REVIEW ARTICLE

Analysis of Quality of Nuchal Translucency Measurements

[1]Carmen Comas, [1]Monica Echevarria, [1]Ignacio Rodríguez, [2]Joan Sabrià, [1]Bernat Serra

[1]Fetal Medicine Unit, Department of Obstetrics and Gynecology, Institut Universitari Dexeus, Barcelona, Spain

[2]Department of Obstetrics and Gynecology, Hospital Sant Joan de Déu, Barcelona, Spain

Correspondence: Carmen Comas, Fetal Medicine Unit, Department of Obstetrics and Gynecology, Institut Universitari Dexeus Gran Vía Carles III 71-75, Barcelona 08028, Spain, Phone: 0034 932274706, Fax: 0034 934187832, e-mail: carcom@dexeus.com

ABSTRACT

Objective: Quantitative analysis of the quality of nuchal translucency (NT) measurements.

Methods: This is a retrospective single-center study. NT was measured according to the Fetal Medicine Foundation (FMF) criteria by 20 trained obstetricians (October 2003-November 2009). The performance of NT measurements was analyzed with regard to several quality control standards. Accuracy according to experience, professional profile, crown-rump length (CRL) values and FMF certification was statistically tested.

Results: A total of 14,978 NT measurements were assessed. (1) The mean operator-specific median NT-MoM values was 0.98. (2) Mean percentage of cases >95th and <5th percentiles were 5.0 and 4.2% respectively. (3) Logarithmic mean and SD of the NT-MoM values were 0.00 and 0.13 respectively. (4) The DR for trisomy 21 at screening time was 90.7% for a FPR of 6.7% for standard screening strategy. (5) According to cumulative SUM (CUSUM) figures, the performance was more acceptable in FMF-certified operators. Operator experience, exclusive dedication to FM, FMF certification and a range of CRL values > 60 mm had a statistical impact improving these standards.

Conclusion: Overall quality standards show optimal NT measurements in our unit. Experience a dedicated profile to fetal medicine ultrasound, CRL over 60 mm and FMF certification has a significant positive impact on the quality standards.

Keywords: Prenatal screening, Down syndrome, Quality control, CUSUM, Fetal medicine.

INTRODUCTION

In the medical field, and more precisely in prenatal ultrasound, the concept of quality assessment and certification has only recently emerged. In clinical laboratories all tests are regularly subjected to quality controls to determine their reliability.[1] However, although rigid standardization of laboratory measurements has been traditionally well-established,[2,3] clinical measurements, such as ultrasound biometries have only recently been object of interest.[1,4-8] Nuchal translucency (NT) measurement has been shown to be a useful marker for Down syndrome in the late first-trimester in both high-risk and low-risk populations, but only when accompanied by targeted training and ongoing quality assessment.[5] The same principle applies to NT measurements where a tight distribution of normal data increases the performance of the screening algorithm for that condition. A continuous monitoring and scrupulous evaluations of individual performance is likely to improve NT measurement procedure. Although the trisomy 21 detection rate remains a priority, this indicator cannot be used as a reliable marker of quality. Efforts in quality assurance should include more reliable, realistic and individualized indicators.

Recently, several statements have addressed guidelines specific to first-trimester screening for DS, providing the necessary information to ensure accurate and reliable DS screening results given a screening protocol. These policy guidelines include recommendations for screening in the first and/or second trimester, ultrasound markers and dating, as well as analyte combinations. Prenatal screening for DS is implemented in the context of a comprehensive program that coordinates preanalytic (patient and provider information, specimen collection and transportation), analytic (specimen processing and storage, assay methodologies and results, analytic validation) and postanalytic (coordinating information with, and collecting data from sonographers, clinical validity, results reporting, clinical utility) components of the process.[6,9-11]

Individual centers are responsible for meeting the quality assurance standards. Therefore, the aim of this study is to quantitative assess the quality of NT measurements over a period of 6 years in our fetal medicine unit, testing different models for quality control.

