CEOG Clinical and Experimental Obstetrics & Gynecology

Interobserver reliability of sonographic fetal biometry in second trimester maternal serum screening

H. Aksoy[1], S. Özyurt[2], Ü. Aksoy[3], Ö. Özdamar[4], E. Mutlu[5], Ş. Tutuş[6], M.A. Babayiğit[7], G. Açmaz[2]

[1] Department of Obstetrics and Gynecology, Kayseri Military Hospital, Kayseri
[2] Department of Obstetrics and Gynecology, Kayseri Education and Research Hospital, Kayseri
[3] Department of Obstetrics and Gynecology, Kayseri Memorial Hospital, Kayseri
[4] Department of Obstetrics and Gynecology, Gölcük Military Hospital, Kocaeli
[5] Department of Obstetrics and Gynecology, Çorlu Military Hospital, Tekirdağ
[6] Department of Radiology, Kayseri Education and Research Hospital, Kayseri
[7] Department of Public Health and Epidemiology, Gülhane Military Faculty of Medicine, Ankara (Turkey)

Summary

Purpose: To examine the interobserver variability for fetal biometry parameters and to investigate whether this variability affects the second-trimester maternal serum screening test (STMSS) results. *Materials and Methods:* A total of 60 singleton pregnancies who were scheduled for STMSS were investigated. Two experienced sonographers performed all examinations at the same visit. The risk calculations of screening were performed according to the each operator's biometric measurements separately. Interobserver variability in measurements of fetal biometrics and the effect of this interobserver variability on the screening results were assessed. *Results:* interobserver reliability for biparietal diameter (BPD) and femur length (FL) were 0.904 and 0.888 ($p < 0.001$), respectively. interobserver reliability coefficients for trisomy 21, trisomy 13/18, and neural tube defect were 0.887, 0.999, and 0.920 ($p < 0.0001$), respectively. *Conclusion:* The present results demonstrate that the interobserver reliability and agreement of ultrasound measurements of fetal biometry in cases of routine prenatal screening are highly reliable.

Key words: interobserver variability; Maternal serum screening; Triple test; Trisomy 21.

Introduction

Trisomy 21 (Down's Syndrome), is the most commonly encountered viable chromosomal abnormality [1], which affects approximately one in 800 live-born babies [2]. It is the commonest cause of mental retardation, and is also associated with a variety of congenital malformations and prenatal screening is now recommended routinely in many countries. Second-trimester maternal serum screening (STMSS) is the predominant non-invasive method of prenatal screening for trisomy 21 (Down syndrome), trisomy 18 (Edwards syndrome) or other type of chromosomal abnormality and neural tube defects (NTDs) in many countries. The STMSS based on maternal serum alpha-fetoprotein (AFP), unconjugated estriol (uE3), and free beta human chorionic gonadotropin (β-hCG) levels at 15 to 20 weeks of gestation detects approximately 70% of pregnancies affected by trisomy 21 for a false-positive rate of about 5% [3]. The results are also combined with the maternal age, maternal weight, ethnicity, and gestational age in order to assess probabilities of potential chromosomal abnormalities. This prenatal screening test allows an estimate of the risk of a pregnancy being affected and provides parents with information to guide them to make decision about definitive invasive testing [4]. Definitive invasive tests [amniocentesis (AS) and chorionic villus sampling (CVS)] allow the diagnosis of Down's syndrome but carry the risk of serious complications such as miscarriage. However, this detection rate can only be achieved consistently if fetal gestational age and the maternal serum markers are measured correctly. While the concentrations of AFP, uE3, and free β-hCG are determined objectively, the process of fetal gestational age assessment is a subjective process. Inaccurate determination of fetal gestational age affects prenatal screening performance. Biparietal diameter (BPD) and femur length (FL) are established measurements for dating the pregnancy [5]. The pregnancy must be dated accurately because errors will affect the assigned risk, causing false-negative and false-positive. Despite the importance of fetal biometry in dating of pregnancy and prenatal screening, few interoperator variability studies have been published. The aim of this study was to examine the interobserver reliability for fetal biometric parameters and to investigate whether this reliability affects the STMSS test results. In this study, the authors compared the interobserver reliability in the gestational age assessment of experienced radiologist and obstetrician in cases of routine STMSS.

