A prospective study to assess the clinical impact of interobserver reliability of sonographic measurements of fetal nuchal translucency and crown-rump length on combined first-trimester screening

Huseyin Aksoy,[1] Sezin Ozyurt,[2] Ulku Aksoy,[3] Erol Mutlu,[4] Sadan Tutus,[5]
Mustafa Alparslan Babayigit,[6] Gokhan Acmaz,[2] Ozge Idem Karadag,[7] Burak Yucel[7]

[1]Department of Obstetrics and Gynecology, Kayseri Military Hospital, Kayseri, Turkey

[2]Department of Obstetrics and Gynecology, Kayseri Training and Research Hospital, Kayseri, Turkey

[3]Department of Obstetrics and Gynecology, Kayseri Memorial Hospital, Kayseri, Turkey

[4]Department of Obstetrics and Gynecology, Corlu Military Hospital, Tekirdag, Turkey

[5]Department of Radiology, Kayseri Training and Research Hospital, Kayseri, Turkey

[6]Department of Public Health and Epidemiology, Gulhane Military Faculty of Medicine, Ankara, Turkey

[7]Department of Obstetrics and Gynecology, Kayseri Acibadem Hospital, Kayseri, Turkey

ABSTRACT

OBJECTIVE: To examine the interobserver variability for fetal biometric parameters and to focus on how this variability affects the combined first-trimester screening test results.

METHODS: Sixty-four singleton pregnancies who were scheduled for combined first-trimester screening 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 individual ultrasonographic fetal measurements. Interobserver variability in measurements of fetal nuchal translucency (NT) and crown-rump length (CRL) and the effect of this interobserver variability on the screening results were assessed.

RESULTS: Interobserver reliabilities for NT and CRL were 0.687 and 0.945 (p<0.001), respectively. Interobserver reliability coefficients for trisomy 21 and trisomy 13/18 were 0.945 and 0.963 (p<0.001), respectively.

CONCLUSION: Our study demonstrated quite high interobserver reliability for CRL measurement, while interobserver agreement for NT was lower. Despite this lower reliability and agreement for NT measurement, interobserver reliability for chromosomal abnormalities was quite high.

Keywords: First-trimester screening; interobserver reliability; nuchal translucency; trisomy 21.

Trisomy 21 (Down syndrome) is the most common viable chromosomal abnormality, affecting approximately 1 in 800 live-born babies [1, 2]. It is the commonest cause of mental retardation, and it is also associated with many congenital malformations. Prenatal screening for Down's syndrome is now offered routinely in many countries including Turkey. First-trimester screening is the predominant noninvasive method of prenatal screening for trisomy 21 (Down syndrome), trisomy 18 (Edwards Syndrome) and other type of chromosome abnormality in many countries. Combined first-trimester screening based on maternal age, fetal nuchal translucency (NT) thickness and maternal serum markers (β-hCG and PAPP-A) at 11–14 weeks of gestation can detect about 81% of the fetuses with trisomy 21 and other major aneuploidies with a false-positive rate of 5 percent [3, 4, 5]. The single most effective marker in prenatal screening for trisomy 21 is fetal NT. Fetal NT alone can identify about 80% of the fetuses with trisomy 21 and other major aneuploidies with a false-positive rate of 5% [4, 6].

The results of maternal serum markers and fetal NT thickness are also combined with the maternal age, maternal weight, ethnicity and gestational age in order to assess probabilities of major potential chromosomal aneuploidies. This prenatal screening test allows an estimate of the risk of a pregnancy being affected and provides parents with information so as to guide them to make decision about definitive invasive testing [7]. Definitive invasive tests [amniocentesis (AS) and chorionic villus sampling (CVS)] allow the physicians to make a diagnosis of Down's syndrome and other major aneuploidies but carry the risk of serious complications including miscarriage. However, this detection rate can only be achieved consistently if fetal gestational age, NT thickness and the maternal serum markers are measured correctly. While the concentrations of β-hCG and PAPP-A can be determined objectively, estimation of gestational age and NT are subjective processes. Both processes are prone to inter-operator variations, but fetal NT measurements are more prone to these operator dependent variations. This issue has been addressed by several studies investigating the repeatability of NT measurements [8,

9, 10, 11, 12, 13, 14, 15]. The majority of previous studies on this issue have focused on reproducibility of the NT measurements but comprehensive studies investigating the repeatability of the CRL measurements are limited. Accurate pregnancy dating and fetal NT measurements offer important clinical benefits in terms of prenatal screening [16]. Inaccurate determination of fetal gestational age and NT thickness affects prenatal screening performance [17]. Crown-rump length (CRL) is an established measurement for dating the pregnancy [18, 19]. The pregnancy must be dated accurately because errors will affect the assigned risk, causing false-negative and false-positive results.

