Ultrasound Obstet Gynecol 2003; **21**: 596–601
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/uog.124

Nuchal translucency thickness measurement: repeatability using a virtual ultrasound scanner

V. R. NEWEY*, D. K. NASSIRI*, A. BHIDE† and B. THILAGANATHAN†

**Department of Medical Physics and Bioengineering and †Fetal Medicine Unit, Academic Department of Obstetrics and Gynaecology, St. George's Hospital, London, UK*

KEYWORDS: measurement repeatability; nuchal translucency; phantom; quality assurance audit; ultrasound image; virtual scanner

ABSTRACT

Aim *To use a PC-based virtual ultrasound scanner (VirUS) in the investigation of inter- and intraoperator nuchal translucency (NT) thickness measurement repeatability of experienced ultrasound operators.*

Methods *Realistic fetal ultrasound images of defined NT thickness were simulated with VirUS with emulation of scanner gain and time-gain compensation and gain-dependent echo size changes. A set of 50 images was generated with uniformly distributed NT thickness (range, 1–5 mm at 1-mm intervals) and translucency angle (mean ± standard deviation of ± 2.52° ± 1.85° about the horizontal). Operators* (n = *13) measured NT thickness in the image set on three occasions separated by at least 1 day, giving 150 measurements per operator (total measurements, 1950).*

Results *Inter- and intraoperator repeatabilities were ± 0.41 mm and ± 0.22 mm, respectively (at the 95% confidence level). There were significant correlations between repeatability and mean measured NT thickness (*r = *− 0.72,* P = *0.005 at 4-mm interval), between gain and mean measured NT thickness (*P ≤ *0.002,* n = *8/13) and between gain and repeatability coefficient (*P < *0.01,* n = *6/13).*

Discussion *VirUS provides a consistent NT audit environment and demonstrates the need to both optimize repeatability vs. mean measured thickness and to set gain consistently. The technique has potential in operator training. Copyright © 2003 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

The measurement of nuchal translucency (NT) thickness in ultrasound images of the fetus at 11–14 weeks' gestation is used, in conjunction with maternal age, as an indicator of the risk of chromosomal defects[1,2]. Nuchal translucency thickness is measured manually by placing the ultrasound scanner cursors on the edges of the translucency echoes in the B-mode sagittal plane. However, this results in subjective measurements due principally to the limitations of vision and gain-dependent error in the scanner[3,4]. The eye has a non-linear response to perceived brightness changes and can detect only one or two dozen different brightness levels when focusing on small sections of a complex image[5]. This makes for better perception of brightness changes at relatively low rather than high brightness levels and for difficulty in detecting subtle brightness changes near the peaks of bright echoes. It is customary, therefore, to use the higher contrast trailing-to-leading edges of translucency echoes for NT measurement. However, the trailing and leading edges have asymmetric rise and fall times, which are processed differently by the scanner to give a perceptible variation in echo size with gain and consequent measurement dependence[3,4]. The problems are compounded by other factors, such as fetal movement, positioning of the neck or an improper scan plane, and variation in display brightness, contrast and ambient lighting. These limitations contribute to the reported NT thickness measurement repeatability coefficients of ± 0.88 mm[6] and ± 0.62 mm[7] for interoperator and ± 0.70 mm[6], ± 0.54 mm[7] and ± 0.22 mm[8] for intraoperator measurements. It is important for measurement quality assurance, therefore, to audit operators at regular intervals.

This paper describes a study of NT thickness measurement repeatability using a PC-based virtual

Correspondence to: Prof. D. K. Nassiri, St. George's Hospital, Department of Medical Physics and Bioengineering, Blackshaw Rd., Tooting, London SW17 OQT, UK (e-mail: nassiri@sghms.ac.uk)

Accepted: 13 February 2003

ultrasound scanner, VirUS, that was designed to provide a consistent and quantitative NT audit environment.

