Does prior knowledge of maternal age affect judgment of operators measuring nuchal translucency?

Czy wcześniejsza znajomość wieku ciężarnej przez badającego może mieć wpływ na pomiar przezierności karku?

Jose Carlos P. Ferreira[1], Miroslaw Wielgos[1], Dariusz Borowski[1], Piotr Sieroszewski[2], Bartosz Czuba[3], Wojciech Cnota[3], Agata Wloch[3], Krzysztof Sodowski[3], Piotr Wegrzyn[1,4]

[1] 1st Department of Obstetrics and Gynecology, Medical University of Warsaw, Warsaw, Poland
[2] Department of Fetal Medicine and Gynecology, Medical University of Lodz, Lodz, Poland
[3] Department of Obstetrics and Gynecology in Ruda Slaska, Medical University of Silesia, Ruda Slaska, Poland
[4] Department of Obstetrics and Perinatology, Medical University of Warsaw, Warsaw, Poland

Abstract

Objectives: *To test the hypothesis that, in real life standard clinical practice, knowledge of maternal age (MA) by operators measuring nuchal translucency (NT) for screening of aneuploidy may influence their judgment, resulting in a tendency to over-measurement in older women.*

Material and methods: *We retrospectively analyzed the correlation between MA and NT MoMs in data from a group of operators from several clinical practices, with different levels of experience.*

Results: *We assessed 66,918 measurements by 41 operators. There was no association between NT and MA in all the measurements analyzed together. In 3 experienced operators (N>1900), there was a significant association between the variables, although all were negative and its effect size was very small (0.004, 0.006 and 0.01). However one of the less experienced operators (N=47) had a statistically significant (p=0.0002) and strong (R2=0.2634) association. We tested the hypothesis that this bias could occur in less experienced operators but time/experience would correct it. We did the same analyses for each set of 50 tests, sorted by date, for each operator, up to the 7th set. No significant progression was identified in association with increase in experience.*

Conclusions: *Our data does not support the hypothesis that operators might be biased towards over-measuring NT in older women.*

Key words: **nuchal translucency** / **maternal age** / **operator bias** / **prenatal ultrasound** / / **aneuploidy screening** /

Corresponding author:
Piotr Wegrzyn
Department of Obstetrics and Perinatology, Medical University of Warsaw
ul. Żwirki i Wigury 63a, 02-091 Warsaw, Poland
e-mail: piotr.wegrzyn@wum.edu.pl

Otrzymano: **03.03.2015**
Zaakceptowano do druku: **01.04.2015**

Streszczenie

Cel pracy: *Celem pracy była ocena hipotezy, że w rutynowej praktyce klinicznej wcześniejsza znajomość wieku ciężarnej przez lekarza wykonującego pomiar przezierności karku w ramach skriningu w kierunku aberracji chromosomowych, może prowadzić do zawyżania pomiaru u starszych kobiet.*

Materiał i metody: *Przeanalizowano retrospektywnie dane z 66918 badań wykonywanych przez 41 lekarzy o różnym stopniu doświadczenia, pracujących w kilku różnych ośrodkach i zbadano korelację pomiędzy pomiarem przezierności karku i wiekiem ciężarnej.*

Wyniki: *Nie stwierdzono związku między zmierzoną wielkością przezierności karku i wiekiem ciężarnej. Natomiast u 3 doświadczonych lekarzy (>1900 wykonanych badań) stwierdzono istotną statystycznie korelację pomiędzy zmiennymi, aczkolwiek była ona negatywna i o niewielkim znaczeniu praktycznym (p=0,004, p=0,006 i p=0,01). Jednakże u jednego z mniej doświadczonych lekarzy (47 zarejestrowanych badań w bazie) stwierdzono istotną (p=0,0002) i silną korelację (R2=0,2634). Zbadano hipotezę, czy taki błąd może występować u mniej doświadczonych lekarzy i zmniejszać się z czasem. W analogiczny przeanalizowano sposób wyniki z 7 kolejnych zbiorów danych liczących po 50 badań dla każdego lekarza. Nie wykazano istotnej zmiany wraz ze wzrostem doświadczenia.*

Wyniki: *Uzyskane wyniki nie potwierdzają hipotezy, że istnieje tendencja do zawyżania pomiarów przezierności karku u starszych kobiet.*

Słowa kluczowe: **przezierność karku** / **wiek ciężarnej** / **błąd pomiaru** / **ultrasonografia prenatalna** / **skrining w kierunku aberracji chromosomowych** /

Introduction

Measurement of the thickness of fetal Nuchal Translucency (NT), a collection of fluid at the back of fetal neck, is an ultrasound based phenotypic marker currently used in the assessment of the risk for fetal developmental anomalies, the most common of which is Down syndrome, caused mostly by Trisomy 21 [1-3]. Increased NT can also be a manifestation of rare genetic syndromes [4].