METHODS

This is a retrospective single-center study started in October 2003 and ended in November 2009. First-trimester combined screening for Down syndrome (DS) was performed on all pregnant women who attended our department during this period, including maternal age, biochemistry (pregnancy-associated plasma protein-A and free β-human chorionic gonadotrophin) and NT. The maternal serum biochemistry was measured using the Kryptor analyzer (Brahms Diagnostica) in

a one or two-step strategy, between 15 and 85 mm of crown-rump length (CRL) measurements. Scans were carried out by 20 trained obstetricians, transvaginally or transabdominally (depending on fetal and maternal conditions). Voluson 730 and E8 (GE Medical Systems, Zipf, Austria) machines with 5 MHz transabdominal and/or a 8 MHz transvaginal probes. NT was measured according to the FMF criteria. For combined risk calculation, the commercial available SBP-software was used. Cytogenetic study was recommended when combined risk index was higher than 1/270 at screening time. Outcome follow-up was retrieved from our database, obtaining 96% completed cases in this specified period. For each individual operator and overall, the performance of the NT measurements were analyzed with regard to the follow quality standards: (1) Median NT-multiples of the median (MoM) values and 5th and 95th percentiles; (2) percentage of cases below and above the 5th and 95th percentiles respectively; (3) logarithmic mean and logarithmic standard deviation (SD) of the NT-MoM values; (4) performance of the screening test [detection rate (DR) and false-positive rate (FPR)]; and (5) cumulative SUMs (CUSUM) tests. CUSUMs test display a sequential monitoring of a cumulative performance measure over time. After each procedure, the CUSUM tests whether the process under scrutiny is "in control", that is if the process is performing at an acceptable level. For this study, K and H values were set at 0.25 and 9.2 respectively, which corresponds to the statistics suggested in the literature.[12] As a reference for the expected median NT, the Nicolaides formula was used.[13] Operator specific NT measurement accuracy according to the experience [sequential number of measurements (first 100 scans compared to the later scans)], chronological period (2003-2006 versus 2007-2009), professional profile (fetal medicine dedicated or general obstetric profile), CRL values (≤ 60 versus > 60 mm) and FMF certification was statistically tested by U Mann-Whitney (median NT-MoM), ANOVA (logarithmic mean) and random effects ANOVA (SD of the logarithmic NT MoM values). The percentage of cases under and over the 5th and 95th percentiles were compared by Pearson's Chi-square test.

RESULTS

A total of 14,978 NT measurements were reviewed. The mean maternal age was 33 (range 17-45, SD 3.8) years and the mean gestation age at scan was 11 weeks (range 10-13.6). The population included 32% over 35 years. Down syndrome was identified in 54 pregnancies. Seven out of 20 operators (35%) had a professional profile dedicated to fetal medicine, and two of them (10%) were FMF certified at the time of the study. Eight operators remained from 2003 to 2009, which represents

Table 1: Multiples of the median NT values, expressed as median, 5th and 95th percentiles, according to the different criteria (operator, period, CRL values, FMF certification and professional profile)

		MoM-NT			
		Median	*5th percentile*	*95th percentile*	*n*
Operator	Obs 1	1.00	0.68	1.63	2.364
	Obs 2	0.99	0.70	1.45	585
	Obs 3	1.13	0.81	1.60	294
	Obs 4	1.02	0.64	1.63	905
	Obs 5	1.00	0.71	1.40	79
	Obs 6	0.83	0.51	1.32	168
	Obs 7	0.92	0.58	1.39	51
	Obs 8	0.92	0.59	1.42	1.232
	Obs 9	0.87	0.41	1.53	114
	Obs 10	1.03	0.71	1.68	2.517
	Obs 11	0.92	0.60	1.48	2.715
	Obs 12	1.04	0.68	1.69	167
	Obs 13	1.00	0.64	1.50	1.571
	Obs 14	0.95	0.70	1.26	225
	Others < 50 US	0.84	0.48	1.52	100
Period	2003-2006	0.97 *	0.62	1.67	5560
	2007-2009	0.99 *	0.67	1.50	6623
CRL (mm)	≤ 60	0.96 *	0.64	1.61	7.620
	> 60	1.00 *	0.65	1.50	5.467
FMF	Certified	1.00 *	0.67	1.47	1.719
	Noncertified	0.97 *	0.64	1.58	11.368
Profile	Dedicated	0.98	0.65	1.56	11.889
	Nondedicated	0.97	0.57	1.48	1.198
	Overall	0.98	0.64	1.56	13.087

