Ultrasound Obstet Gynecol 2009; **33**: 147–151
Published online 13 November 2008 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/uog.6250

Comparison of Prenatal Risk Calculation (PRC) with PIA Fetal Database software in first-trimester screening for fetal aneuploidy

C. HÖRMANSDÖRFER*, A. SCHARF†, M. GOLATTA†, B. VASKE‡, A. CORRAL§, P. HILLEMANNS* and P. SCHMIDT*

*Departments of *Obstetrics and Gynaecology and ‡Biometry and Medical Statistics, Medical University of Hanover, Hanover, †Department of Obstetrics and Gynaecology, University of Heidelberg, Heidelberg and §Department of Obstetrics and Gynaecology, Evangelical Hospital Oberhausen, Oberhausen, Germany*

KEYWORDS: aneuploidy; Down syndrome; first-trimester screening; prenatal risk calculation; trisomy 21

ABSTRACT

Objectives *In February 2007 new software, Prenatal Risk Calculation® (PRC), for calculating the risk of fetal aneuploidy was introduced in Germany. Our aim was to investigate its test performance and compare it with that of the PIA Fetal Database® (PIA) software developed and used by The Fetal Medicine Foundation.*

Methods *Between 31 August 1999 and 30 June 2004 at the Women's Hospital of the Medical University of Hanover in Germany, 3120 singleton pregnancies underwent combined first-trimester screening at 11 + 0 to 13 + 6 weeks of gestation. Calculation of risk for fetal aneuploidy was computed prospectively using the PIA software. In a subsequent retrospective analysis, we recalculated risks for the 2653 of these datasets with known fetal outcome using the PRC software and compared the results.*

Results *Of the 2653 datasets analyzed, 17 were cases of aneuploidy. At a cut-off of 1:230, for the detection of fetal aneuploidy, the respective sensitivity, false-positive rate and positive predictive value were 70.6%, 4.1% and 9.9% for PRC and 76.5%, 2.9% and 14.6% for PIA. At a cut-off of 1:300, the equivalent values were 70.6%, 5.6% and 7.5% for PRC and 76.5%, 4.0% and 11.0% for PIA. The differences in test performance between the two types of software were highly significant ($P < 0.0001$).*

Discussion *The test performance of PRC was inferior to that of PIA, the sensitivity for detection of fetal aneuploidy being lower and the false-positive rate higher. Had PRC been employed prospectively in our study, 40% more women examined would have been offered unnecessarily an invasive procedure for fetal karyotyping. Copyright © 2008 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

Internationally, the gold standard for non-invasive first-trimester fetal aneuploidy screening is that developed by The Fetal Medicine Foundation (FMF)[1]. Its calculation strategy, which is utilised in the PIA Fetal Database® (PIA) software, is based on 100 311 datasets collected prospectively and employs the likelihood ratio method of Palomaki and Haddow[2]. Accordingly, a maternal- and gestational-age dependent background risk is modified sequentially by likelihood ratios, which have been established with respect to the following measurement parameters[1]: nuchal translucency thickness (NT) in mm; pregnancy associated plasma protein-A (PAPP-A) level in multiples of the median (MoM); and free beta-human chorionic gonadotropin (free β-hCG) in MoM. The result represents an individually adapted fetal risk for an aneuploidy, with an indication for invasive testing if a defined cut-off value is exceeded.

In February 2007 new risk calculation software, Prenatal Risk Calculation® (PRC), was introduced in Germany[3–6]. It also utilizes the classic factors of background risk, NT, PAPP-A and free β-hCG. The algorithm is, however, based on a retrospective analysis of 70 030 pregnancies[6,7]. For risk calculation with biochemical values, newly developed 'degrees of extremeness' (DoE) replace the MoM concept[5,7–9]. The

Correspondence to: Dr C. Hörmansdörfer, Department of Obstetrics and Gynaecology, Medical University of Hanover, Carl-Neuberg-Str. 1, 30625 Hanover, Germany (e-mail: cinsanchez@hotmail.com)

Accepted: 18 September 2008

DoE is a ratio of the distance between the median value and actual value and the distance between the median value and 5th centile (when the measured value is below the median) or the distance between the median and 95th centile (when the measured value is above the median)[5,6]. Under this assumption, a DoE is 0 at the median value, 1.0 at the 95th percentile, and −1.0 at the 5th percentile[9]. Additionally, the Bayesian theorem[10] replaces the mathematical concept of sequential likelihood ratios of Palomaki and Haddow[2].