It has been shown that, without proper operator training and quality control, the effectiveness of NT measurement in that risk assessment varies wildly to the point of uselessness [5-8]. Specific training and quality control programs have been developed to address some of the potential operator related biases and to decrease the inter observer variability (e.g. programs of Fetal Medicine Foundation (FMF) - http://www.fetalmedicine.com/fmf/ - and of Perinatal Quality Foundation - http://www.perinatalquality.org.

However other potential operator related biases have been hypothesized, especially in the setting of conference proceedings discussions. One of those hypothesized biases is described as a possible tendency of operators to over measure NT in mothers whose age is perceived to be advanced. According to the proponents of this hypothesis, this would possibly be an unconscious tendency to avoid false negatives, at the expenses of an increase in the false positive rate for older women. Although there is evidence that no correlation exists between maternal age and the NT thickness, when corrected for Gestational Age (GA) [9], that evidence is, not only inconsistent [10] but also argued to be obtained in the setting of prospective research studies. In real-life medical practice situations, the referred bias would be likely to occur.

Since, as per our knowledge, there hasn't been any study trying to assess this hypothesis, we decided to test it ourselves.

Aim of the study

To test the hypothesis that, in real life standard clinical practice, knowledge of maternal age (MA) by operators measuring nuchal translucency (NT) for screening of aneuploidy may influence their judgment, resulting in a tendency to over-measurement in older women.

Material and methods

This was a retrospective descriptive and correlation study of anonymized data obtained from several FMF certified clinical practices.

Given the methodology of this study (retrospective, anonymized, non-interventional), as per the local rules, there was no need for review by a research ethics board.

If there is an operator bias causing an over measurement of the nuchal translucency, a positive correlation would be expected between NT and maternal age (MA). Therefore, to test the hypothesis we assessed the correlation between those two variables.

All patients, in Poland are offered common aneuploidy screening, although the costs must be supported by all patients less than 35 y.o. NT measurement data, spanning a period of 11 years, from patients having their screening in the first trimester, was extracted from record databases from 7 prenatal outpatient clinics in Poland all managed with Astraia software (several versions were included, Astraia Gmbh, Munich, Germany). Several ultrasound equipments were used (GE Voluson 730 PRO, GE Voluson 730 Expert, GE Voluson E8, Aloka F75, Aloka Alpha 10). Measurements included in the study were obtained from operators with different levels of experience, although all received their Fetal Medicine Foundation (FMF) NT certificate at the appropriate experience level and were enrolled in the foundation's NT measurement's quality control program. The measurements were selected if the operator had an arbitrarily

selected minimum of 40 measurements experience. All operators entered the demographics of the patient into the database prior to the start of the exam. Therefore all operators had prior knowledge of the maternal age while performing the NT measurement.

Data used for the analyses included date of exam, patient's date of birth (used to calculate MA at the time of the exam), Crown-Rump Length (CRL) (used to calculate predicted median), NT, Operator ID, and database ID number.

NT measurements were excluded if some of the required data was not available or if they were extremely high (more than 3.5 mm, corresponding to 99th percentile) because we wanted to exclude the clearly abnormal cases (more frequent in older woman) and we also think that in those extreme values the type of bias we were testing is unlikely to occur. The NT units used in the analyses were the multiples of the median (MoM), calculated from the ratio between the NT measurement and the predicted median for the CRL. The predicted median NT was calculated for the CRL measurement using the formula derived by Wright and al., 2008 [10]. MA was measured in years rounded up to 2 decimals.

We analyzed the linear correlations between NT and MA for all selected measurements, for each individual operator and for sets of arbitrarily chosen 50 measurements sampled at different levels of experience for each operator.

We used linear regression analysis. The statistical significance of the correlations was assessed by p value, which was considered significant if lower than the multiple test Bonferroni corrected base level of 0.05. The strength of the correlations analyzed was measured by the square of the R correlation index (R^2), the measure of association of linear correlations. We also analyzed the direction of the correlation, positive or negative, by the sign of the slope of the linear correlation line.

