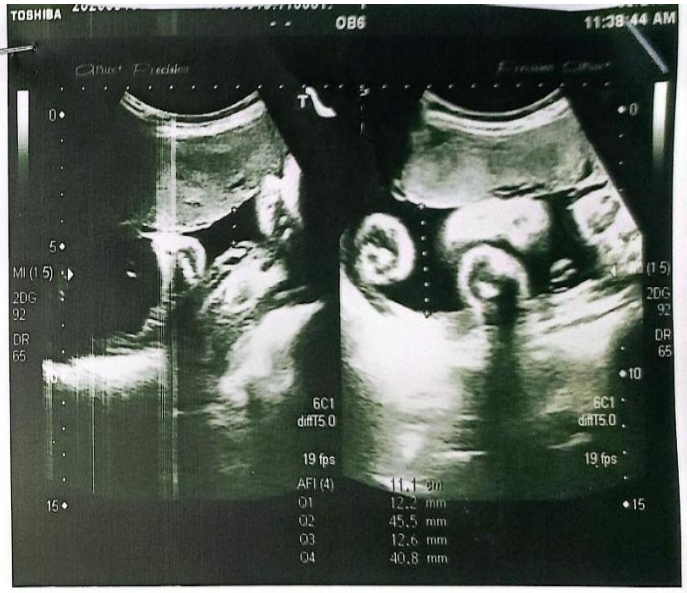
Describe and carry out competency-based assessment modules, protocol based training programmes for sonographers, especially for women with increased BMI to optimise technique.

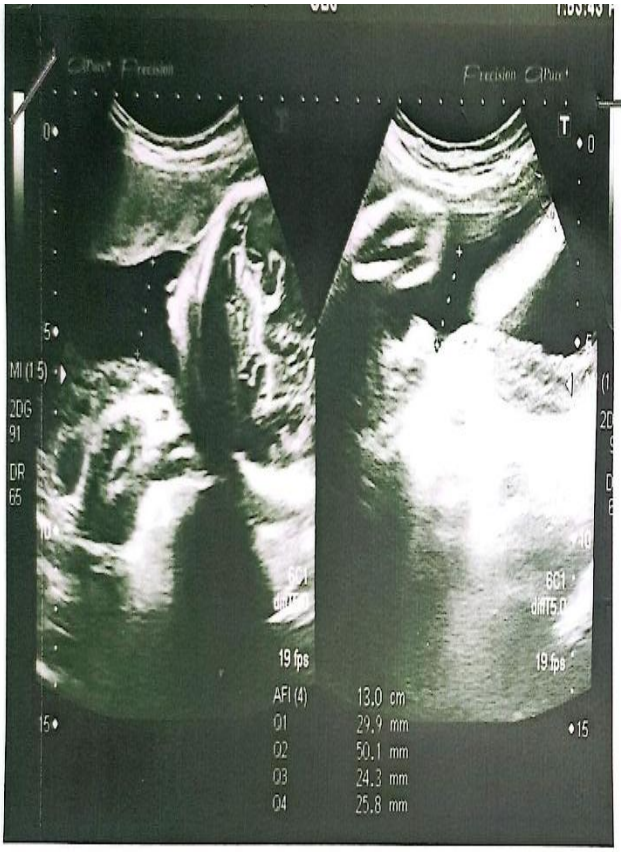
As standard practice, detail maternal BMI for each AFI measurement taken to support contextual clinical interpretation and audit of variability.