Despite the importance of CRL and NT measurements in prenatal screening, and dating the pregnancy, few studies have been published, so far [8, 9, 10, 11, 12, 13, 14, 15]. However, there is no comprehensive study specific to the possible impact of interobserver variability of CRL and NT measurements on screening. The aim of this study is to examine the interobserver reliability for CRL and NT measurements and to focus on how this reliability affects the combined first-trimester screening test results. In this study we compared the interoperator reliability of the CRL and NT measurements performed by an experienced radiologist and an obstetrician in cases of routine combined first-trimester screening.

MATERIALS AND METHODS

A prospective clinical study was performed at our Obstetrics and Gynecology Clinic of Kayseri Education and Research Hospital, a tertiary referral centre in Turkey between November 2013 and August 2014. The study was approved by the institutional ethics committee and performed in accordance with the Helsinki Declaration of the World Medical Association. All participants signed an informed consent form regarding participation in the present study. A total of 64 singleton pregnants in their gestational age between 11 and 14 weeks (77–98 days of gestation) who were scheduled for combined first-trimester screening and admitted to our obstetrics clinic for routine prenatal care were inves-

tigated. Prenatal evaluation before first-trimester screening was consistent with the protocol of the clinic and included comprehensive medical and obstetric examination along with obstetric ultrasound to determine the gestational age (CRL) as well as to exclude any other pelvic and obstetric pathologies.

Inclusion criteria for this study were as follows: age 18 years or older; acceptance of combined first-trimester screening 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 7 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 can affect test results, multiple pregnancies, and known fetal congenital anomalies were excluded from the study. Moreover, pregnancies under 11 weeks' of gestation and pregnancies over 14 weeks' of gestation were also excluded from the study.

Two experienced sonographers (one experienced radiologist and one experienced obstetrician with an average knowledge in obstetric ultrasound) performed all examinations at the same visit. Fetal CRL and NT measurements were obtained during each ultrasonographic examination using the same ultrasound machine in cases of routine first-trimester screening. Each measurement was performed once by each operator. The initial measurement was recorded by the first sonographer who is an experienced obstetrician (U.A.). Subsequently, a second sonographer who is an experienced radiologist (S.T.), blinded to the results of the first sonographer, performed the same measurements. Numeric displays on the screen were always deleted from the ultrasound screen so that the operators were blinded to the results obtained by the other operators. The operators were not allowed to present in the ultrasound room during each other's examinations so as to discard any possible influence exerted by the second operator when generating the image and measuring the fetal CRL and NT. Each operator was blinded to any preexisting measurements. All ultrasound examinations were performed in a single room.

Standard ultrasound planes for fetal measurements were used as described previously [20, 21]. Briefly, CRL was measured in the mid-sagittal longitudinal plane in the absence of fetal movements while the fetus was in a neutral position. The measurement was taken along a straight line extending from the top of the head to the bottom of the caudal end of the fetus. Manual measurement of NT was performed according to the guidelines of the Fetal Medicine Foundation (2011). The maximum thickness of the nuchal translucency was measured from a sagittal plane of the fetal profile in supine position, using a magnified frozen image image with fetus filling at least 75% of the viewable screen area so that each increment in the distance between the calipers was 0.1 mm. After the fetal amnion was clearly observed and discriminated from the fetal skin; the measuring calipers were placed to obtain an inner diameter and maximum NT measurements. All ultrasouographic examinations were performed transabdominally using a commercially available ultrasound system (Toshiba Xario, Toshiba Medical Systems Corporation, Japan) equipped with a 4–7 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 combined first-trimester screening tests. Blood samples were obtained for measurements of β-hCG and PAPP-A concentrations. In the first-trimester prenatal screening the measurements of concentrations of maternal serum biochemical markers were converted into Multiple of Median (MoM) of unaffected pregnancies at the same gestation. The combined first-trimester screening was performed according to calculation of a risk based on maternal age, previous history of Down's syndrome, measurements of biochemical markers of maternal serum samples, and fetal CRL and NT measurements. The risk calculations of screening were performed according to the each sonographer's individual ultrasonographic measurements. That's why, two separate risks were calculated for each patient and two separate prenatal screening test results for each patient were obtained. The resultant risks were compared with a threshold and, in cases where