The aim of this study was to investigate the test performance of the new PRC software and compare it with that of the PIA software developed and used by The FMF.

METHODS

Between 31 August 1999 and 30 June 2004 at the Women's Hospital of the Medical University of Hanover in Germany, 3120 singleton pregnancies underwent combined first-trimester screening at 11 + 0 to 13 + 6 weeks of gestation. Ultrasonographic measurement of NT was obtained according to The FMF standard protocol[11] by four experienced and FMF-certified examiners. The following high-resolution ultrasound devices were used: GE Logiq 500, GE Logiq 700 (GE Healthcare Ltd., Chalfont St. Giles, UK), Hitachi EUB 8500 and Hitachi EUB 6500 (Hitachi Ltd., Tokyo, Japan). At the Institute for Prenatal Diagnosis and Human Genetics in Peine, Germany, a FMF-certified and quality-controlled laboratory, the concentrations of both biochemical parameters PAPP-A and free β-hCG were determined with a Brahms Kryptor system (Brahms AG, Heringsdorf/Berlin, Germany) and converted to MoM. An adjusted risk calculation for fetal aneuploidy was computed prospectively with PIA Fetal Database software (GE-Viewpoint, Wessling, Germany), which uses the original first-trimester screening algorithm proposed by Nicolaides *et al.*[12]. If the adjusted risk reached or exceeded a defined cut-off value of 1 : 300, women were offered individually adapted counseling concerning further invasive testing (amniocentesis or chorionic villus sampling). After delivery, all participating women were asked to report their baby's health status in a questionnaire. Further fetal outcomes were retrieved from birth protocols and neonatal examination reports from the respective hospitals. In order to produce test performance results which could be compared to those of PRC, the adjusted risks generated by PIA were analyzed at cut-off values of 1 : 230 as well as 1 : 300.

We excluded 467 (15%) datasets due to unknown fetal outcome. In a retrospective analysis, risks of the remaining 2653 datasets were recalculated with the PRC software (version 1.0.6.1, Nexus/GMT GmbH, Frankfurt/Main, Germany) at the recommended cut-off values of 1 : 230[4–6] and 1 : 300[7]. NT values were entered in mm. Absolute values of the biochemical parameters PAPP-A and fβ-hCG were converted to DoE.

In a four-fold table analysis, test performance parameters (sensitivity, specificity, false-positive rate, and positive predictive value) of both algorithms were compared. Significance values were obtained by Pearson's chi-square and McNemar tests, and by receiver–operating characteristics (ROC) curves. For statistical analysis, Analyse-it for Microsoft Excel (version 1.73, Analyse-it Software Ltd., Leeds, UK) was utilized. Results were validated by the Department for Medical Statistics and Biometry, Medical University of Hanover.

RESULTS

For the 2653 complete data sets of singleton pregnancies available for analysis, the mean (range) maternal age was 31.2 (16.5–45.7) years. The study cohort included 17 cases of aneuploidy: 10 of trisomy 21, four of trisomy 18, one of trisomy 13, one of Turner syndrome, and one of triploidy. One case with Dandy-Walker syndrome and one case with a Turner/triple X mosaic were considered unremarkable, because they are not typically detected in first-trimester screening.

Measured NT values ranged from 0.5 to 8.0 mm (mean, 1.5 mm; median, 1.5 mm). Absolute values of PAPP-A ranged from 0.2 U/L to 24.8 U/L (mean, 3.7 U/L; median, 3.1 U/L), corresponding to 0.4–7.8 MoM (mean, 1.1 MoM; median, 1.0 MoM). Absolute free β-hCG values ranged from 3.2 U/L to 398.3 U/L (mean, 45.6 U/L; median, 36.3 U/L), corresponding to 0.1–11.1 MoM (mean, 1.3 MoM; median, 1.1 MoM).