METHODS

Realistic fetal images were simulated with VirUS[9] using ultrasound properties of soft tissue and tissue boundaries derived from a real fetal ultrasound image (Figure 1a). The ultrasound properties were obtained by reducing individual echoes in the fetal image to points[10] (Figure 1b and 1c) forming a fetal structure image on which all subsequent simulations were based. The resulting points were at the same locations and of the same amplitudes as were the peaks of echoes in the fetal image. The brightness of points was inversely corrected from the scanner settings to give the actual backscattering of different structures. The points in the structure image were edited to give a straight, parallel edge NT (Figure 1c) with thickness defined by adjusting the vertical separation of the rows of points that represented the translucency edges. The edited structure image was scanned by a simulated ultrasound beam of known axial and lateral resolution[11] and the resulting radio frequency data were processed to give a B-mode image with user management of gain and time-gain compensation (TGC). An example of the final simulation output is shown in Figure 2 demonstrating an increased NT thickness compared to Figure 1a.

Fifty realistic fetal images with uniformly distributed NT thickness were generated for the study. Translucency thickness ranged between 1 and 5 mm at 1-mm intervals (constrained by the image resolution), with 10 images at each interval. A mean ± standard deviation clinical range of NT angles of $\pm 2.52° \pm 1.85°$ about the horizontal was included for greater clinical realism. Additionally, a random amplitude artifact was placed on one translucency edge to simulate slight narrowing over a portion of the nuchal area. The artifact gave some variation in translucency appearance in each image over approximately half of the defined thickness region, whilst leaving a well-defined portion of adequate length for measurement. This required a greater degree of concentration from the operators than would be necessary with a consistent translucency appearance. The 50 images were shuffled to produce a final test sequence common, apart from gain and TGC settings, to all operators.

Thirteen operators experienced in NT thickness measurement participated in the study and received a VirUS training session before starting the study proper. Each operator measured one image set per day on three different occasions separated by at least 1 day, giving 150 measurements per operator and a total of 1950. The scanner gain was set to zero on the presentation of each image and NT thickness was measured after operators had readjusted the gain as required. The TGC controls were reset once at the start of each test session. Operators played no role in generating or manipulating images other than to adjust gain and TGC and place measurement cursors. A fine cross-hair (PC Cross_m.cur) was used for the cursor and no indication