Obs: Observer or operator
US: Ultrasound scans
FMF: Fetal Medicine Foundation
*p < 0.05

a series of 13,840 measurements (6,615 in the first period and 7,225 in the second one). Six operators performed less than 50 NT measurements and were considered all together for statistic analysis. Epidemiological monitoring involved computing five quality measurements, overall and for each operator:

1. The mean of all operator-specific median NT-MoM values was 0.98 (targeted value 1.0) (Table 1). Overall, experience (comparing the first and the second chronological period), CRL > 60 mm and FMF certification had a significant statistical impact improving this standard. FMF-certified operators had a more accurate median NT-MoM (mean of operator-specific medians of 1.00) as compared to the non-certified sonographers (mean of medians was 0.97) (p < 0.05). During the study period, the median of NT-MoM of all operators rose significantly from 0.97 to 0.99 (p < 0.05). A professional profile dedicated to US fetal medicine had a tendency to improve the accuracy of measurements, although nonstatistically significant.

2. Mean percentage of cases over the 95th and below the 5th percentiles were 5.0 and 4.2% respectively (targeted value 5%) (Table 2). Values of CRL < 60 mm and exclusive dedication to fetal medicine had a statistically significant

impact improving this standard. As the experience increased, there was a tendency to move the values closer to the median, decreasing the values over the 95th and below the 5th percentiles. This tendency was also observed in FMF-certified operators.

3. Logarithmic mean and logarithmic SD of the NT-MoM values were 0.00 and 0.13 respectively (mean and SD expected to be 0.00 and 0.08-0.13 respectively)[6] (Table 3). Values of CRL > 60 mm and a dedicated profile had a statistical significant impact improving this standard. Experience and FMF certification had a tendency to reduce the SD (lower dispersion of values), although not statistically significant.

4. The DR for DS at screening time was 90.7% for a FPR of 6.7% for standard screening strategy (maternal age, NT and biochemistry).

5. Figures 1A and B show the CUSUM graph of consecutive NT measurements for each operator, during the last three months, according to FMF certification (excluding measurements > 3 mm). Figure 1A shows the CUSUM graph for the non-FMF certified operators. Figure 1B displays the same chart for the FMF certified operators.

Table 2: Distribution of NT values, expressed in percentiles, according to the different criteria (operator, number of consecutive scans, period, CRL values, FMF certification and professional profile)

NT		≥ 95th percentile		≤ 5th percentile		Total
		n	*%*	*n*	*%*	*n*
Operator	Obs 1	186	6.7 **	52	1.9 **	2.776
	Obs 2	17	2.5 **	9	1.3 **	669
	Obs 3	20	5.7	0	0.0 **	352
	Obs 4	62	6.3 **	38	3.9	987
	Obs 5	3	2.1	1	0.7 **	141
	Obs 6	5	2.5	41	20.6 **	199
	Obs 7	40	2.8 **	92	6.5 **	1.407
	Obs 8	8	5.1	38	24.4 **	156
	Obs 9	200	6.7 **	43	1.4 **	2.968
	Obs 10	107	3.6 **	190	6.4 **	2.969
	Obs 11	15	7.0	8	3.8	213
	Obs 12	79	4.6	76	4.4	1.712
	Obs 13	1	1.8	9	16.4 **	55
	Obs 14	1	0.4 **	8	3.3	243
	Others < 50 US	5	3.8	19	14.5 **	131
No. US	First 100 US	63	4.2	117	7.9 *	1.486
	>100 US	686	5.1	507	3.8 *	13.492
Period	2003-2006	430	6.5 *	328	5.0 *	6.615
	2007-2009	281	3.9 *	172	2.4 *	7.225
CRL (mm)	≤ 60	466	5.6 *	381	4.6 *	8.287
	> 60	283	4.2 *	243	3.6 *	6.691
FMF	Certified	61	3.2 *	42	2.2 *	1.878
	Noncertified	688	5.3 *	582	4.4 *	13.100
Profile	Dedicated	691	5.1 *	500	3.7 *	13.488
	Nondedicated	58	3.9 *	124	8.3 *	1.490
	Overall	749	5.0	624	4.2	14.978