Irrespective of the chosen cut-off value, both PIA and PRC correctly identified nine out of 10 trisomy 21 cases (sensitivity, 90%), and the case of Turner syndrome (sensitivity, 100%). However, PIA detected three out of four trisomy 18 cases (sensitivity, 75%), while PRC detected only two (sensitivity, 50%). One case each of trisomy 21 and trisomy 18, as well as the cases of trisomy 13 and triploidy, attained false-negative results with both software programs. In these cases, an unremarkable NT and/or biochemical value led to risk values that were far lower than the cut-off value. The measurement values and the test results of all aneuploidies are displayed in Table 1.

Test performance of PIA

At a cut-off value of 1 : 300, PIA reached an overall sensitivity of 76.5% (95% CI, 50.1–93.2%), a false-positive rate of 4.0% (95% CI, 3.3–4.8%) and a positive predictive value of 11.0% (95% CI, 6.0–18.1%). At a cut-off value of 1 : 230, PIA reached an overall sensitivity of 76.5% (95% CI, 50.1–93.2%), a false-positive rate of 2.9% (95% CI, 2.3–3.6%) and a positive predictive value of 14.6% (95% CI, 8.0–23.7%). The results, in terms of screening performance, were highly significant ($P < 0.0001$). The respective contingency tables are displayed in Tables 2 and 3.

Test performance of PRC

At a cut-off value of 1 : 300, PRC attained an overall sensitivity of 70.6% (95% CI, 44.0–89.7%), a false-positive rate of 5.6% (95% CI, 4.7–6.5%) and a positive predictive value of 7.5% (95% CI, 4.0–12.8%). At a cut-off value of 1 : 230, PRC attained an overall sensitivity of 70.6% (95% CI, 44.0–89.7%), a false-positive rate of 4.1% (95% CI, 3.4–5.0%) and a positive predictive value of 9.9% (95% CI, 5.2–16.7%). The results, in terms of screening performance, were highly significant ($P < 0.0001$). Tables 4 and 5 illustrate the respective contingency tables.

Table 1 Test results and comparison of PIA Fetal Database® (PIA) and Prenatal Risk Calculation® (PRC) software in 17 first-trimester pregnancies with aneuploidy

Diagnosis	*MA (years)*	*CRL (mm)*	*NT (mm)*	*PAPP-A (MoM)*	*Free β-hCG (MoM)*	*PIA*	*PRC*
Trisomy 21	39	74.9	7.9	0.51	1.34	+	+
Trisomy 21	36	66.0	2.4	0.22	6.18	+	+
Trisomy 21	32	68.0	3.4	0.71	5.16	+	+
Trisomy 21	32	68.2	3.0	0.48	3.03	+	+
Trisomy 21	37	65.0	1.9	0.41	4.85	+	+
Trisomy 21	25	54.0	5.8	0.14	1.79	+	+
Trisomy 21	23	77.3	4.7	0.29	1.43	+	+
Trisomy 21	35	59.0	1.9	0.23	1.86	+	+
Trisomy 21	29	47.0	4.0	2.56	1.91	+	+
Trisomy 21	33	61.8	1.2	0.56	1.10	−	−
Trisomy 18	42	45.0	4.1	0.22	0.26	+	+
Trisomy 18	36	61.9	2.0	0.07	0.09	+	−
Trisomy 18	24	56.4	3.5	0.15	0.35	+	+
Trisomy 18	42	57.0	1.5	0.39	0.12	−	−
Trisomy 13	23	47.0	1.2	0.54	0.52	−	−
Turner (X0)	34	56.5	8.0	0.52	0.42	+	+
Triploidy	31	51.9	1.9	0.35	0.12	−	−

+, test positive (detected); −, test negative (not detected); β-hCG, beta-human chorionic gonadotropin; CRL, crown–rump length; MA, maternal age; NT, nuchal translucency thickness; PAPP-A, pregnancy-associated plasma protein A.

Table 2 Contingency table for detection of aneuploidy by PIA Fetal Database® (PIA) software at a cut-off value of 1 : 300

	Aneuploidy		
Test result	*Yes*	*No*	*Total*
Positive	13	105	118
Negative	4	2531	2535
Total	17	2636	2653

$\chi^2 = 208.83$ ($P < 0.0001$).