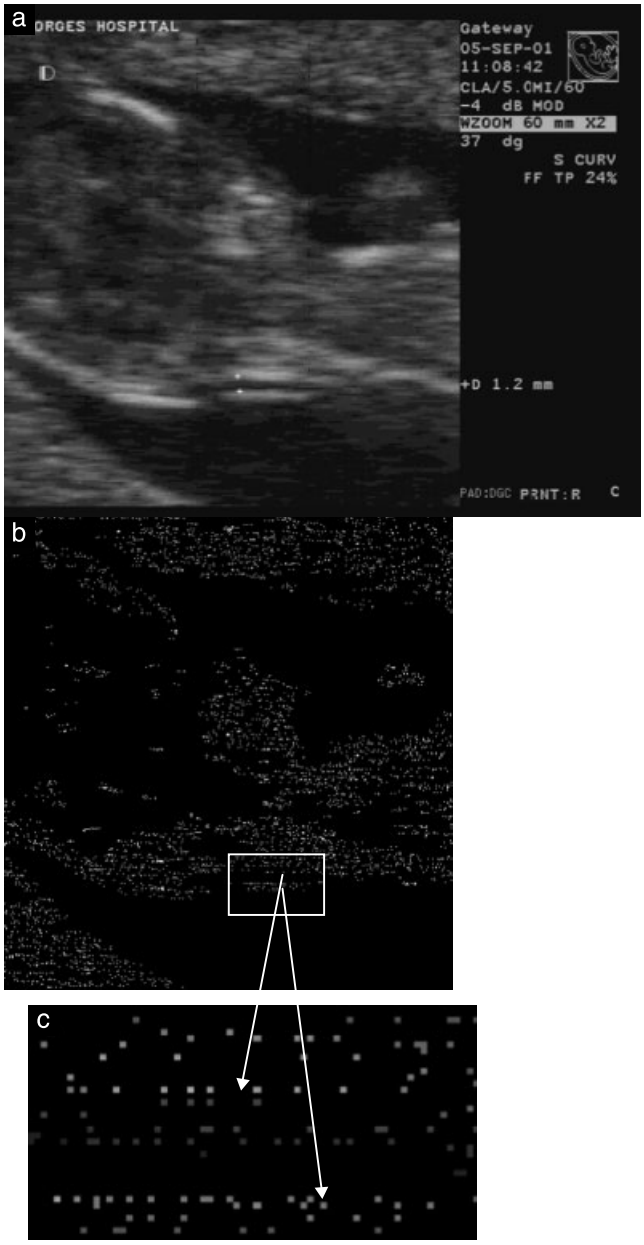

Figure 1 (a) Original clinical image of fetus. (b) Structure image. (c) Enlarged portion of structure image. Note that points in the structure image have been brightened for publication.

of measured distance was available. An 'undo' control was provided to correct erroneous cursor placement. The scanner zoom setting for the images was 123 pixels/cm giving a minimum resolution of 0.08 mm. Pythagoras's theorem was used to calculate the distance between cursor points resulting in higher resolution at some translucency angles.

The intraoperator repeatability was calculated as the standard deviation of the differences between measurements made by the same observer at each measurement interval pooled across all observers[12]. The interoperator repeatability was calculated as the standard deviation of differences between measurements made on the same images by randomly selected operator pairs. Intra- and interoperator repeatabilities and mean

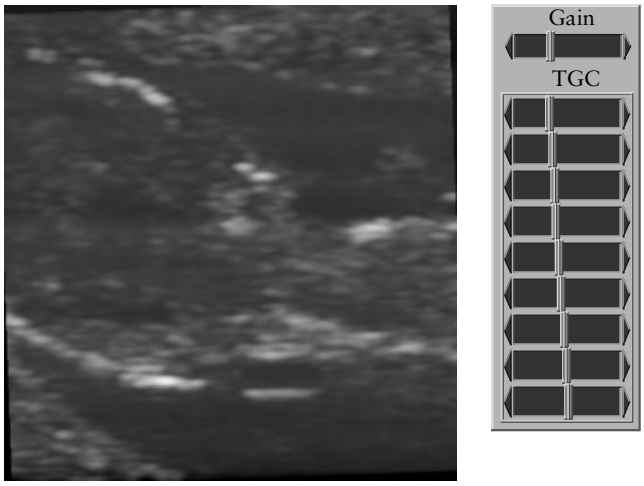

Figure 2 VirUS with gain/TGC control panel and final output from the example of Figure 1 demonstrating an increased NT thickness compared to Figure 1a.

measured NT thickness were calculated as a function of NT thickness. The Pearson correlations for scanner gain vs. mean measured thickness and scanner gain vs. repeatability were calculated for individual operators at each interval.

Software for the project was written using LabVIEW and IMAQ (National Instruments, Austin, TX, USA), and a Mitsubishi Diamond Plus 200 monitor (NEC Mitsubishi Electric Visual Systems Corp., Tokyo, Japan) with fixed, digitally controlled brightness and contrast settings, was used for NT thickness measurements.

RESULTS

All operators successfully completed the study, taking approximately 10–15 min per operator to cover an image set. Three of 1950 measurements in the study were clearly erroneous (cursor spatial coordinates were recorded for all measurements) and were replaced by random selection of the appropriate intraoperator thickness. Inter- and intraoperator repeatabilities were both weakly dependent on NT thickness. Therefore, they have been expressed simply as a mean value for the thickness range, giving repeatabilities at the 95% confidence level of ±0.41 (range, 0.35–0.46) mm for interoperator and ±0.22 (range, 0.21–0.23) mm for intraoperator repeatability (Figure 3). The histogram for intra- and interoperator differences is shown in Figure 4 and their respective distributions are shown in Figures 5 and 6. Figure 7 shows individual operator measurements (mean ± 1.96 × standard deviation) ranked from left to right according to mean measured NT thickness averaged over the 3- and 4-mm intervals. The trend lines at the 4-mm interval in Figure 7 indicate a tendency ($r = -0.72$, $P = 0.005$) for the measurement repeatability to increase as the mean measured NT thickness decreases.