the risk is at or above the threshold, the test was deemed screen-positive. Otherwise, it was deemed as screen-negative. The current policy in Turkey is to use a risk threshold of 1 in 250 for risk assessment in the second trimester of pregnancy. Interobserver variability in measurements of fetal CRL and NT and the effect of this interobserver variability on the screening results were assessed.

Statistical analysis

Collected data were analyzed by Statistical Package for Social Sciences version 15.0 (SPSS Inc., Chicago, IL, USA). Continuous variables were expressed as mean±standard deviation. Interobserver comparisons were done by reliability tests (Cronbach's alpha and intraclass correlation coefficients). Two-way mixed effects model was used in cases where people effects were random and measured effects were fixed (Absolute Agreement Definition) Two-tailed p value less than 0.05 was accepted to be statistically significant.

RESULTS

A total of 64 patients who underwent a first trimester ultrasound scan for routine combined first-trimester

Table 1. Some clinical characteristics of the group (n=64)

	Mean±SD
Age (years)	26.91±5.22
BMI (kg/m²)	26.41±4.63
Gravidity	2.14±1.20
Parity	1.02±.96
Gestational age (day)	87.72±5.27

screening were included in the study. All pregnancies were examined by two operators. Interobserver variability of CRL and NT was assessed in the 64 patients with measurements performed by both observers. Maternal age ranged from 18 to 39 years and the mean age of the patients was 26.91±5.22 years. All patients were Caucasian. All serum parameters and risk calculations were obtained successfully. The mean gestational age was 87.72±5.27 days. Some demographic and clinical characteristics of the patients are illustrated in Table 1.

Descriptive statistics for the fetal CRL, NT and screening test results obtained by each observer are

Table 2. Descriptive statistics for CRL, NT, Tr 21 and Tr 13/18 (n=64)

Parameter	Observer	Mean	SD	Median	Minimum	Maximum	p
CRL							
	1 (Rad)	88.59	4.68	89.0	77	97	0.699**
	2 (Obs)	88.48	5.15	88.0	77	98	
NT							
	1 (Rad)	1.39	0.41	1.4	0.70	3.41	0.246***
	2 (Obs)	1.35	0.36	1.4	0.80	3.20	
Tr 21							
	1 (Rad)	19590.42	18445.06	11400.0	27	50000	0.696***
	2 (Obs)	18849.04	18202.07	9685.0	22	50000	
Tr 13/18							
	1 (Rad)	93008.75	20142.33	99000.0	1760	99000	0.401***
	2 (Obs)	92175.62	20233.77	99000.0	8240	99000	

CRL: Crown-rump length; NT: Nuchal translucency; Tr 21: Trisomy 21; Tr 13/18: Trisomy 13/18; **Paired Samples t-test; ***Wilcoxon Signed-Rank Test; p values <0.05.