Revised manuscript accepted for publication August 3, 2015

Clin. Exp. Obstet. Gynecol. - ISSN: 0390-6663
XLIV, n. 3, 2017
doi: 10.12891/ceog3246.2017

7847050 Canada Inc.
www.irog.net

Materials and Methods

A prospective clinical study was conducted at the present Obstetrics and Gynecology Clinic of Kayseri Education and Research Hospital, a tertiary referral centre in Turkey between September 2013 and March 2014. The study was approved by the institutional ethics committee and all participants signed an informed consent form to participate in the present study. A total of 60 singleton pregnancies with a gestational age between 16 and 18 weeks (112-126 days of gestation) who were scheduled for STMSS admitted to this Obstetrics clinic for routine prenatal care were investigated. The prenatal evaluation before STMSS was consistent with the clinic protocol and included comprehensive medical and obstetric examination along with obstetric ultrasound to determine the gestational age (BPD and FL) as well as to exclude any other pelvic and obstetric pathology.

Inclusion criteria for this study were: age 18 years or older; accepting STMSS as the method of prenatal screening for Down's syndrome, and other chromosomal abnormalities; a single viable intrauterine pregnancy confirmed by precise date of the last menstrual period and an ultrasound scan (up to seven weeks of gestation by crown-to-rump length); written approval, and willingness to comply with the study. Patients with any known high-risk conditions, including medical problems that could affect test results, multiple pregnancies, and known fetal congenital anomalies, were excluded from participation in the study. Moreover, pregnancies under 16 weeks' of gestation and pregnancies over 18 weeks' of gestation were also excluded from the study.

Two experienced sonographers (one experienced radiologist and one experienced obstetrician with average knowledge in obstetric ultrasound) performed all examinations at the same visit. Fetal biometrical measurements (BPD and FL) were obtained by each ultrasonographic examination using the same ultrasound machine in cases of routine STMSS. Each measurement was performed once by each operator. The initial measurement was recorded by the first sonographer who is an experienced obstetrician (S.O). Subsequently, a second sonographer who is an experienced radiologist (S.T), blinded to the results of the first sonographer, performed the same measurements. The fetal biometric measurements of the first sonographer were always removed from the ultrasound screen, after a hard copy had been made, before the second sonographer entered the room. The operators were not allowed to present in the ultrasound room during each other's examinations to remove any possible influence by the second operator when generating the image and measuring the fetal biometry. Each operator was blinded to any pre-existing measurements. All ultrasound examinations were performed in a single room. For BPD a transverse section of the head was used with both the lateral ventricles symmetrical in view with a horizontal midline. The BPD measurement was made perpendicular to the midline outer to outer at the widest point. Finally, the FL was measured by including only the femoral diaphysis length, excluding the hypoechogenic cartilaginous structures at either end of the femur. All measurements were obtained at the appropriate levels described elsewhere. The fetal biometrics calculations were made by using the formula based on Hadlock descriptions [6, 7]. Ultrasound examinations were performed transabdominally using a commercially available ultrasound system equipped with a four- to seven-MHz curved, high frequency, curved array transducer for all attendants.

During the same visit, after obtaining fetal biometric measurements, patients were sent to the biochemistry laboratory for prenatal maternal serum screening test. Blood samples were obtained for measurements of AFP, uE3, and free β-hCG concentrations. In serum screening using maternal serum biochemical markers, the measured concentration of the markers was converted into multiple of median (MoM) of unaffected pregnancies at the same gestation. The STMSS was performed according to calculation of a risk based on maternal age, previous history of Down's syndrome, measurements of biochemical markers obtained from maternal serum samples, and fetal biometric measurements. The risk calculations of screening were performed according to the each operator's biometric measurements separately. This is the reason why two separate risk was calculated for each patient and two separate prenatal screening results for each patient were obtained. The resultant risks were compared with a threshold and, in cases where the risk was at or above the threshold, the test was deemed screen-positive. Otherwise, it was deemed screen-negative. The current policy in Turkey is to use a risk threshold of one in 250 for risk assessment in the second trimester of pregnancy. Interobserver variability in measurements of fetal biometrics and the effect of this interobserver variability on the screening results were assessed.