Table 3 Contingency table for detection of aneuploidy by PIA Fetal Database® (PIA) software at a cut-off value of 1 : 230

	Aneuploidy		
Test result	*Yes*	*No*	*Total*
Positive	13	76	89
Negative	4	2560	2564
Total	17	2636	2653

$\chi^2 = 282.12$ ($P < 0.0001$).

Comparison of test performances

At both cut-off values, the PRC software displayed an inferior test performance in comparison with the established PIA software. Expressed as relative differences, at a cut-off value of 1 : 300, sensitivity, specificity and positive predictive value obtained by PRC were lower by 7.7%, 1.7% and 31.5%, respectively, compared with PIA values. In particular, with an additional 42 false-positive cases, the false-positive rate generated by PRC was 40.0% higher compared with that generated by PIA. At a cut-off value of 1 : 230, sensitivity, specificity and positive predictive value were lower by 7.7%, 1.3% and 32.1%, respectively, compared with PIA values. With an additional 33 false-positive cases, the false-positive rate generated by PRC was 43.4% higher compared with that generated by PIA. The observed differences were highly significant ($P < 0.0001$). The results are demonstrated as a comparison table (Table 6) and two ROC curves (Figure 1).

DISCUSSION

Test performance

Although PRC has been promoted widely in Germany during the last 18 months, this is the first study comparing the test performance of PRC with that of the PIA software developed and used by the FMF.

Table 4 Contingency table for detection of aneuploidy by Prenatal Risk Calculation® (PRC) software at a cut-off value of 1 : 300

	Aneuploidy		
Test result	*Yes*	*No*	*Total*
Positive	12	147	159
Negative	5	2489	2494
Total	17	2636	2653

$\chi^2 = 126.71$ ($P < 0.0001$).

Table 5 Contingency table for detection of aneuploidy by Prenatal Risk Calculation® (PRC) software at a cut-off value of 1 : 230

	Aneuploidy		
Test result	*Yes*	*No*	*Total*
Positive	12	109	121
Negative	5	2527	2532
Total	17	2636	2653

$\chi^2 = 171.36$ ($P < 0.0001$).

Table 6 Comparison of test performance of PIA Fetal Database® (PIA) and Prenatal Risk Calculation® (PRC) software for aneuploidy screening in 2653 first-trimester pregnancies

	PIA		*PRC*		*Relative difference (%)*	
Test performance	*1 : 300*	*1 : 230*	*1 : 300*	*1 : 230*	*1 : 300*	*1 : 230*
Sensitivity (%)	76.5	76.5	70.6	70.6	−7.7	−7.7
Specificity (%)	96.0	97.1	94.4	95.9	−1.7*	−1.3*
False-positive rate (%)	4.0	2.9	5.6	4.1	+40.0*	+43.4*
Positive predictive value (%)	11.0	14.6	7.5	9.9	−31.5*	−32.1*

*Highly significant ($P < 0.0001$).

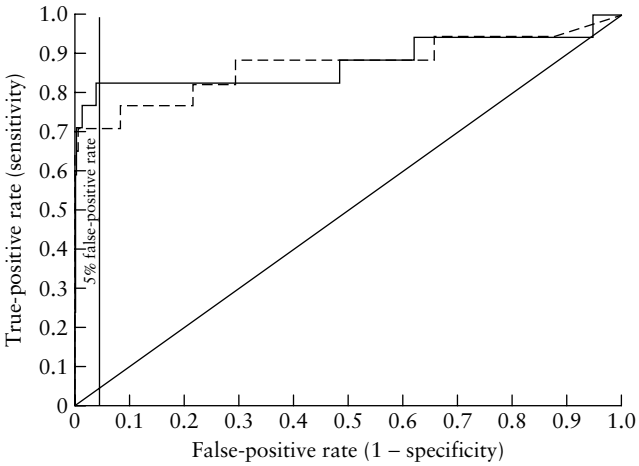

Figure 1 Receiver–operating characteristics (ROC) curves for The Fetal Medicine's PIA Fetal Database® software (——) and the new Prenatal Risk Calculation® (PRC) software (– – –). The PIA ROC curve is closer to the test optimum than is the PRC one. At a false-positive rate of 5%, the vertical line shows a 12% difference in sensitivity between PIA and PRC. The rest of the ROC curves (> 5% false-positive rate) are of negligible clinical relevance. Using PRC, risks that are smaller than 1 : 20 000 are displayed by the software as '< 1 : 20 000'. Therefore, its ROC curve demonstrates a slope for the part of the study population with the smallest risks.