Obs: Observer or operator
No. US: Number of scans performed
FMF: Fetal Medicine Foundation
*p < 0.05 (comparison between criteria)
**p < 0.05 (comparison between obstetricians and expected binomial distribution)

Table 3: Logarithmic mean and logarithmic standard deviation (SD) of the NT-MoM values, according to the different criteria (operator, period, CRL values, FMF certification and professional profile)

		log NT		
		mean	SD	n
Operator	Obs 1	0.01	0.12	2.364
	Obs 2	0.00	0.10	585
	Obs 3	0.05	0.10	294
	Obs 4	0.01	0.13	905
	Obs 5	0.00	0.09	79
	Obs 6	− 0.09	0.15	168
	Obs 7	− 0.04	0.11	51
	Obs 8	− 0.03	0.13	1.232
	Obs 9	− 0.07	0.18	114
	Obs 10	0.02	0.12	2.517
	Obs 11	− 0.03	0.13	2.715
	Obs 12	0.03	0.13	167
	Obs 13	0.00	0.12	1.571
	Obs 14	− 0.02	0.08	225
	Others < 50 US	− 0.07	0.16	100
Period	2003-2006	0.00	0.14	5.560
	2007-2009	0.00	0.11	6.623
CRL (mm) *	≤ 60	− 0.01	0.13	7.620
	> 60	0.00	0.11	5.467
FMF **	Certified	0.00	0.11	1.719
	Noncertified	− 0.01	0.13	11.368
Profile *	Dedicated	0.00	0.13	11.889
	Nondedicated	− 0.01	0.14	1.198
	Overall	0.00	0.13	13.087

Obs: Observer or operator
US: Ultrasound scans
FMF: Fetal Medicine Foundation
*$p < 0.05$
**$p < 0.1$
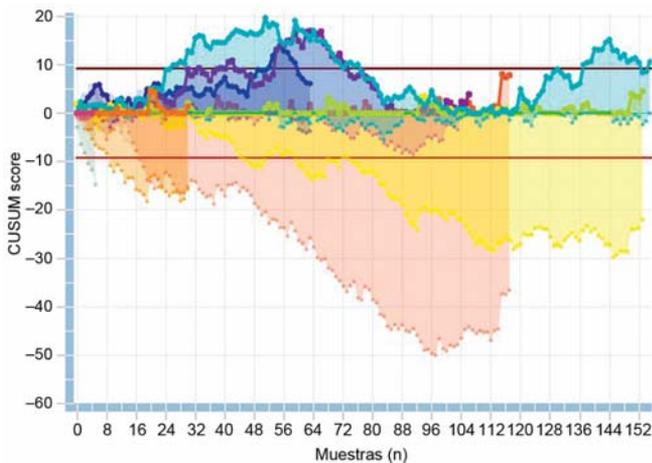
Fig. 1A: Individualized CUSUM chart for the last three-month consecutive NT measurements in the non-FMF certified operator, excluding measurements > 3 mm
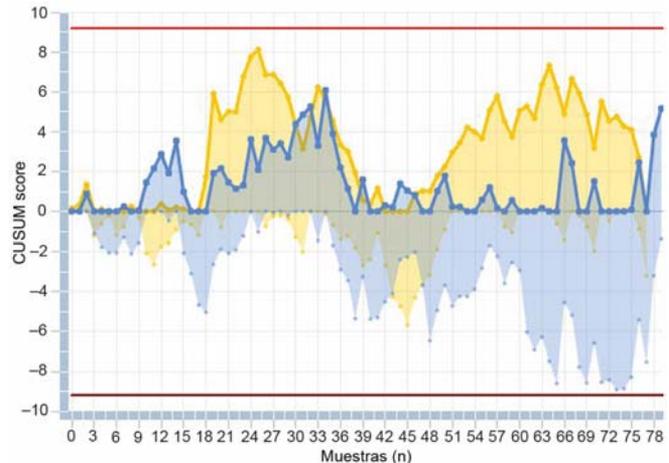
Fig. 1B: Individualized CUSUM chart for the last three-month consecutive NT measurements in the FMF certified operator, excluding measurements > 3 mm