TABLE 3. Interobserver and reliability coefficients with 95% confidence intervals for CRL, NT, Tr 21 and Tr 13/18 (n=64)
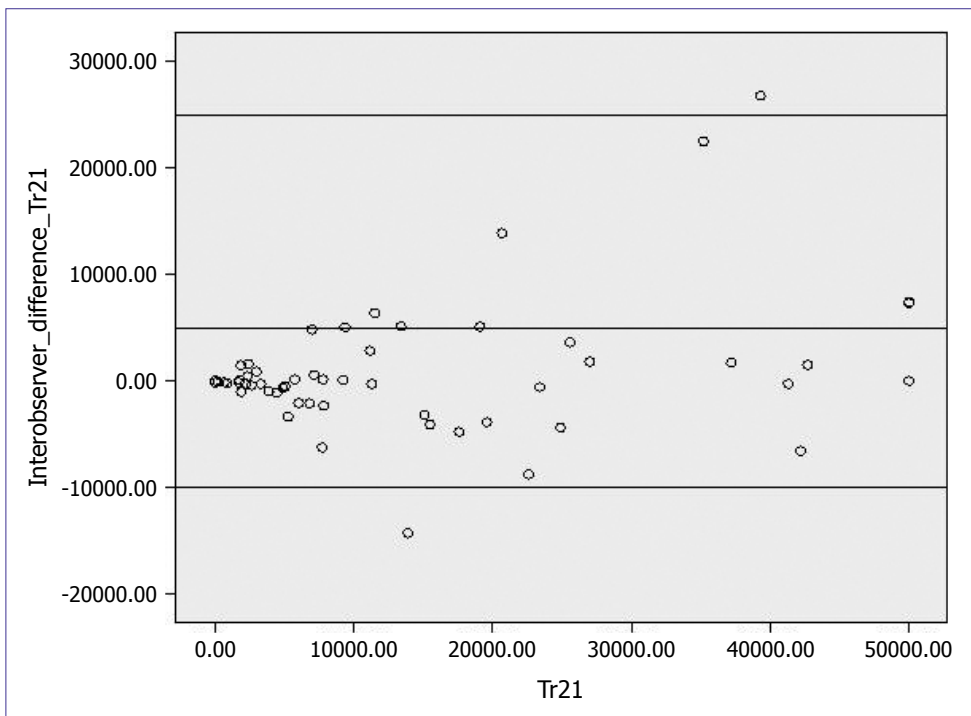
	Interobserver reliability coefficients**	95% confidence interval	p
CRL	0.945	0.910–0.967	<0.0001
NT	0.687	0.485–0.810	<0.0001
Tr 21	0.945	0.957–0.984	<0.0001
Tr 13/18	0.963	0.939–0.977	<0.0001

CRL: Crown-rump length; NT: Nuchal translucency; Tr 21: Trisomy 21; Tr 13/18: Trisomy 13/18; **Intraclass correlation coefficient test.

presented in Table 2. When the resultant risks were compared with the threshold values, according to the CRL and NT measurements made by the obstetrician in two cases, the test was found screen-positive for trisomy 21. According to the CRL and NT measurements made by the radiologist, in three cases the test was deemed screen-positive for trisomy 21. In only one case, screen-negative test result according to the measurements of the obstetrician returned to positive when measured by the radiologist. When the resultant risks according to CRL and NT 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 CRL and NT were 0.945 (0.910–0.967, p<0.001) and 0.687 (0.485–0.810, p<0.001), respectively. Interobserver reliability coefficients for trisomy 21 and trisomy 13/18 were 0.945 (0.957–0.984, p<0.0001) and 0.963 (0.939–0.977, p<0.0001), respectively. Interobserver and reliability coefficients with 95% confidence intervals for fetal CRL, NT and screening test results are summarized in Table 3.

FIGURE 1. Scatter plot of the inter-operator differences plotted against the mean of the two measurements.

Scatter plot diagram of inter-operator differences in combined first-trimester screening for trisomy 21 is shown in Figure 1. Limits of the agreement were plotted on the graph.

Interobserver reliability for first trimester CRL measurement was quite high, which reflects a good agreement of the measurements between observers, while interobserver agreement for NT was lower. Despite this lower interobserver reliability and agreement for NT measurement, resultant risks according to CRL and NT measurements obtained by the two operators were quite similar.

DISCUSSION

In this study, we compared the interoperator reliability of the CRL and NT measurements by an experienced radiologist and an obstetrician in cases of routine combined first-trimester screening. We also evaluated the possible effect of this interobserver variability on the first-trimester screening results. Interobserver reliability for the first trimester CRL measurement was quite high, which reflects a good agreement of the measurements between observers, while interobserver reliability for NT measurement was lower. Despite this lower interobserver reliability for NT measurement, which reflects a moderate agreement of the measurements between observers, resultant risks according to CRL and NT measurements obtained by the two operators were quite similar.

Sonographic measurement is an established screening method for Down syndrome and other major aneuploidies in the first trimester of pregnancy. Accurate assessment of NT thickness in the first-trimester pregnancies can be challenging [22, 23]. CRL is used to standardize the fetal and biochemical parameters of screening [24]. Accurate estimation of fetal NT and CRL measurements by ultrasonography is an integral part of the combined first-trimester screening. Therefore, the screening performance is directly depends on the accuracy and precision of these sonographic measurements [17]. Unfortunately, fetal NT and CRL measurements are often subjective processes secondary to operator's experience and ultrasound technology.