Table 1. — *Clinical characteristics of the included patients (n=60).*

Characteristics	Mean ± SD
Age, years	26.68 ± 5.90
BMI, kg/m^2	25.85 ± 5.01
Gravidity, n	2.15 ± 1.03
Parity, n	1.07 ± 0.97
Gestational age, days	118.17 ± 5.57

Statistical analysis

Collected data were analyzed by Statistical Package for Social Sciences version 15.0. Continuous variables were expressed as mean ± standard deviation. Interobserver comparisons were done by reliability tests (Cronbach alpha and intraclass correlation coefficients). Two-way mixed effects model where people effects are random and measures effects are fixed (absolute agreement definition) was used. Two-tailed *p*-value less than 0.05 was accepted to be statistically significant.

Results

A total of 60 patients who underwent a second trimester ultrasound scan for routine STMSS were included in the study. All pregnancies were examined by two operators. Interobserver variability of BPD and FL was assessed in the 60 patients with measurements performed by both observers. Maternal age ranged from 18 to 41 years and the mean age of patients was 26.68 ± 5.90 years. All patients were Caucasian. All serum parameters and risk calculations were obtained successfully. The mean gestational age was 118.17 ± 5.57 days. Some demographic and clinical characteristics of patients are illustrated in Table 1.

Descriptive statistics for the fetal biometrics and screening test results obtained by each observer are presented in Table 2. When the resultant risks were compared with the threshold, according to the biometric measurements made by radiologist in two cases, the test was found screen-positive for trisomy 21. According to the biometric measurements made by obstetrician, in three cases the test was deemed screen-positive for trisomy 21. In only one case,

Table 2. — *Descriptive statistics for BPD, FL, NTD, Tr 21, and Tr 13/18 (n=60).*

Parameter	Observer	Mean	SD	Median	Minimum	Maximum	*p*
BDP (day)	1 (Rad)	120.90	6.31	120.0	107	139	0.017*
	2 (Obs)	119.80	5.75	119.0	109	137	
FL (day)	1 (Rad)	119.41	7.93	118.0	106	139	< 0.001*
	2 (Obs)	116.53	6.64	115.5	104	137	
NTD (1/risk)	1 (Rad)	21661.07	18076.59	17150.0	16	65900	0.466*
	2 (Obs)	22659.78	20419.76	15477.0	11	64700	
Tr 21 (1/risk)	1 (Rad)	16111.43	18357.59	5630.0	39	50000	0.031**
	2 (Obs)	19511.22	19908.15	7745.0	39	50000	
Tr 13/18 (1/risk)	1 (Rad)	83638.83	31374.65	99000	798	99000	0.729**
	2 (Obs)	83836.36	31610.22	99000	800	99000	

BDP: biparietal diameter; FL: femur length; NTD: neural tube defect; Tr 21: trisomy 21; Tr 13/18: trisomy13/18. *Paired Samples T-test. **Wilcoxon Signed Ranks Test.

Table 3. — *Interobserver and reliability coefficients with 95% confidence intervals for BPD, FL, NDT, Tr 21, and Tr 13/18 (n=60).*

	ICC	95% Confidence interval	*p*
BDP	0.904	0.832 - 0.944	< 0.001
FL	0.888	0.616 - 0.953	< 0.001
NTD	0.920	0.866 - 0.952	< 0.001
Tr 21	0.887	0.807 - 0.934	< 0.001
Tr 13/18	0.999	0.998 - 0.999	< 0.001

ICC: intraclass correlation coefficients; BDP: biparietal diameter; FL: femur length; NTD: neural tube defect; Tr 21: trisomy 21; Tr 13/18: trisomy 13/18.