Correlation analyses were done using Statistica 64 Software (Version 10, Stat Soft. Inc., Tulsa, OK, USA). Plots were done with Systat 64 Software (Version 13.1, Systat Soft. Inc. San Jose, California, USA).

Results

All the NT measurements smaller than the 99th percentile, taken by 41 operators corresponding to our selection criteria, from 7 prenatal clinics in Poland were selected for our analyses. The NT measurements had been taken from Jan 2004 to Feb 2014. A total of 66,918 measurements were complete and thus were selected for the analysis.

There was no correlation between NT MoM and MA for all the exams (p value was 0.42 with a R^2 of zero).

We then analyzed each operator individually.

The number of NT measurements per operator ranged from 44 to 5316, with a median of 1006. Only 6 operators had less than 300 exams, of which 3 had 40 to 50 measurements (here labelled as inexperienced). The lower and upper quartiles were, respectively, 384 and 2464 NT measurements.

With a multiple test Bonferroni corrected p value of 0.0012, 4 operators had a statistically significant association between NT MoMs and MA (figure 1A). However, with the exception of only one of these 4 operators, the effect size of the remaining 3 was very small (R^2 of 0.004, 0.006 and 0.01). The effect size of all the other 40 operators had also very low R^2 (the maximum was 0.02, for a p value of 0.32) (Figure 1A). The low p values of the 3 operators is likely explained by their high number of exams (respectively N=3,501, 1,981 and 3,940) (Figure 1B), which can make very small effect sizes statistically significant. Furthermore, the slope of the correlation line for those 3 operators is negative (Figure1C). This means that the correlation goes the opposite direction of the hypothesis under testing, that the prior knowledge of maternal age leads the operator to over measure the NT. Interestingly enough, there were 24 operators with positive correlations and 17 with negative correlations, which is not statistically significantly different from random variation (Figure 1C).

However there was an inexperienced operator (N = 47) who showed a statistically significant (p value = 0.002) and strong (R^2 = 0.26) positive association between NT MoM and MA. To verify if inexperience was the reason for the bias we sampled sets of 50 measurements of each operator, collected over time, from the earliest measurements (measurement 1 to 50) to the arbitrarily chosen 7th set, corresponding to exams 300 to 350, and we analyzed the correlation under study for each set of each operator. We analyzed 257 samples of 50 sets. The mean and the median of the R^2 measures of association for all the sets was, respectively 0.028 and 0.015, with a 95th percentile of 0.028. There was no tendency towards a decrease, with experience (here measured by number of exams accumulated with time), of the strength of the association (R^2) (Figure 2). Furthermore there was only one more measure of association that was strong and statistically significant, coincidently with the same rounded R^2 and p value as the one referred above. This second strong correlation occurred in the sixth set of a different operator in whom all the other sets did not show the association under study.

Discussion

Fetal ultrasound is a potentially excellent method for screening and diagnosis of fetal anomalies. However it relies heavily on the training of the operator. Fetal NT measurement, one of the parameters frequently used in the risk assessment of fetal defects in the first trimester, has also been shown to have its effectiveness affected by operator related factors [5]. Specific operator training and QC programs have been shown to resolve this problem [6-7]. It is still possible that some operator or even non-operator related bias might interfere with the performance of NT measurements based aneuploidy screening. For instance, recently a serious bias attributed to the equipment has been described [12].

Mainly in the setting of discussion at conference proceedings, the possibility that NT measuring operators are biased by prior knowledge of maternal age, in a way that may lead them to over measure the NT in pregnant patients perceived to be older, and thus with higher risk for Trisomy 21, has been tossed around. The end result of this would be a bias attributable component of the increased false positive rate in older patients, seen when a predicted risk value is used as a cut-off for considering a test as positive [13]. Against this hypothesis there are reports of studies demonstrating that there is no correlation between the NT measurement and MA [9]. However, that evidence is obtained from prospective studies, that is, the data is obtained in non-natural, research settings. As per the proponents of that hypothesis, such bias would occur in less controlled, real-life common clinical practice.