DISCUSSION

Evaluation of prenatal ultrasound measurements for purpose of quality assurance, as known from clinical laboratories analyses,[2,3,9] has only recently received attention from the fetal medicine community.[5,7,8,14-16] From a historical perspective, the NT measurement quality certification policy from the FMF (UK) represents the pioneer and more relevant experience in that field.[17,18] In the United States (US), the Society for Maternal

Fetal Medicine created the nuchal translucency quality review (NTQR) program to similarly provide education and quality review. Despite their common objectives, these two programs differ in their approach to quality control. The FMF has had a continuing reassessment and recertification process in place for nearly a decade, while the NTQR, in part by virtue of his shorter history, is still developing its program for continual monitoring and assessment. Evans has recently demonstrated that a rigid

oversight of NT measurements, as practiced by the UK system, is more effective than the currently practiced in the US system.[1] Recent published papers have pointed the effect of deviation of NT measurements on the performance of screening.[1,4,19,20] Other publications have showed less success for DS screening, probably related to the lack of standardization of NT measurements.[21,22] Our aim was to implement a high level of quality control for DS screening program, particularly for NT screening. Traditionally, the more effective quality controls have relied in the biochemical screening parameters as compared to ultrasound measurements of NT. Our experience demonstrates that strict quality control of clinical measurements, such as NT is possible and as reliable as biochemistry.

In prenatal screening policies, although the highest achievable DR for DS remains a priority, this indicator cannot be used as a reliable marker of quality. The image-scoring methods for quality assessment have been previously introduced in DS quality controls, although they have been showed poorly reproducible, too much time consuming and therefore too expensive to apply on a large scale.[4,5] Extensive qualitative analysis cannot be recommended for ongoing quality control in a NT screening program.[4,6] These systems might be of more value during the initial training period or when quantitative assessment indicates the need for further scrutiny. Epidemiological quantitative quality monitoring is a more practical solution. The advantages of quantitative assessment are the relative simplicity and potential for automation. Moreover, the process of continual assessment and feedback, a well-established principle in business and technology assessments, could be easily adapted to aneuploidy screening in the first trimester. The current cross-sectional study analyzes the performance of NT measurement with regard to several quality standards, individually and for the overall group.

Overall, the quality standards show optimal NT measurements, in terms of satisfactory performance of the screening program and percentage of cases over the 95th percentile. Moreover, data from our center are representative of the expected distribution of NT compared to the Nicolaides reference curve. But epidemiological monitoring of NT measurements shows that there are differences in each of the quality measures chosen, as previously published.[6,7,15] A detailed analysis of the results demonstrates that several operator (experience, dedicated professional profile and FMF certification) and fetal parameters (range of CRL measurements) have a significant impact on the quality standards. More experienced operators, particularly those with a professional profile focused on fetal medicine, CRL over 60 mm and FMF certification have a significant positive impact on the quality standards. In our series, typically and as previously published, there was a tendency to move the measurements closer to the median as the experience increases, with lower dispersion of the extreme values.[7,16,23] CRL range also has an impact, and measurements of NT in fetuses over 60 mm length seems to be more accurate (in terms of median MoM and logarithmic SD),