Interobserver variability in fetal NT and CRL measurements is the most important issue affecting the screening performance. This issue of repeatability has been addressed by several studies [8, 9, 10, 11, 12, 13, 14, 15]. Although most of the previous studies on this issue have focused on repeatability of NT measurements, few studies have addressed the issue of reproducibility of CRL. Surprisingly, despite the pivotal importance of CRL and NT measurements in the first-trimester prenatal screening, there is no comprehensive study specific to the possible impact of interobserver variability of CRL and NT measurements on combined first-trimester screening test results.

This study showed that interobserver reliability coefficients with 95% confidence intervals for NT and CRL were 0.687 (0.485–0.810, p<0.001) and 0.945 (0.910–0.967, p<0.001), respectively. Interobserver reliability coefficients for Trisomy 21 and Trisomy 13/18 were 0.945 (0.957–0.984, p<0.0001) and 0.963 (0.939–0.977, p<0.0001), respectively. This means that interobserver reliability for the first trimester CRL measurement was quite high, thus demonstrating acceptable high reliability and agreement. However, interobserver reliability and agreement for NT was found to be lower, which reflects a moderate agreement of the measurements between observers. First-trimester screening test results were not affected by this moderate reliability of NT measurements. In the literature, previous clinical studies investigating interobserver repeatability of CRL measurements reported that fetal CRL measurements recorded by different operators were reproducible and showed a good agreement [13, 14, 15]. Our results were in agreement with these prior findings of relevant studies. In 2011, Pexsters et al. evaluated the reproducibility of CRL in 54 singleton pregnancies with a gestational age between 6 and 9 weeks using transvaginal sonography [13]. They found that CRL was a highly reproducible biometric measurement for dating the gestational age and the intraclass coefficient (ICC) of this study for CRL was 0.993. In our study, the ICC for CRL was 0.945 (0.910–0.967, p<0.001) and it was similar to that observed by Pexsters et al. Verburg et al. performed a prospective analysis

of 21 singleton pregnancies with a gestational age between 9 and 14 weeks and demonstrated good reproducibility of fetal CRL measurement by abdominal ultrasound. In this prospective study, interobserver ICC for CRL was 0.995 (0.998–0.982, p<0.001) [14]. Recently, Souka et al. conducted a prospective observational study on 592 pregnancies between 8 and 14 gestational weeks. Authors reported that CRL was a highly reproducible parameter of fetal biometry and the ICC for CRL's interobserver variability was 0.968 (0.952–0.978, p<0.001) [15]. Although almost all studies evaluating the reproducibility of CRL measurements demonstrated good reproducibility, any error in CRL measurements may have an impact on combined first-trimester screening. Salomon et al. performed a simulation study using a simulation model to evaluate the impact of error in CRL measurements in cases of sequential combined screening for Down syndrome [24]. After more than 3200 simulated cases were analyzed, they reported that Down syndrome screening might be highly sensitive to errors in CRL measurements. Therefore, quality control of CRL measurements should be performed together with quality control of NT measurements in order to provide an optimal screening performance and highest standard of care. This study of Salomon et al. was different from our study in terms of study design. In our study, risk calculations were performed based on the real operator's actual ultrasonographic measurements separately, but Salomon et al. used a simulation model for the measurements. Another difference between studies was the difference of the serum markers used. Salomon et al. used early second-trimester maternal serum markers (hCG and alpha-fetoprotein (AFP). We used first-trimester maternal serum markers (hCG and PAPP-A).

Hitherto, most of the previous studies investigating the repeatability of fetal NT measurements have indicated that fetal NT measurements were highly reproducible and showed a good agreement between operators [8, 9, 10, 11, 12]. Pajkrt et al. performed a prospective analysis of 161 women with singleton pregnancies at their 10–14 weeks of gestation. In this prospective study, it was found that fetal NT measurements are reproducible, as long