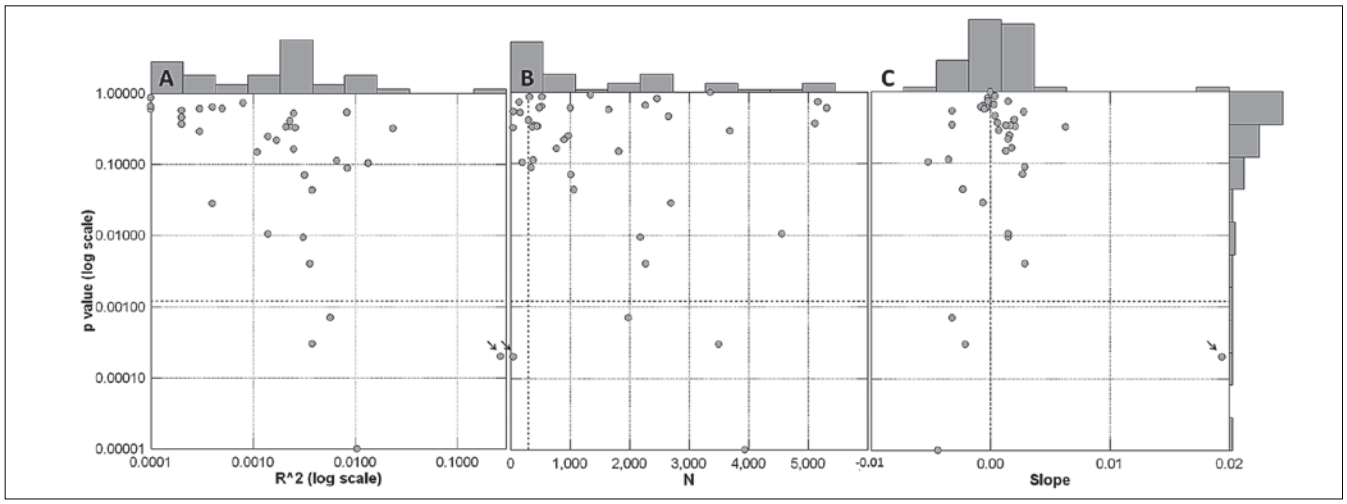

Figure 1.
Scatter plots showing, for each operator, represented by each dot, the p values for the associations between Maternal Age (MA) and Nuchal Translucency (NT) Multiples of the Median (MoMs), in the Y axes.

The X axes are different in each of the plots as explained below for each plot. The p values Y axis is in logged scale to make it easier to see the differences between very similar values.

The horizontal traced line marks the multiple testing Bonferroni corrected significant p value (0.0012). Histogram plots in both axes show the density distribution of the 2 measures represented.

The p values were mostly non-significant, with the majority bigger than 0.1. Only 4 operators had significant p values.

The arrowed dot in all the plots corresponds to the single operator described in detail in the text.

Plot A shows, for each operator, in the X axis, the respective measure of association (R2) between MA and NT MoMs. This is also in logged scale. This plot shows that the effect sizes, including 3 of the 4 operators with significant associations, were very small and with 2 density peaks, one between 0.001 and 0.01 and the other close to 0.0001. Only one of the operators (arrow) showing a statistically significant association had a relatively high effect size.

Plot B shows, for each operator, in the X axis, the respective number of exams (N). The vertical traced line corresponds to N=300. Approximately half of the operators had less than 1,000 exams performed. This plot shows that, of the 4 operators with significant associations between MA and NT MoMs, 3 of them had very high number of datapoints, which makes it possible to demonstrate statistical significance for very small effect sizes (see Figure 2A for the effect sizes), unlikely to have any clinical impact.

Plot C shows, for each operator, in the X axis, the respective slope of the linear correlation between MA and NT MoMs. This plot shows that there were almost as many operators with a negative as with a positive correlation and, of the 4 operators with significant associations between MA and NT MoMs, 3 of them, the ones with very high number of exams and very low effect sizes, had a negative slope, which is the opposite direction of the tested hypothesis.