Since this study considered a range of aneuploidies, the sensitivity of detection was relatively low at 76.5% (PIA) and 70.6% (PRC). However, with a sensitivity of 90%, both software programs demonstrated high effectiveness in detecting trisomy 21 cases. For both software programs, altering the cut-off value from 1 : 300 to 1 : 230 did not modify sensitivity, since no true positive case had attained a risk value in between these cut-off values. However, the number of false-positive cases dropped from 105 (cut-off 1 : 300) to 76 (cut-off 1 : 230) for PIA and from 147 (cut-off 1 : 300) to 109 (cut-off 1 : 230) for PRC. In comparison to using PIA, the prospective use of PRC would have generated at both cut-off values a false-positive rate that was at least 40% higher. In accordance with the modus operandi in our hospital, each of the women concerned would have been offered invasive testing unnecessarily. This redundant exposure to risk of complications, including fetal loss, entails ethical, legal and economic implications[13,14].

Cut-off value

In this study, risk calculation with PRC was performed with two recommended cut-off values, 1 : 300 and 1 : 230[3–6]. A cut-off value of 1 : 300 is recommended by a FMF-certified laboratory which participated significantly in the establishment of the database for the PRC software program[7]. In two communications, it recommended lowering the cut-off value from 1 : 230 to 1 : 300, in order not to miss aneuploidy cases in clinical practice[7]. The PIA database uses a cut-off value of 1 : 300, which has been validated on more than 100 000 cases examined prospectively[11] and in a wide range of international studies. The retrospective use of an additional cut-off value of 1 : 230 in our study merely served as a reference value in order to be able to compare the results with PRC. At both examined cut-off values (1 : 300 and 1 : 230), PIA showed a better test performance in comparison to PRC. However, the superiority of PIA is not limited to these cut-off values. The ROC curve of PIA was, in the clinically relevant section (false-positive rate $\leq 5\%$), always closer to the test optimum in comparison to the ROC curve of PRC, irrespective of the chosen cut-off value (Figure 1).

Quality management

The FMF algorithm was developed on the basis of large prospective studies[1] and has been validated and revalidated continuously in comprehensive quality management processes[11,12,15–18]. In recent years, several alternative software programs have been introduced to the market[5,7,19,20]. The PRC software in particular underwent a relatively short development process before its release, thus substantially reducing the options for a methodical evaluation in a prospective setting and/or a systematic comparison against the gold standard. Nevertheless, as in other fields of medicine, prenatal screening methods should not be exempt from extensive research and testing. High standards in measurement techniques, risk calculation, and quality management have already been set by The FMF's first-trimester screening program. Innovative methods will need to prove their utility by passing through a stringent process of quality control, in order to demonstrate their superiority in comparison to the existing gold standard.

In conclusion, in our study cohort, the test performance of PRC was inferior to that of the software developed by

The FMF, the sensitivity for detection of fetal aneuploidy being lower and the false-positive rate higher. Had PRC been employed prospectively in our study, 40% more women examined would have been offered an unnecessary invasive procedure for fetal karyotyping.