Table 1 shows the individual operator correlation coefficients (Pearson's r) and their significance for scanner gain vs. mean measured NT thickness and gain vs. repeatability at the 4-mm interval. There was a significant correlation between scanner gain and mean measured NT thickness for eight of the 13 operators with $P \leq 0.002$. For gain vs. repeatability there was a significant correlation for six operators with $P < 0.01$ (Table 1).

DISCUSSION

The VirUS technique produces realistic ultrasound images as demonstrated by Figure 2, with simulation of the gain-

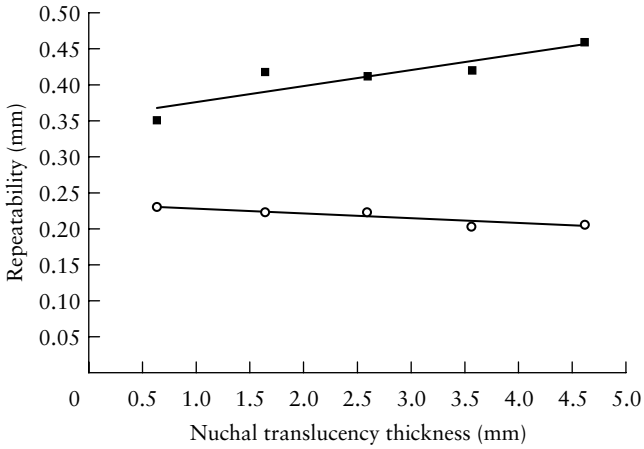

Figure 3 Intraoperator (○) and interoperator (■) repeatability (95% confidence level) as a function of nuchal translucency thickness.

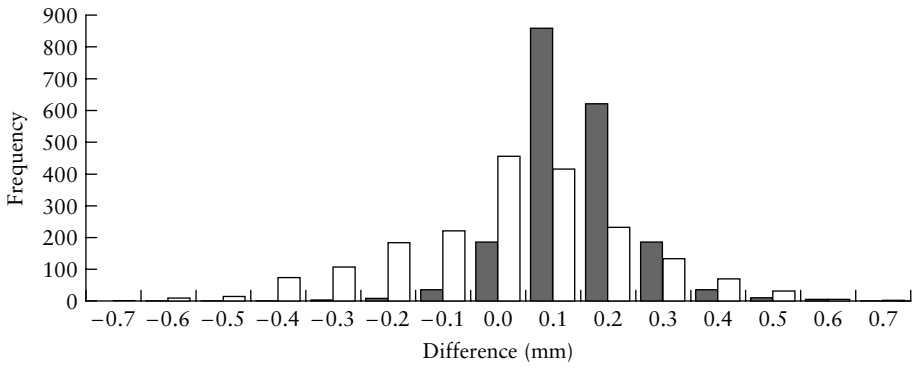

Figure 4 Histogram of intraoperator (■) and interoperator (□) differences.