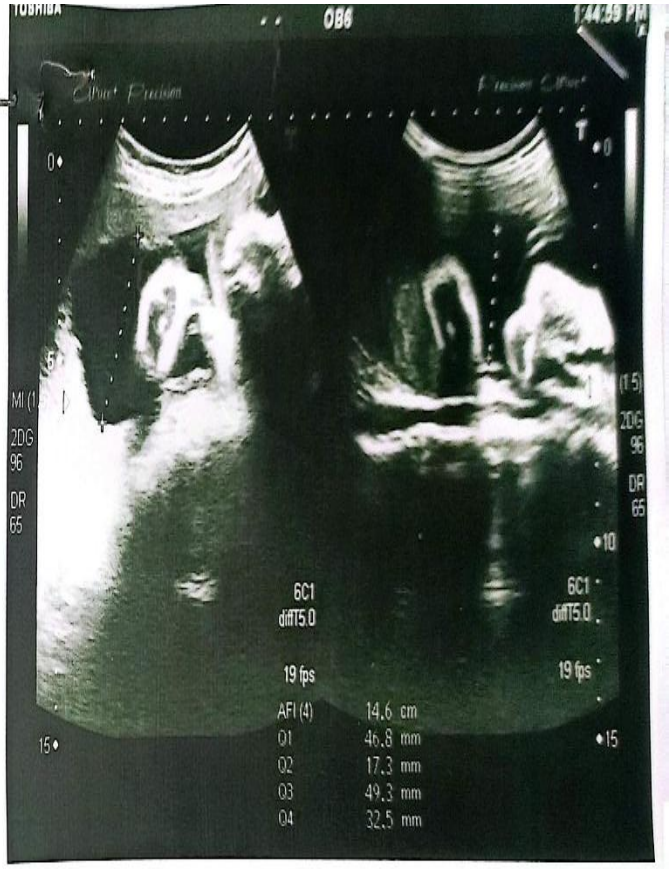
For obese women, use technical adjustments such as probe frequency, gain adjustment, depth optimization and multipocket assessment to optimize the image and minimize measurement error.
Technically challenging obstetric populations should be studied in future multicentre studies with larger subgroups of obese patients to further define inter-observer agreement.
AFI levels on the borderline of any BMI group must be evaluated with the clinical history, biophysical profile and non-stress test.