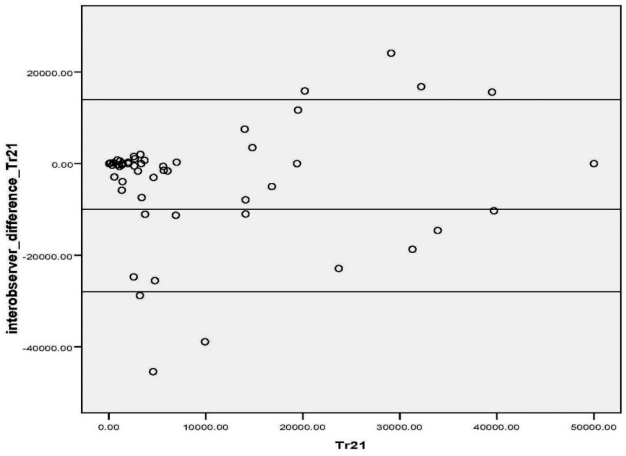

Figure 1. — Scatter plot of the interobserver differences plotted against the mean of the two measurements.

screen-negative test result according to measurements of radiologist returned to negative with measurements of obstetrician. When the resultant risks according to biometric measurements obtained by the two operators were evaluated, in all cases the test was found screen-negative for trisomy 13 and 18.

Interobserver reliability coefficients with 95% confidence intervals for BPD and FL were 0.904 (0.832-0.944, $p < 0.001$) and 0.888 (0.616-0.953, $p < 0.001$), respectively. Interobserver reliability coefficients for trisomy 21, trisomy 13/18, and NTD were 0.887 (0.807-0.934, $p < 0.0001$), 0.999 (0.998-0.999, $p < 0.0001$), and 0.920 (0.866-0.952, $p < 0.0001$), respectively. Interobserver and reliability coefficients with 95% confidence intervals for fetal biometrics and screening test results are summarized in Table 3. Scatter plot of inter-operator differences in second trimester serum screening for trisomy 21 are presented in Figure 1. Limits of the agreement were plotted on the graph.

Interobserver reliability and agreement in second trimester BPD and FL measurements were quite high. Due to high reliability and agreement, resultant risks according to biometric measurements obtained by the two operators were quite similar.

Discussion

The detection rate of Down's Syndrome from STMSS has been reported to range from 60%-75% with a false-positivity rate of 5% [8, 9]. The poor interobserver reliability may be one of the explanation of these conflicting results. This may adversely affect the detection rate of the prenatal screening test.

The present study demonstrated that interobserver reliability coefficients with 95% confidence intervals for BPD and FL were 0.904 (0.832-0.944, $p < 0.001$) and 0.888 (0.616-0.953, $p < 0.001$), respectively. Interobserver reliability coefficients for trisomy 21, trisomy 13/18, and NTD were 0.887 (0.807-0.934, $p < 0.0001$), 0.999 (0.998-0.999, $p < 0.0001$), and 0.920 (0.866-0.952, $p < 0.0001$), respectively. This indicates that interobserver reliability and agreement in second trimester BPD and FL measurements were quite high, thus demonstrating acceptable high reliability and agreement.

In literature, there are many clinical studies investigating

interobserver variability of fetal biometrics by using 2D and 3D ultrasound. These clinical studies reported that fetal biometric measurements obtained by different operators using both 2D and 3D ultrasound were reproducible and revealed a good agreement [10-12]. The present results were in agreement with these prior findings of relevant studies. Another study conducted by Souka *et al.* showed that BPD is a highly reproducible biometric measurement. The ICCs of this study for BPD was 0.968 (0.953-0.978, $p < 0.001$) [13]. In the present study, the ICCs for BPD was 0.904 (0.832-0.944, $p < 0.001$) and is similar to finding observed by Souka *et al.*