as they are performed by operators well-trained in NT measurement techniques [8]. In this study, the inter-operator repeatability showed a variation in ICC from 0.51 to 0.95 for NT measurements. The inter-operator repeatability showed ICCs varying from 0.74 to 0.95 for pairs of experienced operators and 0.68 to 0.74 for pairs of inexperienced operators. This finding was consistent with the ICC of 0.687 (0.485–0.810, p<0.001) for NT reported in our study. In another study, Pandya et al. indicated that NT measurements are highly reproducible when the NT thickness is measured by well-trained operators [9]. Although NT measurements have been found highly reproducible by many researchers, there are studies demonstrating that NT measurement is quite difficult and repeatability of the measurement is low. Roberts et al. conducted a prospective observational study to assess the reproducibility of NT measurements in 1004 women [25]. The authors addressed the difficulty with measuring NT and claimed that the repeatability of the measurement was poor. In 2005, Malone et al. using the data of The First And Second Trimester Evaluation of Risk (FASTER) Trial, which is a National Institute of Child Health and Human Development-sponsored prospective study providing the opportunity to evaluate NT measurement quality assessment, reported that obtaining consistent, accurate and reproducible NT measurements is quite difficult [23]. In this comprehensive study, Malone et al. analyzed 38167 unselected patients with singleton pregnancies with a gestational age between 10+3 and 13+6 weeks from 1999 to 2001. In this comprehensive study, well-trained operators, in the carefully regulated setting of a prospective research trial, could not perform adequate NT measurements in 7.4% of the cases. D'Alton et al. using the data of the FASTER Trial have also confirmed these results in 2009 [22]. It was found that despite initial training and continuous image review, changes in NT measurements occur over time. Therefore to maintain accuracy in NT measurements, ongoing quality assessment is mandatory. These studies using the data of the FASTER Trial have demonstrated the importance of ongoing quality assessment. Large-scale epidemiological monitoring data

indicated that intense and unified training for all sonographers and continuous image review are not enough to prevent changes in NT measurements which occur over time for unexplained reasons [22, 23]. Outside the strictly controlled setting of a clinical research trial, the variability in NT measurements may be even more marked. Therefore, given the significance of the decisions being made on the basis of these measurements, interobserver variability must be taken into consideration.

Measuring fetal NT and CRL accurately is essential for optimal combined first-trimester screening performance and prenatal care. Prenatal counseling and invasive interventions for high risk pregnant women as well as the avoidance of unnecessary invasive interventions such as AS and CVS depend on the accuracy of NT and CRL measurements. However, inaccurate measurements may result in misinterpretation of the screening test results. The aim of the study was to investigate effect of interobserver reliability on the combined first-trimester screening test results. Thereby, the importance of the present study is that it discarded the possibility of misinterpretation of the screening test results and unnecessary invasive interventions. This study is the first study which investigated interobserver reliability of fetal NT and CRL measurements in combination and its impact on the combined first-trimester test results.

There are some limitations to our study. Lack of the intraobserver variability for ultrasonographic measurements is the potential limitation of the study. Ideally, the study should have intraobserver variability for NT and CRL measurements. The absence of time required for ultrasound measurements is another limitation of this study. In this current study time required for ultrasound measurements was not measured, thus comparison between groups regarding the time required for ultrasound examination could not be made. Data collection from a single obstetrics clinic is another important limitation of the study. Therefore, generalizability of our findings is limited. Small size of the study sample is another weakness of the study.

In conclusion, this study has demonstrated acceptable high interobserver reliability and agreement for fetal CRL measurements in cases of routine combined first-trimester screening. However, interobserver reliability and agreement for NT was found to be lower, which reflected a moderate agreement of the measurements between observers. Resultant risks according to measurements obtained by the two operators were quite similar. Despite this moderate reliability and agreement for NT measurements, interobserver reliability for the assessment of chromosomal abnormalities was quite high. Ultrasonographic measurements of fetal NT and CRL obtained by experienced operators canbe use for combined first-trimester screening.

Acknowledgements

REFERENCES

1. Birth defect surveillance data from selected states, 1989 - 1996. Teratology 2000;61:86–160.

2. Cuckle HS, Wald NJ, Thompson SG. 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;94:387–402.

3. Crossley JA, Aitken DA, Cameron AD, McBride E, Connor JM. Combined ultrasound and biochemical screening for Down's syndrome in the first trimester: a Scottish multicentre study. BJOG 2002;109:667–76.

4. Niemimaa M, Suonpää M, Perheentupa A, Seppälä M, Heinonen S, Laitinen P, et al. Evaluation of first trimester maternal serum and ultrasound screening for Down's syndrome in Eastern and Northern Finland. Eur J Hum Genet 2001;9:404–8.