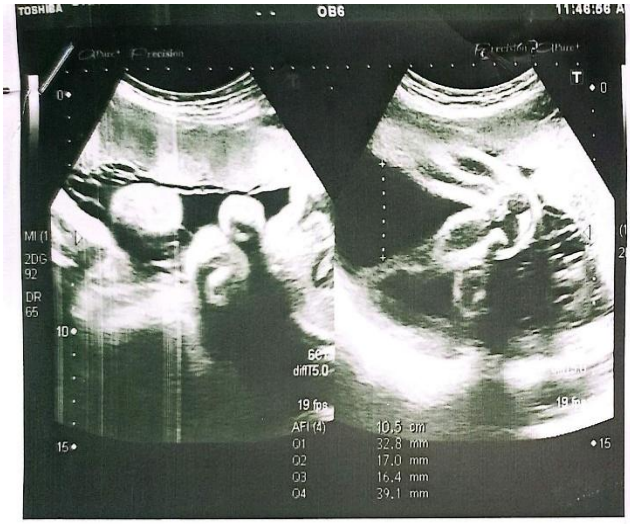
Ultrasonographic Images

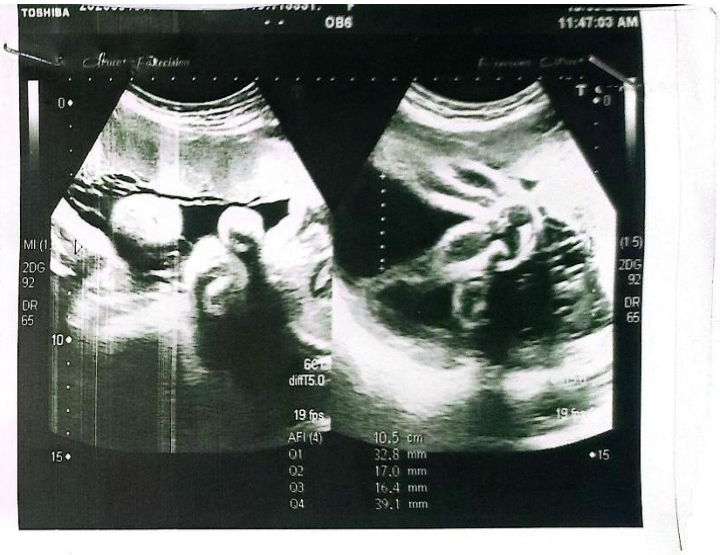


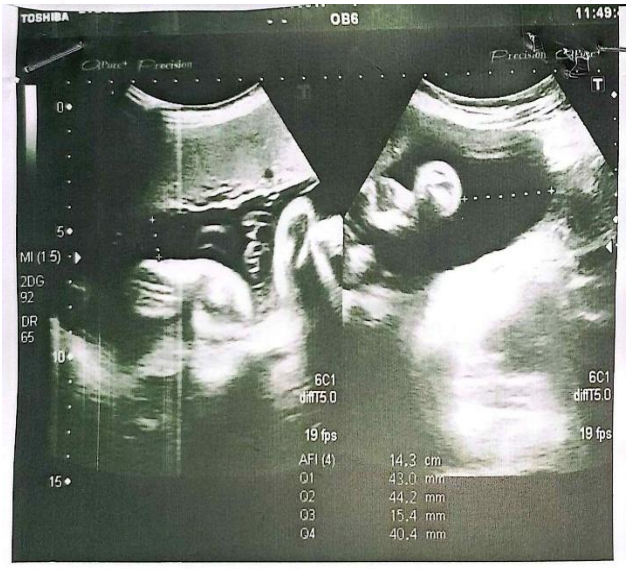


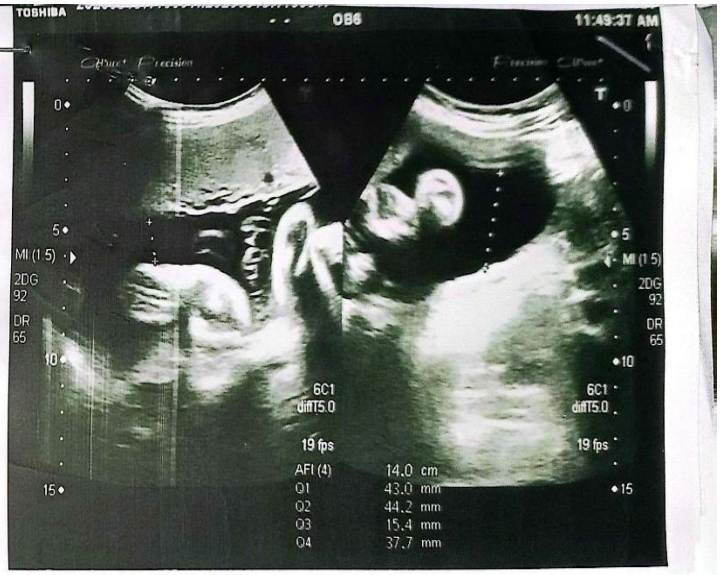


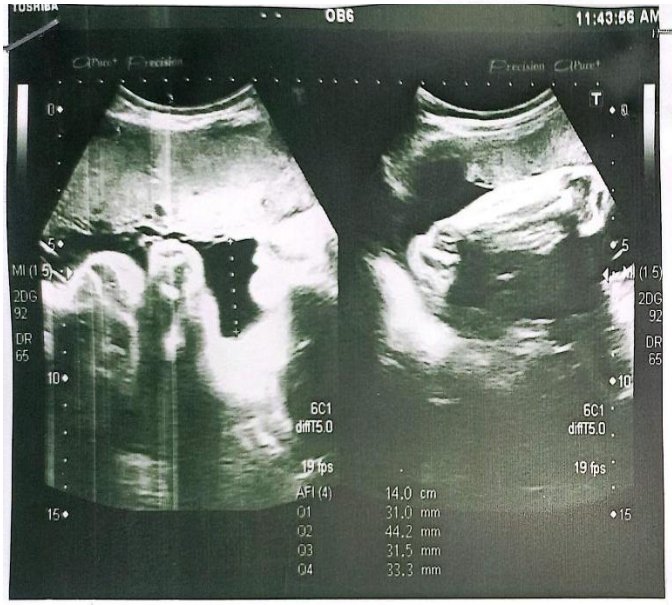


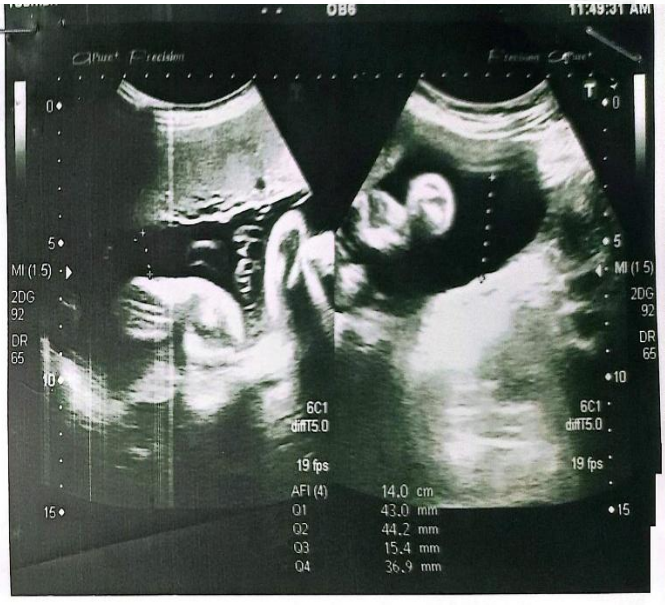

TOSHIBA
OB6
11:46:56 A
6C1
diffT5.0
19 fps
AFI (4) 10.5 cm