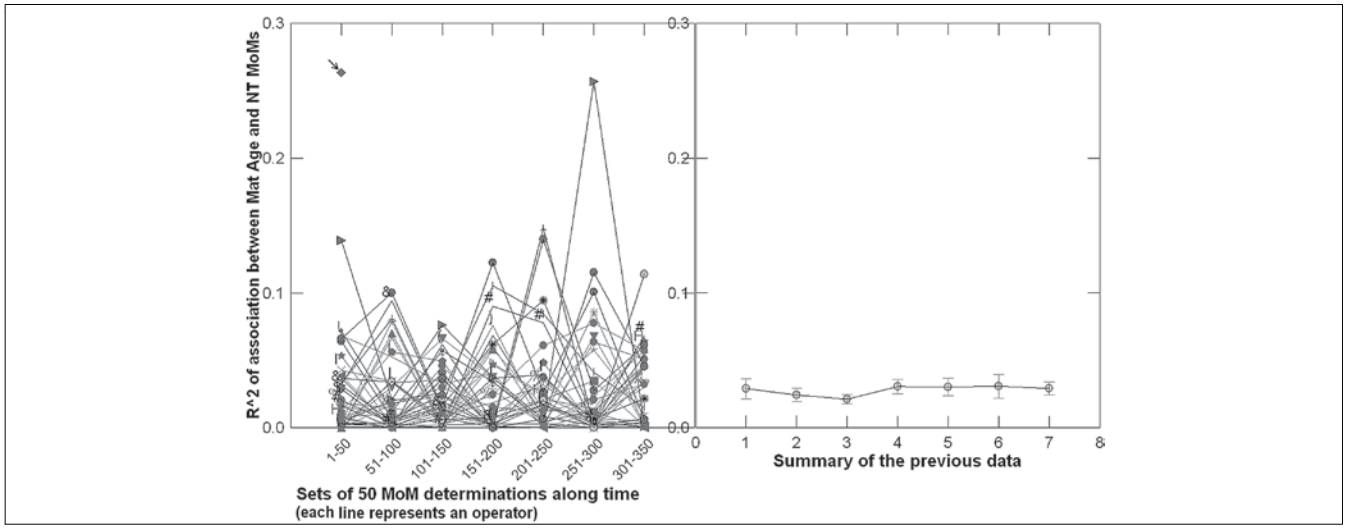

Figure 2.
Repeated measures plot of the measure of association (R2) between Maternal Age and Nuchal Translucency Multiples of the Median (Y axis), in each of the samples of 50 NT measurements for each of the operators along time, here measured in sets of exams ordered by date. The plot on the left has all the values for all the operators. Each datapoint represents a set of 50 measurements for a given operator. Each operator is represented by a different symbol. The lines link the sampled sets of each operator along time. In the X axis, each mark represents a set of 50 exams, being the first set represented by exams 1 to 50, the second set by exams 51 to 100, etc.

The plot on the right shows the summary of the first plot as the mean and standard error of the sets of datapoints in each of the timed sets of 50 exams. This is to show that there is no tendency to an improvement (decrease) in the measure of the association, as experience (number of exams) increases. It also shows that the operator for which a significant strong association between MA and NT MoMs was found (arrow), is likely to represent random variation of the measure of the association among samples of 50 exams, as another set for another operator reached a similarly high level of association in one of his samples of 50 measurements, even at a higher experience level (set 251-300).

In this work we tested this possibility by assessing the correlation between NT measurements, in the form of NT MoM, and maternal age, in data obtained from such real-life clinical settings. If this hypothesis was true and it indeed affected, in a clinically relevant manner, the effectiveness of this measurement, we would find a statistically significant and strong correlation between these two variables. We did several types of analyses – an aggregate analysis of 66,918 measurements, an operator specific analysis of 41 operators of different levels of experience, and an operator and level of experience specific analysis.

The first, aggregate, analysis, not showing the correlation expected under the hypothesis, did not support it.

We decided to analyze each operator individually to see if some operators could indeed be biased as per the tested hypothesis. Our operator sample included a broad range of experience level and most operators had a reasonable to large experience.

Only in one operator a statistically significant positive strong correlation was demonstrated. For all the others there was no statistically significant correlation or the effect sizes were too small, unlikely to have any clinical effect.

These data shows that the vast majority of operators (40 out of 41) are not biased as per the tested hypothesis.

The only operator that was consistent with the bias (statistically significant strong positive correlation between NT MoM and MA, likely to have a clinical effect) was inexperienced (N = 47). That led us to hypothesize that inexperience with NT measurement may indeed predispose to this bias. As experience is acquired the bias will tend to disappear. If this was the case, the measure of this association could be an indirect QC assessment of the experience of the operator and its absence a requirement for certification.

Against this second hypothesis, there was no progressive decrease in the value of the correlations with increase in experience. And, in total, only two instances of relevant associations were found; one indeed was observed in the first set (inexperienced set) of measurements of one of the operators but the second was observed in a set in which inexperience was unlikely. These two identified relevant associations are thus likely to represent just random variation that may occur when enough sets of 50 exams are randomly sampled. The identified association found in the inexperienced operator is thus unlikely to be explained by inexperience related bias.

The results of our study, showing no consistent strong significant positive correlations between NT MoM and MA, for any of 41 operators, regardless of experience, did not prove the hypothesis.