A study of Callis *et al.* illustrated that when using measurement units (mm) to express differences, both intra- and interobserver variability increased with gestational age [14]. The present authors excluded pregnancies between 18-20 weeks of gestation therefore the present results were not affected from interobserver variability of fetal biometric measurements. Additionally, standardization of fetal biometric measurements should be done to improve the uniformity and quality of the data. In literature Sarris *et al.* claimed that even for experienced sonographers, a standardization exercise before starting a study of fetal biometry using multiple sonographers can improve the consistency of the measurements [15]. Concomitantly a study reported that nine local paramedics from four health clinics in rural Bangladesh, with no prior exposure to ultrasonography, were trained to conduct ultrasound examinations for fetal biometry during prenatal visits related to a prenatal intervention trial. They claimed that with intense training, paramedics with no prior exposure to ultrasonography can provide accurate and precise measures of fetal biometry [16]. Two sonographers of the present study, S.O. and S.T. are experienced on these measurements and they have been working in the same center for a long period.

An estimated risk is calculated and adjusted for the maternal age; maternal weight and ethnicity [17]. The fetal gestational age is another important parameter that should be used in adjustments [18]. Each of these factors affects the levels of the substances being measured and the interpretation of the screening results. Knowing gestational age (GA) accurately is essential for optimal prenatal screening test result and prenatal care. The accurate interpretation of prenatal screening test results rest on the accurate estimated GA, as do prenatal counseling and invasive interventions for high risk pregnant women, as well as the avoidance of unnecessary invasive interventions such as AS and CVS. However, inaccurate estimation of fetal gestational age may result in misinterpretation of the screening test results. As a result, accurate dating of pregnancy is subject interoperator variation. The aim of the study was to investigate effect of interobserver reliability on the STMSS test results. Thereby, the importance of present study is based on avoiding misinterpretation of the screening test results and possibility of unnecessary invasive interventions. In literature many of the authors investigated the effect of interobserver reliability on the first trimester screening [19-22]. This study is the first which investigates interobserver reliability of fetal biometric measurements and its impact on the STMSS test results.

In conclusion, this study demonstrated that fetal biometric measurements for routine prenatal screening by abdominal ultrasound are highly reliable. Interobserver reliability and agreement in second trimester BPD and FL measurements were quite high. Due to high reliability and agreement, resultant risks according to biometric measurements obtained by the two operators were quite similar. The ultrasound measurements of fetal biometry obtained by experienced operators can be safely used for second trimester maternal serum screening.

References

[1] “Birth defect surveillance data from selected states, 1989 - 1996”. *Teratology*, 2000, 61, 86.

[2] Cuckle H.S., Wald N.J., Thompson S.G.: “Estimating a woman’s risk of having a pregnancy associated with Down's syndrome using her age and serum alpha-fetoprotein level”. *Br. J. Obstet. Gynaecol.*, 1987, *94*, 387.

[3] Alldred S.K., Deeks J.J., Guo B., Neilson J.F., Alfirevic Z.: Second trimester serum tests for Down’s Syndrome screening. *Cochrane Database Syst. Review,* 2000, CD009925.

[4] Alfirevic Z., Neilson J.P.: “Antenatal screening for Down’s syndrome”. *BMJ,* 2004, *329*, 811.

[5] Mongelli M., Wilcox M., Gardosi J.: Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates. *Am. J. Obstet. Gynecol*., 1996, *174*, 278.

[6] Shepard M., Filly R.A.: “A standardized plane for biparietal diameter measurement”. *J. Ultrasound. Med.*, 1982, *1*, 145.

[7] Hadlock F.P., Harrist R.B., Deter R.L., Park S.K.: “Fetal femur length as a predictor of menstrual age: sonographically measured”. *AJR. Am. J. Roentgenol.,* 1982, *138*, 875.

[8] Haddow J.E., Palomaki G.E., Knight G.J., Williams J., Pulkkinen A., Canick J.A., *et al.*: Prenatal screening for Down’s syndrome with use of maternal serum markers. *N. Engl. J. Med.*, 1992, *327*, 588.