Q1 32.8 mm
Q2 17.0 mm
Q3 16.4 mm
Q4 39.1 mm

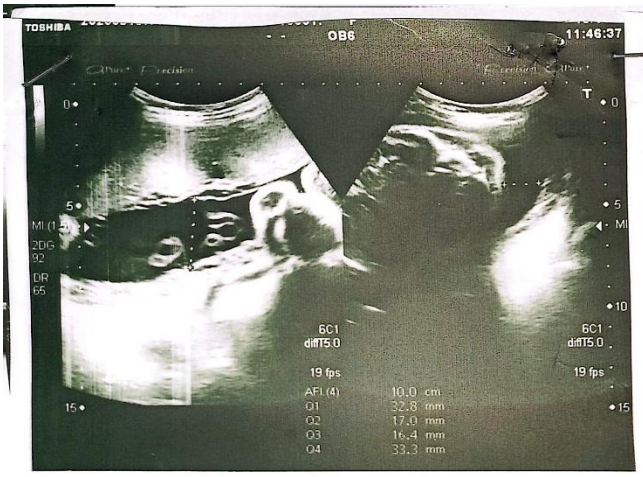

TOSHIBA
OB6
11:47:03 AM
6C1
diffT5.0
19 fps
AFI (4) 10.5 cm
Q1 32.8 mm
Q2 17.0 mm
Q3 16.4 mm
Q4 39.1 mm

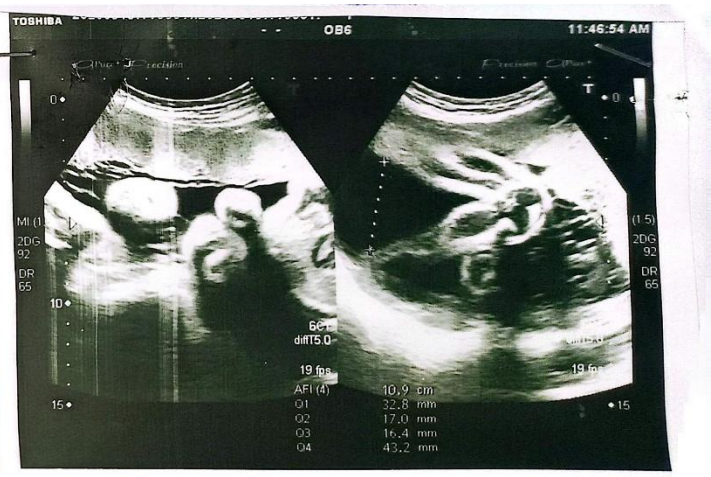

TOSHIBA
OB6
6C1
diffT5.0
19 fps
AFI (4) 14.3 cm
Q1 43.0 mm
Q2 44.2 mm
Q3 15.4 mm
Q4 40.4 mm