similar to previous published experiences.[7] Interestingly, each sonographer has the opportunity to compare its own measurements over time with the average measurements performed at the center. Theoretically, an increase in the variation of NT measurements would lead to suboptimal screening results. Accordingly, we focused on the distribution of these measurements in order to discover systematic differences or changes for the individual examiner and for the total group. Our findings show that when well-trained examiners perform NT screening, continuous evaluation of the distribution of the NT-MoM is a good method to assess the quality of the center and may also be useful to identify individual examiners deviating from the mean performance. Moreover, the CUSUM method has recently received attention in the medical literature owing to its simple formulation and very intuitive representation.[12] This statistical toll graphically presents outcomes of consecutive procedures, estimates the putative factors diminishing the accuracy of the procedure and assesses the competence of the operator over a certain period of time, focusing on systematic and random errors. When applying the CUSUM method, the target, the properties and the control limits should be prospectively defined, less stringent at the beginning of the learning process and recalculated according to stricter standards once the initial rates are achieved. In our series, these settings are designed to detect a shift of half the SD (0.125 mm), which corresponds to the statistics suggested in the literature.[12] As shown in the Figures 1A and B, the performance was more acceptable in FMF-certified operators compared with noncertified operators. Figure 1A shows the CUSUM graph in the noncertified group, where the CUSUM raises quick alarm in almost all operators. On the contrary, Figure 1B displays the CUSUM chart for the two-certified operators, with all measurements in control during the same three-month period. Besides its relative simplicity and potential for automation, the main advantages of this method include the early detection of deviations of the measurement compared to other standard quality control indicators. As shown in this study, CUSUM test can be used as a prospective quality control procedure to continuously monitor the performance of sonographers as they assess NT in DS screening.

CONCLUSION

Our experience in epidemiologic monitoring data shows that: (1) Quality standards show optimal NT measurements in our unit; (2) there are differences in each of the quality measures chosen; (3) several operator and fetal parameters have a significant impact on the quality standards; (4) despite intense and unified training for all sonographers, NT measurements drift over time for explained and unexplained reasons. Despite appropriate training, experience, accredited certification and optimal quality standards achieved, continuous monitoring and scrupulous evaluations of individual operators is likely to lead to a better performance of NT screening program. This study demonstrates the importance of ongoing quality assessment.

Subsequently, in our center, regular quality assurance by means of simple and automatic quantitative analysis is going to be conducted at regular 6-month intervals, including prospective CUSUM tool.

Commitment to ongoing quality assessment by our scientific community is needed if early screening for DS can maintain high detection rates and low screen-positive rates in clinical practice.