REFERENCES

1. Snijders RJM, 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; **351**: 343–346.
2. Palomaki GE, Haddow JE. Maternal serum alpha-fetoprotein, age, and Down syndrome risk. *Am J Obstet Gynecol* 1987; **156**: 460–463.
3. Merz E. 11–14 SSW Screening – Zertifizierte Ultraschalluntersuchung und zertifizierter biochemischer Test in der Frühgravidität. *Ultraschall in Med* 2002; **23**: 161–162.
4. Eiben B, Thode C, Glaubitz R, Merz E. Das neue Ersttrimesterscreening-Programm PRC der FMF-Deutschland – Erste Erfahrungen zur Trisomie-21-Entdeckungsrate. *Frauenarzt* 2007; **48**: 468–470.
5. Merz E. First trimester screening – a new algorithm for risk calculation of chromosomal anomalies developed by FMF Germany. *Ultraschall in Med* 2007; **28**: 270–272.
6. Merz E, Thode C, Wellek S, Alkier A, Eiben B, Hackelöer BJ, Hansmann M, Huesgen G, Kozlowski P, Pruggmaier M. Fetal Medicine Foundation Germany (FMF-D): a new approach to calculating the risk of chromosomal abnormalities with first-trimester screening data (11 + 1 to 14 + 0 weeks). *Ultrasound Obstet Gynecol* 2007; **30**: 542–543.
7. Schmidt P, Scharf A, Hörmansdörfer C, Elsässer M, Hillemanns P. Unterschiedliche Berechnungsmethoden für das Ersttrimester Screening. *Frauenarzt* 2007; **48**: 1089–1092.
8. Spencer K, Souter V, Tul N, Snijders R, Nicolaides KH. Screening program for trisomy 21 at 10–14 weeks using fetal nuchal translucency, maternal serum free β-human chorionic gonadotropin and pregnancy-associated plasma protein-A. *Ultrasound Obstet Gynecol* 1999; **13**: 231–237.
9. Schmidt P, Hörmansdörfer C, Staboulidou I, Hillemanns P, Scharf A. Using 'Degrees of Extremeness ' instead of 'Multiples of Median' in first trimester risk assessment for Down syndrome – an improved method or just a gimmick in face of political motivations? *Arch Gynecol Obstet* 2008; **278**: 119–124.
10. Schetinin V, Fieldsend JE, Partridge D, Coats TJ, Krzanowski WJ, Everson RM, Bailey TC, Hernandez A. Confident interpretation of Bayesian decision tree ensembles for clinical applications. *IEEE Trans Inf Technol Biomed* 2007; **11**: 312–319.
11. Nicolaides KH, Sebire N, Snijders RJM. Nuchal translucency and chromosomal anomalies. In *The 11–14 Week Scan: the Diagnosis of Fetal Abnormalities*, Nicolaides KH, Sebire N, Snijders RJM (eds). Pathenon Publishing: Carnforth, UK, 1999; 3–72.
12. Nicolaides KH, Spencer K, Avgidou K, Faiola S, Falcon O. Multicenter study of first-trimester screening for trisomy 21 in 75 821 pregnancies: results and estimation of the potential impact of individual risk-orientated two-stage first trimester screening. *Ultrasound Obstet Gynecol* 2005; **25**: 221–226.
13. Gilbert RE, Augood C, Gupta R, Ades AE, Logan S, Sculpher M, van Der Meulen JH. Screening for Down's syndrome: effects, safety, and cost effectiveness of first and second trimester strategies. *BMJ* 2001; **323**: 423–425.
14. Georgsson Öhman S, Saltvedt S, Waldenström U, Grunewald C, Olin-Lauritzen S. Pregnant women's responses to information about an increased risk of carrying a baby with Down syndrome. *Birth* 2006; **33**: 64–73.
15. Spencer K, Spencer CE, Power M, Moakes A, Nicolaides KH. One stop clinic for assessment of risk for fetal anomalies: a report of the first year of prospective screening for chromosomal anomalies in the first trimester. *BJOG* 2000; **107**: 1271–1275.
16. Wojdemann 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**: 641–644.
17. Spencer K. Accuracy of Down syndrome risks produced in a first-trimester screening programme incorporating fetal nuchal translucency thickness and maternal serum biochemistry. *Prenat Diagn* 2002; **22**: 244–246.
18. Sheppard C, Platt LD. Nuchal translucency and first trimester risk assessment: a systematic review. *Ultrasound Q* 2007; **23**: 107–116.
19. Schmidt P, Staboulidou I, Soergel P, Wüstemann M, Hillemanns P, Scharf A. Comparison of Nicolaides' risk evaluation for Down's syndrome with a novel software: An analysis of 1463 cases. *Arch Gynecol Obstet* 2007; **275**: 469–474.
20. Schmidt P, Hörmansdörfer C, Pruggmayer M, Schütte C, Neumann A, Gerritzen A, Vaske B, Hillemanns P, Scharf A. Improved test performance for prenatal aneuploidy screening by the novel advanced first-timester screening algorithm – a multicenter study of 10,017 pregnancies. *J Clin Ultrasound* 2008; **36**: 397–402.