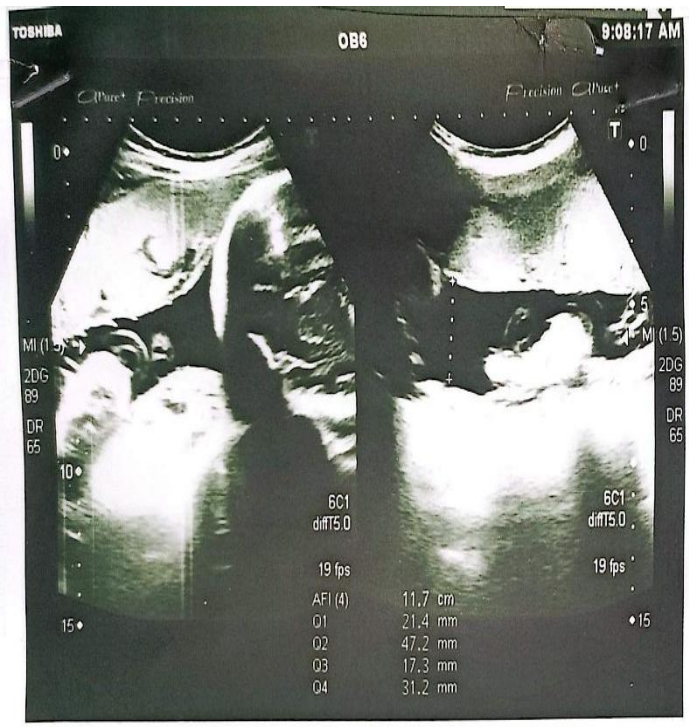

TOSHIBA
OB6
11:49:37 AM
6C1
diffT5.0
19 fps
AFI (4) 14.0 cm
Q1 43.0 mm
Q2 44.2 mm
Q3 15.4 mm
Q4 37.7 mm

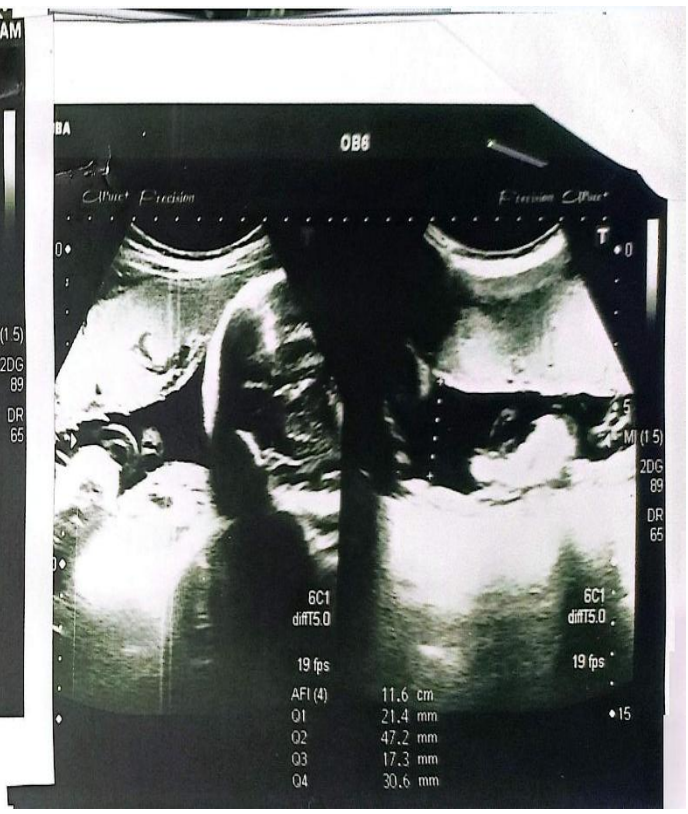

TOSHIBA
OB6
11:43:66 AM
6C1
diffT5.0
19 fps
AFI (4) 14.0 cm
Q1 31.0 mm
Q2 44.2 mm
Q3 31.5 mm
Q4 33.3 mm