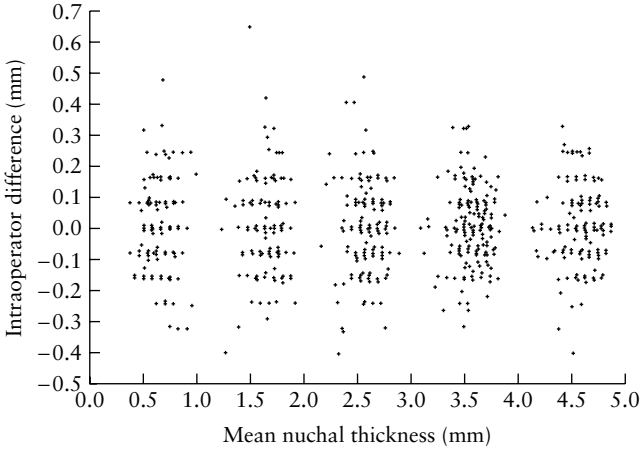

Figure 5 Distribution of intraoperator differences.

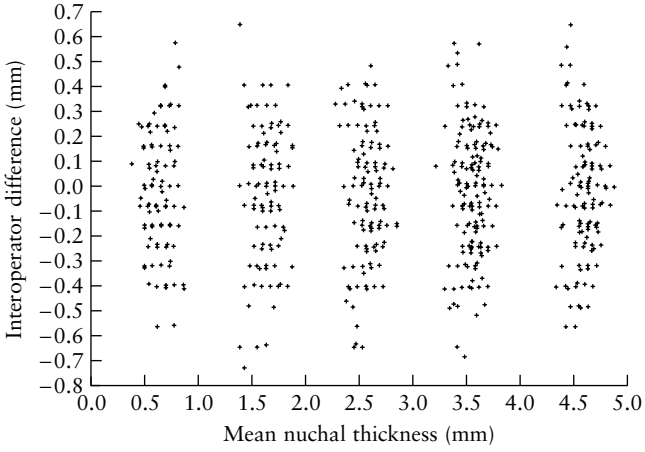

Figure 6 Distribution of interoperator differences.

and TGC-dependent echo size changes that influence the measurement repeatability in a clinical setting. Performance of individual operators is easily distinguished using the method as shown in Figure 7. This figure shows two parameters of operator performance, mean measured NT thickness and repeatability (mean ± standard deviation × 1.96). The trend lines at the 4-mm interval (similar for all NT thickness intervals) indicate a tendency for the measurement repeatability to increase as the mean measured thickness decreases ($r = -0.72$, $P = 0.005$). The mean measured NT thickness parameter is directly related to the corresponding mean risk of chromosomal defects predicted by a given operator. If this parameter is relatively high i.e. when the cursors are positioned very close to echo peaks, the measured values will be overestimated while the repeatability will improve. Therefore, the optimum relationship between these parameters must be established. In order to do so the ideal NT thickness for each interval based on established measurement protocols (the known NT thickness intervals are based on echo peak-to-peak distances) must first be decided by expert judgement. An example is shown for the 3-mm interval by the horizontal dashed line in Figure 7 where NT thickness based on the established protocol is measured as 2.6 mm. Measurements above and below such calibration levels represent over- and underestimation, respectively, relative to established chromosomal risk assessment.