REFERENCES

1. Evans MI, Pergament E. Impact of quality of nuchal translucency measurements on detection rates of trisomies 13 and 18. Fetal Diagn Ther 2010;27:68-71.
2. Evans MI, Belsky RL, Greb AE, Dvorin E, Drugan A. Wide variation in maternal serum alpha-fetoprotein reports in one metropolitan area: Concerns for the quality of prenatal testing. Obstet Gynecol 1988;72:342-45.
3. Knight GJ, Palomaki GE. Epidemiologic monitoring of prenatal screening for neural tube defects and Down syndrome. Clin Lab Med 2003;23:531-51.
4. Wøjdemann KR, Christiansen M, Sundberg K, Larsen SO, Shalmi A, Tabor A. Quality assessment in prospective nuchal translucency screening for Down syndrome. Ultrasound Obstet Gynecol 2001;18:641-44.
5. Snijders RJ, Thom EA, Zachary JM, Platt LD, Greene N, Jackson LG, et al. First-trimester trisomy screening: Nuchal translucency measurement training and quality assurance to correct and unify technique. Ultrasound Obstet Gynecol 2002;19:353-59.
6. Palomaki GE, Neveux LM, Donnenfeld A, Lee JE, McDowell G, Canick JA, et al. Quality assessment of routine nuchal translucency measurements: A North American laboratory perspective. Genet Med 2008;10:131-38.
7. Koster MP, Wortelboer EJ, Engels MA, Stoutenbeek PH, Elvers LH, Visser GH, Schielen PC. Quality of nuchal translucency measurements in the Netherlands: A quantitative analysis. Ultrasound Obstet Gynecol 2009;34:136-41.
8. Salomon LJ, Ville Y. The science and art of quality in obstetric ultrasound. Curr Opin Obstet Gynecol 2009;21:153-60.
9. Palomaki GE, Lee JES, Canick JA, McDowell GA, Donnenfeld AE. For the ACMG Laboratory Quality Assurance Committee. Technical standards and guidelines: Prenatal screening for Down syndrome that includes first-trimester biochemistry and/or ultrasound measurements. Genet Med 2009;11:669-81.
10. ACOG Committee on Practice Bulletin. ACOG practice bulletin number 77: Screening for fetal CA. Obstet Gynecol 2007;109:217-27.
11. Summers AM, Langlois S, Wyatt P, Wilson RD. Prenatal screening for fetal aneuploidy. J Obstet Gynaecol Can 2007;29:146-79.
12. Biau DJ, Porcher R, Salomon LJ. CUSUM: A tool for ongoing assessment of performance. Ultrasound Obstet Gynecol 2008;31:252-55.
13. Nicolaides KH, Snijders RJ, Cuckle HS. Correct estimation of parameters for ultrasound nuchal translucency screening. Prenat Diagn 1998;18:519-23.
14. Herman A, Maymon R, Dreazen E, Caspi E, Bukovsky I, Weinraub Z. Nuchal translucency audit: A novel image-scoring method. Ultrasound Obstet Gynecol 1998;12:398-403.
15. D'Alton ME, Cleary-Goldman J, Lambert-Messerlian G, Ball RH, Nyberg DA, Comstock CH, et al. Maintaining quality assurance for sonographic nuchal translucency measurements: Lessons from the FASTER Trial. Ultrasound Obstet Gynecol 2009;33:142-46.
16. Evans MI, Van Decruyes H, Nicolaides KH. Nuchal translucency measurements for first-trimester screening: The 'price' of inaccuracy. Fetal Diagn Ther 2007;22:401-04.
17. Snijders RJ, Noble P, Sebire N, Souka A, Nicolaides KH. UK multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal translucency thickness at 10-14 weeks of gestation. Fetal Medicine Foundation First Trimester Screening Group. Lancet 1998;352:343-46.
18. Nicolaides KH. The 11-13 + 6 weeks scan. London, Fetal Medicine Foundation 2004.
19. Kagan KO, Wright D, Etchegaray A, Zhou Y, Nicolaides KH. Effect of deviation of nuchal translucency measurements on the performance of screening for trisomy 21. Ultrasound Obstet Gynecol 2009;33:657-64.
20. Evans M, Krantz D, Hallahan T, Sherwin J. Skewed to the left: Under measurements of NT's and implications for screening efficiency. Am J Obstet Gynecol 2009;201:S140.
21. Kornman LH, Morssink LP, Beekhuis JR, De Wolf BT, Heringa MP, Mantingh A. Nuchal translucency cannot be used as a screening test for chromosomal abnormalities in the first trimester of pregnancy in a routine ultrasound practice. Prenat Diagn 1996;16:797-805.
22. Bewley S, Roberts LJ, Mackinson AM, Rodeck CH. First trimester fetal nuchal translucency: Problems with screening the general population. Br J Obstet Gynaecol 1995;102:386-88.
23. Wapner R, Thom E, Simpson JL, Pergament E, Silver R, Filkins K, Platt L, et al. First trimester maternal serum biochemistry and fetal nuchal translucency screening (BUN) study group. First-trimester screening for trisomies 21 and 18. N Engl J Med 2003;9;349:1405-13.