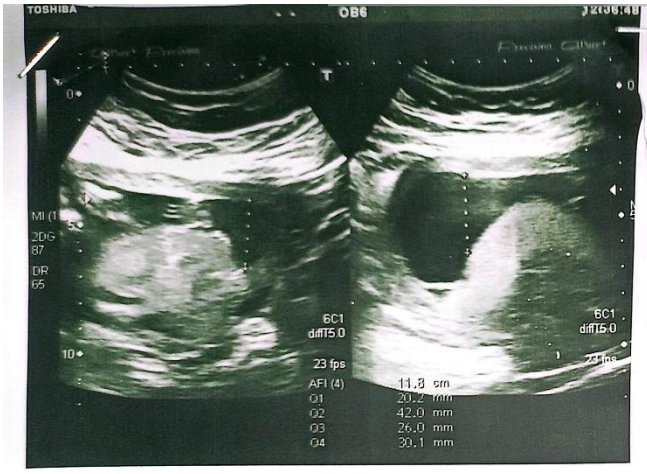

OB6
6C1
diffT5.0
19 fps
AFI (4) 14.0 cm
Q1 43.0 mm
Q2 44.2 mm
Q3 15.4 mm
Q4 36.9 mm

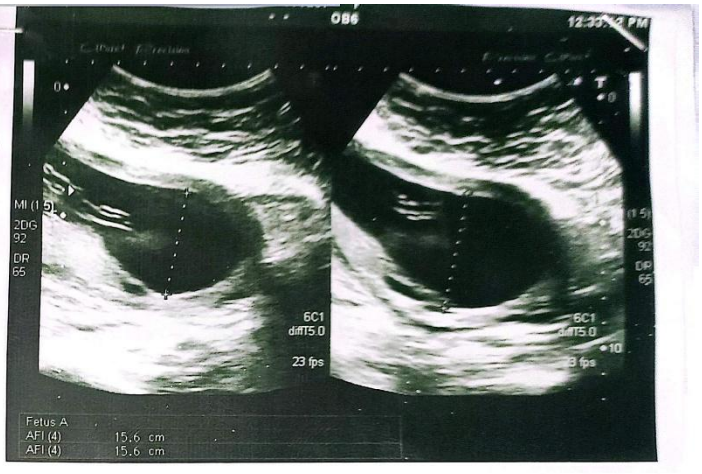

TOSHIBA
OB6
11:46:37
6C1
diffT5.0
19 fps
AFI (4) 10.0 cm
Q1 32.8 mm
Q2 17.0 mm
Q3 16.4 mm
Q4 33.3 mm

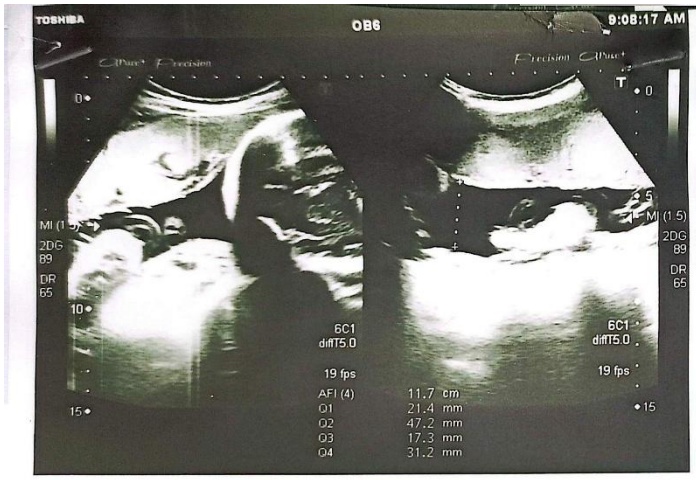

TOSHIBA
OB6
11:46:54 AM
6C1
diffT5.0
19 fps
AFI (4) 10.9 cm
Q1 32.8 mm
Q2 17.0 mm
Q3 16.4 mm
Q4 43.2 mm