5. Nicolaides KH. Screening for fetal aneuploidies at 11 to 13 weeks. Prenat Diagn 2011;31:7–15.

6. Kagan KO, Wright D, Valencia C, Maiz N, Nicolaides KH. Screening for trisomies 21, 18 and 13 by maternal age, fetal nuchal translucency, fetal heart rate, free beta-hCG and pregnancy-associated plasma protein-A. Hum Reprod 2008;23:1968–75.

7. Alfirevic Z, Neilson JP. Antenatal screening for Down's syndrome. BMJ 2004;329:811–2.

8. Pajkrt E, Mol BW, Boer K, Drogtrop AP, Bossuyt PM, Bilardo CM. Intra- and interoperator repeatability of the nuchal translu-

cency measurement. Ultrasound Obstet Gynecol 2000;15:297–301.

9. Pandya PP, Altman DG, Brizot ML, Pettersen H, Nicolaides KH. Repeatability of measurement of fetal nuchal translucency thickness. Ultrasound Obstet Gynecol 1995;5:334–7.

10. Newey VR, Nassiri DK, Bhide A, Thilaganathan B. Nuchal translucency thickness measurement: repeatability using a virtual ultrasound scanner. Ultrasound Obstet Gynecol 2003;21:596–601.

11. 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;36:417–22.

12. Kagan KO, 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;31:1229–33.

13. Pexsters A, Luts J, Van Schoubroeck D, Bottomley C, Van Calster B, Van Huffel S, et al. Clinical implications of intra- and interobserver reproducibility of transvaginal sonographic measurement of gestational sac and crown-rump length at 6-9 weeks' gestation. Ultrasound Obstet Gynecol 2011;38:510–5.

14. Verburg BO, Mulder PG, Hofman A, Jaddoe VW, Witteman JC, Steegers EA. Intra- and interobserver reproducibility study of early fetal growth parameters. Prenat Diagn 2008;28:323–31.

15. Souka AP, 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;32:1158–65.

16. Whitworth M, Bricker L, Neilson JP, Dowswell T. Ultrasound for fetal assessment in early pregnancy. Cochrane Database Syst Rev 2010;4:CD007058.

17. Schmidt P, Staboulidou I, Elsässer M, Vaske B, Hillemanns P, Scharf A. How imprecise may the measurement of fetal nuchal translucency be without worsening first-trimester screening? Fetal Diagn Ther 2008;24:291–5.

18. Grange G, Pannier E, Goffinet F, Zorn JR, Cabrol D. Dating biometry with crown-rump length, precision of a routine practice. [Article in French] J Gynecol Obstet Biol Reprod (Paris) 2003;32:221–6. [Abstract]

19. Papageorghiou AT, Kennedy SH, Salomon LJ, Ohuma EO, Cheikh Ismail L, Barros FC, et al. International standards for early fetal size and pregnancy dating based on ultrasound measurement of crown-rump length in the first trimester of pregnancy. Ultrasound Obstet Gynecol 2014;44:641–8.

20. Robinson HP, Fleming JE. A critical evaluation of sonar "crown-rump length" measurements. Br J Obstet Gynaecol 1975;82:702–10.

21. Verburg BO, Steegers EA, De Ridder M, Snijders RJ, Smith E, Hofman A, et al. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. Ultrasound Obstet Gynecol 2008;31:388-96.

22. D'Alton ME, Cleary-Goldman J, Lambert-Messerlian G, Ball RH, Nyberg DA, Comstock CH, et al. Maintaining quality assurance for sonographic nuchal translucency measurement: lessons from the FASTER Trial. Ultrasound Obstet Gynecol 2009;33:142–6.

23. Malone FD. Nuchal translucency-based Down syndrome screening: barriers to implementation. Semin Perinatol 2005;29:272–6.

24. Salomon LJ, Bernard M, Amarsy R, Bernard JP, Ville Y. The impact of crown-rump length measurement error on combined Down syndrome screening: a simulation study. Ultrasound Obstet Gynecol 2009;33:506–11.

25. Roberts LJ, Bewley S, Mackinson AM, Rodeck CH. First trimester fetal nuchal translucency: problems with screening the general population. 1. Br J Obstet Gynaecol 1995;102:381–5.