[9] Burton B.K., Prins G.S., Verp M.S.: “A prospective trial of prenatal screening for Down syndrome by means of maternal serum alpha-fetoprotein, human chorionic gonadotropin, and unconjugated estriol”. *Am. J. Obstet. Gynecol*., 1993, *169*, 526.

[10] Yang F., Leung K.Y., Lee Y.P., Chan H.Y., Tang M.H.: “Fetal biometry by an inexperienced operator using two- and three-dimensional ultrasound”. *Ultrasound. Obstet. Gynecol*., 2010, *35*, 566.

[11] Lima J.C., Miyague A.H., Filho F.M., Nastri C.O., Martins W.P.: “Biometry and fetal weight estimation by two-dimensional and three-dimensional ultrasonography: an intraobserver and interobserver reliability and agreement study”. *Ultrasound. Obstet. Gynecol.*, 2012, *40*, 186.

[12] Chan L.W., Fung T.Y., Leung T.Y., Sahota D.S., Lau T.K.: “Volumetric (3D) imaging reduces inter- and intraobserver variation of fetal biometry measurements”. *Ultrasound. Obstet. Gynecol.,* 2009, *33*, 447.

[13] Souka A.P., Pilalis A., Papastefanou I., Salamalekis G., Kassanos D.: “Reproducibility study of crown-rump length and biparietal diameter measurements in the first trimester”. *Prenat. Diagn*., 2012, *32*, 1158.

[14] 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.

[15] Sarris I., Ioannou C., Dighe M., Mitidieri A., Oberto M., Qingqing W., *et al.*: Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements". *Ultrasound. Obstet. Gynecol*., 2011, *38*, 681.

[16] Neufeld L.M., Wagatsuma Y., Hussain R., Begum M., Frongillo E.A.: "Measurement error for ultrasound fetal biometry performed by paramedics in rural Bangladesh". *Ultrasound. Obstet. Gynecol*., 2009, *34*, 387.

[17] Wald N.J., Kennard A., Hackshaw A., McGuire A.: "Antenatal screening for Down's syndrome". *Health. Technol. Assess.*, 1998, *2*, 1.

[18] Wald N.J., Cuckle H., Brock J.H., Peto R., Polani P.E., Woodford F.P.: "Maternal serum-alpha-fetoprotein measurement in antenatal screening for anencephaly and spina bifida in early pregnancy. Report of U.K. collaborative study on alpha-fetoprotein in relation to neural-tube defects". *Lancet,* 1977, *1*, 1323.

[19] Kagan K.O., Hoopmann M., Baker A., Huebner M., Abele H., Wright D.: "Impact of bias in crown-rump length measurement at first-trimester screening for trisomy 21". *Ultrasound. Obstet. Gynecol*., 2012, *40*, 135.

[20] Abele H., Hoopmann M., Wright D., Hoffmann-Poell B., Huettelmaier M., Pintoffl K., *et al.*: "Intra- and interoperator reliability of manual and semi-automated measurement of fetal nuchal translucency by sonographers with different levels of experience". *Ultrasound. Obstet. Gynecol*., 2010, *36*, 417.

[21] Kagan K.O., Abele H., Yazdi B., Böer B., Pintoffl K., Wright D., *et al.*: Intraoperator and interoperator repeatability of manual and semi-automated measurement of increased fetal nuchal translucency according to the operator's experience". *Prenat. Diagn*., 2011, *31*, 1229.

[22] Newey V.R., Nassiri D.K., Bhide A., Thilaganathan B.: "Nuchal translucency thickness measurement: repeatability using a virtual ultrasound scanner". *Ultrasound. Obstet. Gynecol*., 2003, *21*, 596.

Corresponding Author:
Ö. ÖZDAMAR, M.D., Ph.D.
Gölcük Military Hospital
Department of Obstetrics and Gynecology
Golcuk Naval Base
41650 Gölcük, Kocaeli (Turkey)
e-mail: ozkan_ozdamar@yahoo.com