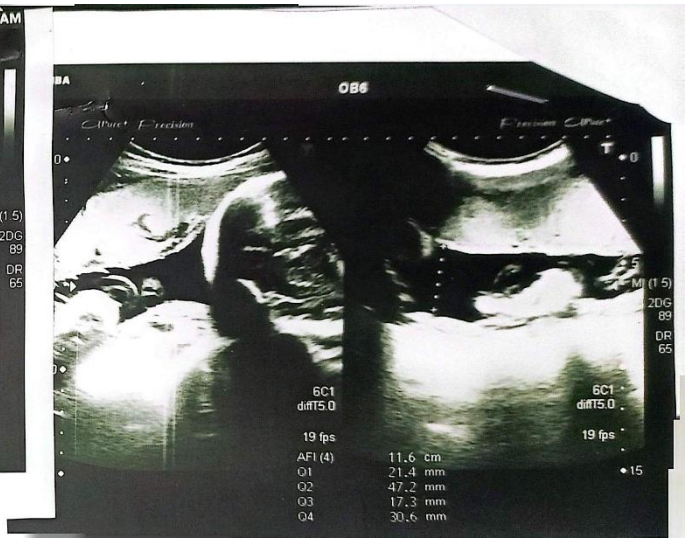

TOSHIBA
OB6
9:08:17 AM
6C1
diffT5.0
19 fps
AFI (4) 11.7 cm
Q1 21.4 mm
Q2 47.2 mm
Q3 17.3 mm
Q4 31.2 mm

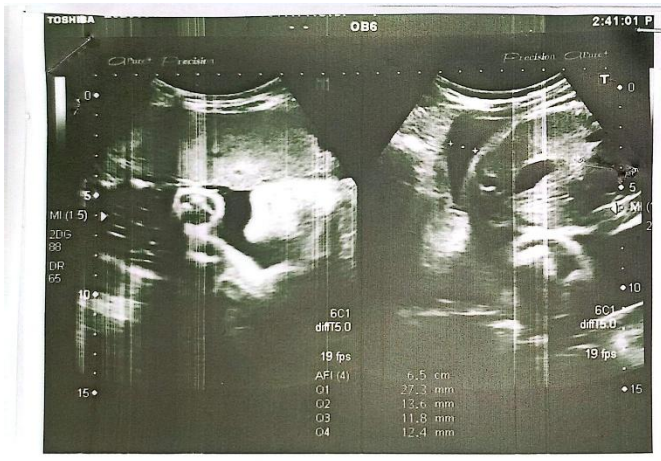

OB6
6C1
diffT5.0
19 fps
AFI (4) 11.6 cm
Q1 21.4 mm
Q2 47.2 mm
Q3 17.3 mm
Q4 30.6 mm

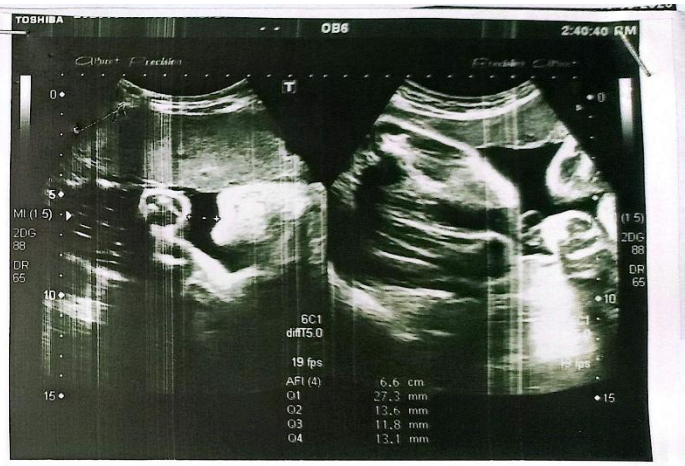

TOSHIBA
OB6
6C1
diffT5.0
23 fps
AFI (4) 11.8 cm
Q1 20.2 mm
Q2 42.0 mm
Q3 26.0 mm
Q4 30.1 mm