Conclusion

Our data does not support the hypothesis that there is clinically relevant operator related bias towards an overestimation of NT in older women.

Oświadczenie autorów:

1. Jose Carlos P. Ferreira – analiza statystyczna wyników, przygotowanie manuskryptu i piśmiennictwa
2. Mirosław Wielgoś – ostateczna weryfikacja i akceptacja manuskryptu
3. Dariusz Borowski – współautor protokołu, zebranie materiału
4. Piotr Sieroszewski – zebranie materiału, przygotowanie manuskryptu
5. Bartosz Czuba – zebranie materiału, opracowanie wyników badań
6. Wojciech Cnota – zebranie materiału, opracowanie wyników badań
7. Agata Włoch – zebranie materiału, opracowanie wyników badań
8. Krzysztof Sodowski – zebranie materiału, przygotowanie manuskryptu
9. Piotr Węgrzyn – autor koncepcji i założeń pracy, przygotowanie manuskryptu i piśmiennictwa, autor zgłaszający i odpowiedzialny za manuskrypt.

Źródło finansowania:

Część projektu finansowanego z grantu NCN nr 1233/B/P01/2011/40. This project was supported by The National Science Centre (Grant No. 1233/B/P01/2011/40). The supporting source had no involvement in the study design, in the collection, analysis and interpretation of data, in the writing of the report and in the decision to submit the report for publication.

Konflikt interesów:

Autorzy nie zgłaszają konfliktu interesów oraz nie otrzymali żadnego wynagrodzenia związanego z powstawaniem pracy.

References

1. Nicolaides KH. Nuchal translucency and other first-trimester sonographic markers of chromosomal abnormalities. *Am J Obstet Gynecol.* 2004, 191, 45-67.
2. Kagan KO, Avgidou K, Molina FS, [et al.]. Relation between increased fetal nuchal translucency thickness and chromosomal defects. *Obstet Gynecol.* 2006, 107 (1), 6-10.
3. Chasen ST, Sharma G, Kalish RB, Chervenak FA. First-trimester screening for aneuploidy with fetal nuchal translucency in a United States population. *Ultrasound Obstet Gynecol.* 2003, 22 (2), 149-151.
4. Olejniczak T, Niepsuj-Biniaś J, Rabiega-Gmyrek D, [et al.]. Zespół Escobara u płodu – opis przypadku. *Ginekol Pol.* 2014, 85 (8), 629-632.
5. Snijders R, Smith E. The role of fetal nuchal translucency in prenatal screening. *Curr Opin Obstet Gynecol.* 2002, 14 (6), 577-585.
6. Snijders RJ, Thom EA, Zachary JM, [et al.]. First-trimester trisomy screening: nuchal translucency measurement training and quality assurance to correct and unify technique. *Ultrasound Obstet Gynecol.* 2002, 19, 353-359.
7. D'Alton ME, Cleary-Goldman J, Lambert-Messerlian G, [et al.]. Maintaining quality assurance for sonographic nuchal translucency measurement: lessons from the FASTER Trial. *Ultrasound Obstet Gynecol.* 2009, 33, 142-146.
8. D'Alton ME, Cleary-Goldman J. Education and quality review for nuchal translucency ultrasound. *Semin Perinatol.* 2005, 29, 380-385.
9. Biagiotti R, Periti E, Brizzi L, [et al.]. Comparison between two methods of standardization for gestational age differences in fetal nuchal translucency measurement in first-trimester screening for trisomy 21. *Ultrasound Obstet Gynecol.* 1997, 9, 248-52.
10. Orlandi F, Damiani G, Hallahan TW, [et al.]. First-trimester screening for fetal aneuploidy: biochemistry and nuchal translucency. *Ultrasound Obstet Gynecol.* 1997, 10, 381-386.
11. Wright D, Kagan KO, Molina FS, [et al.]. A mixture model of nuchal translucency thickness in screening for chromosomal defects. *Ultrasound Obstet Gynecol.* 2008, 31, 376-383.
12. Axell RG, Gillett A, Pasupathy D, [et al.]. The accuracy of nuchal translucency measurement depends on the equipment used and its calibration. *Ultrasound Obstet Gynecol.* 2014. 44, 31-37.
13. Schmidt P, Rom J, Maul H, [et al.]. Advanced first trimester screening (AFS): an improved test strategy for the individual risk assessment of fetal aneuploidies and malformations. *Arch Gynecol Obstet.* 2007, 276, 159-166.