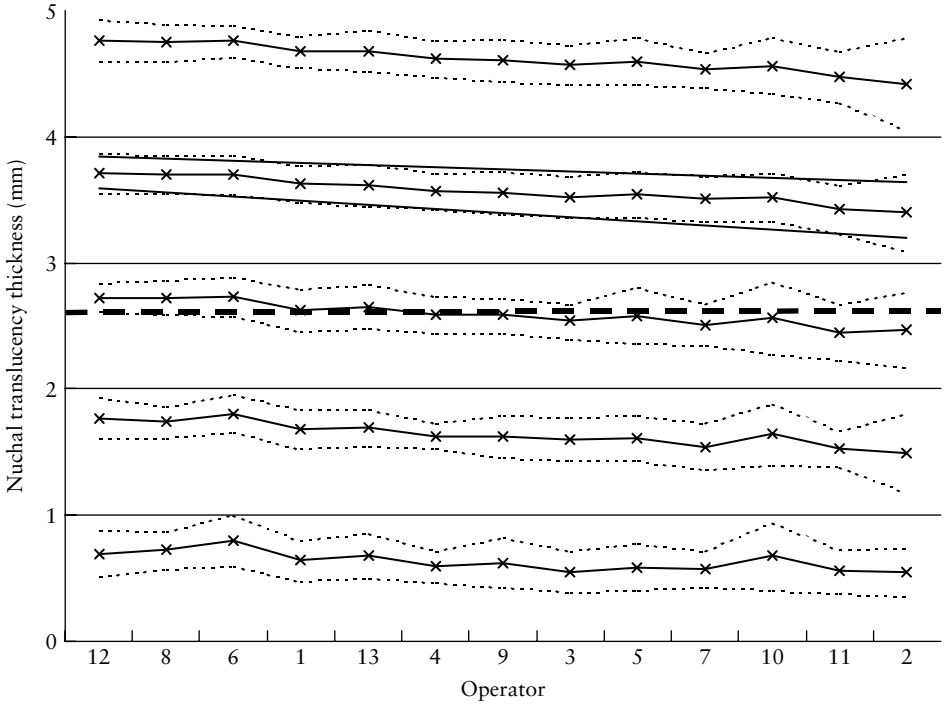

Figure 7 Individual operator performance (mean ± 1.96 standard deviations) for 1–5-mm test intervals. The operators are ranked from left to right, according to the mean measured NT thickness, averaged over the 3- and 4-mm intervals. Repeatability trend lines, an example of which is shown at the 4-mm interval, were similar for other intervals and have been omitted for clarity. The horizontal dashed line represents the conventional trailing-to-leading edge ideal measurement for the simulation at the 3-mm interval.

Table 1 Correlation between gain and mean measured thickness and between gain and repeatability for individual operators at the 4-mm interval

	Gain correlated with			
	Mean thickness		*Repeatability*	
Operator	r*	P	r*	P
1	− 0.68	0.000	0.61	0.000
2	− 0.06	0.739	0.60	0.001
3	− 0.60	0.000	0.37	0.045
4	− 0.62	0.000	0.59	0.001
5	− 0.71	0.000	0.40	0.027
6	0.10	0.616	0.24	0.200
7	− 0.61	0.000	0.55	0.002
8	− 0.60	0.001	0.51	0.004
9	0.01	0.986	0.35	0.055
10	− 0.58	0.001	0.33	0.078
11	0.13	0.470	0.15	0.437
12	− 0.24	0.203	0.49	0.006
13	− 0.55	0.002	0.45	0.012

*Pearson's *r*.

There was a significant correlation between scanner gain and mean measured thickness for eight of the 13 operators with $P \leq 0.002$, and for gain vs. repeatability coefficient with $P < 0.01$ for six operators (Table 1). Scanner gain changes cause slight variation in echo size, the magnitude of which depends upon several factors, such as the scanner process setting (e.g. logarithmic, s-curve) and axial pulse shape. Hence, in clinical practice the mean measured NT thickness and repeatability may exhibit greater scanner gain dependence. It is therefore important to set gain consistently and to standardize variables such as scanner process setting.

The inter- and intraoperator repeatabilities of ± 0.41 mm and ± 0.22 mm (95% confidence level), respectively, in this study are approximately half those reported by Pandya *et al.*[7] (± 0.62 mm and ± 0.54 mm) and Pajkrt *et al.*[6] (± 0.88 mm and ± 0.70 mm). The intra-operator results are more compatible with those obtained in the transvaginal ultrasound study of Braithwaite and Economides[8] who cite better image clarity and resolution of transvaginal scanning, with an intraoperator repeatability of ± 0.22 mm compared with ± 0.40 mm for transabdominal scanning. VirUS has a consistent image quality with relatively mild gain compression characteristics and all simulations were based on the same fetal B-mode image. This gave consistent scan orientation with little variability in translucency echo quality other than that caused by VirUS gain settings and the small random artifact. Differences in study designs could also contribute to the lower repeatabilities . This study had a uniformly distributed sample, whereas studies cited here[6–8] had few samples > 2 mm in thickness, and 13 experienced operators were tested here, compared with one[8], four[7] and five[6]. Repeatability in this study was relatively independent of NT thickness, indicating that operators took similar care over the clinical range. However, in clinical practice access to the scanner NT thickness readout may influence repeatability around the critical NT thickness threshold of 3 mm, at which some operators subconsciously minimize patient risk.