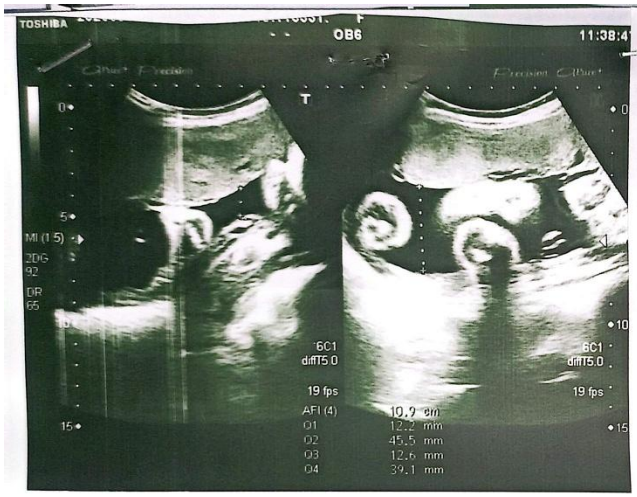

OB6
6C1
diffT5.0
23 fps
Fetus A
AFI (4) 15.6 cm
AFI (4) 15.6 cm

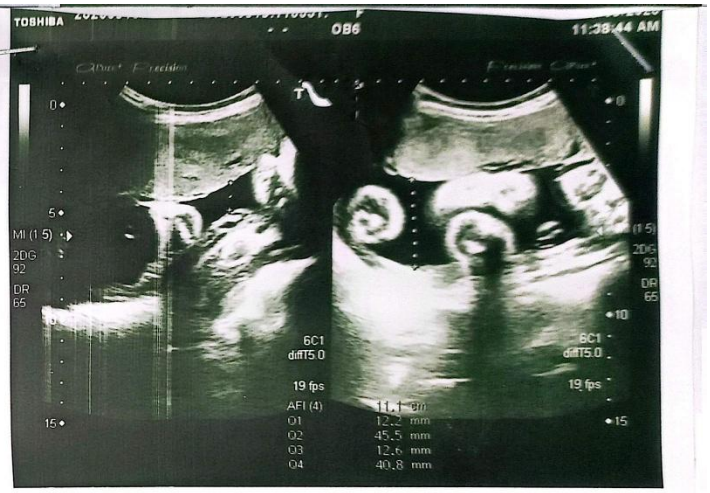












Declaration

Ethical Approval:

The study was conducted after approval from the relevant institutional authority. Written informed consent was obtained from all participants before data collection.

Conflict of Interest:

The authors declare no conflict of interest.

Funding:
No external funding was received for this study.

Data Availability:
Data supporting the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions:
All authors contributed to study design, data collection, statistical analysis, manuscript preparation, and final approval of the manuscript.

References

Phelan JP et al. Amniotic Fluid Index measurements during pregnancy.
Moore TR, Cayle JE. The amniotic fluid index in normal human pregnancy.
Magann EF et al. Comparison of AFI and single deepest pocket techniques.
Nabhan AF, Abdelmoula YA. Amniotic fluid index versus single deepest vertical pocket.
American College of Obstetricians and Gynecologists. Antepartum fetal surveillance guidelines.
Society for Maternal-Fetal Medicine. Oligohydramnios and polyhydramnios practice recommendations.
Martins WP, et al. Interobserver variation in ultrasound measurements during pregnancy.
Bhide A, et al. Reproducibility of amniotic fluid measurements across BMI categories.
Odeh M, et al. Observer agreement in AFI and SDP measurements.
Coomarasamy A, et al. Effect of operator experience on obstetric ultrasound reproducibility.
Khalil A, et al. Maternal obesity and ultrasound image quality in pregnancy.
Hebbar S, et al. Clinical applications and limitations of AFI measurement.
Sande JA, et al. Interobserver variability of AFI in different trimesters.
Kendall JL, et al. Impact of maternal BMI on ultrasound visualization.
Gupta N, et al. Technical challenges of AFI assessment in obese pregnant women.