Current NT thickness measurement repeatability audits are conducted under variable conditions and with skewed NT thickness distributions[6–8]. VirUS simulates fetal images of accurately defined NT thickness, uniform distribution and with gain-dependent echo size. In addition to the repeatability the mean under- or overestimate of chromosomal risk can be calculated, as the NT thickness measurements in VirUS simulations can be related to established risk data. Operator repeatability is more easily isolated from the effects of other variables, such as fetal position, that may influence the study accuracy. VirUS can detect a high measurement repeatability sooner than is possible with clinical audits, due to the time required to collect an adequate clinical sample. Clinical audits can create considerable resource management problems. VirUS, however, requires no patient or scanner access and audits can be conducted at the operator's convenience with tests performed under consistent conditions. The method provides a reasonably realistic simulation that can be refined, for example by increasing the range of translucency quality and dissimilar fetal images, and by simulating fetal movement to assess scan plane acquisition. The technique has potential as a quantitative training aid in NT thickness measurement.

REFERENCES

1. Nicolaides KH, Azar G, Byrne D, Mansur C, Marks K. Fetal nuchal translucency: ultrasound screening for chromosomal defects in the first trimester of pregnancy. *BMJ* 1992; **304**: 867–869.
2. Pandya PP, Kondylios A, Hilbert L, Snijders RJM, Nicolaides KH. Chromosomal defects and outcome in 1015 fetuses with increased nuchal translucency. *Ultrasound Obstet Gynecol* 1995; **5**: 15–19.
3. Li S, McDicken WN, Hoskin PR. Blood vessel diameter measurement by ultrasound. *Physiol Meas* 1993; **14**: 291–297.
4. Montauban van Swijndregt AD, The SH, Gussenhoven EJ, CT, Rijsterborgh H, de Groot E, van der Steen AF, Bom N, Ackerstaff RG. An *in vitro* evaluation of the line pattern of the near and far walls of carotid arteries using B-mode ultrasound. *Ultrasound Med Biol* 1996; **22**: 1007–1015.
5. Gonzalez RC, Wintz P. *Digital Image Processing*. Addison Wesley: Reading, MA, 1987.
6. Pajkrt E, Mol BWJ, Boer K, Drogtrop AP, Bossuyt PM, Bilardo CM. Intra- and interoperator repeatability of the nuchal translucency measurement. *Ultrasound Obstet Gynecol* 2000; **15**: 297–301.
7. Pandya PP, Altman DG, Brizot ML, Petterson H, Nicolaides KH. Repeatability of measurement of fetal nuchal translucency thickness. *Ultrasound Obstet Gynecol* 1995; **5**: 334–337.
8. Braithwaite JM, Economides DL. The measurement of nuchal translucency with transabdominal and transvaginal sonography – success rates, repeatability and levels of agreement. *Br J Radiol* 1995; **68**: 720–723.
9. Newey VR, Nassiri DK. Analysis of variability in visual B-mode Image evaluation; application of a virtual scanner and phantom. *Eur J Ultrasound* 2000; **13**: 1S.

10. Newey VR. *Classical & neural network analysis of ultrasound images: estimation of fat & fibrosis content in diffuse liver disease* MPhil thesis. Kingston University, London, 1996.
11. Nassiri DK, Nicholas D, Hill CR. B-scan texture classification: a study using physical and theoretical models. In *3rd meeting of the World Federation for Ultrasound in Medicine & Biology*, Lerski RA, Morley P (eds). Pergamon Press: Oxford, 1983; 133–139.
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–310.

