Monitoring Quality Control of Nuchal Translucency

Howard Cuckle, MSc, DPhil

KEYWORDS

• Nuchal • Screening • Down syndrome • Quality • Prenatal

Nuchal translucency (NT) is by far the most discriminatory marker of fetal Down syndrome available today and is the main component of the most effective multimarker screening strategies. However, the quality control of ultrasound markers is more difficult than that of maternal serum markers, and this leads to practical difficulties. This article discusses the importance of maintaining the quality of NT and the different ways of achieving this quality.

MAIN SCREENING MARKERS

Typically, screening markers have considerable overlap in the distribution of results between affected and unaffected individuals. The potential utility in the screening of a given marker depends on the extent of separation between the 2 distributions, which can be expressed as the absolute difference between the means of the distribution divided by the average standard deviation for the 2 distributions, a form of Mahalanobis distance.

The levels of all commonly used Down syndrome screening markers change with gestation: NT, maternal serum pregnancy-associated plasma protein (PAPP)-A, α-fetoprotein (AFP), and unconjugated estriol (uE_3) increase steadily; human chorionic gonadotropin (hCG) and the free β subunit of hCG decrease rapidly to a plateau; inhibin-A decreases to a nadir and increases thereafter. To allow for these changes with gestation, marker levels are expressed in multiples of the gestation-specific median for unaffected pregnancies, derived by regression. Early ultrasonography studies of NT did not allow for gestation at all, but levels are now being reported in either multiples of median (MoMs) or deviations from the gestation-specific normal median (delta-NT).

Unlike NT,all serum markers have a negative correlation between the MoM level and maternal weight. This negative correlation is largely because of dilution; a fixed mass of chemical produced in the fetoplacental unit is diluted by a variable volume in the

Department of Obstetrics and Gynecology, Columbia University Medical Center, 622 West 168th Street, New York, NY 10032, USA
E-mail address: hsc2121@columbia.edu

Clin Lab Med 30 (2010) 593–604
doi:10.1016/j.cll.2010.04.012 **labmed.theclinics.com**
0272-2712/10/$ – see front matter

maternal unit. It is standard practice to adjust for this dilution by dividing the observed MoM by the expected value for the maternal weight derived by regression. Many centers also adjust serum MoMs, but not NT, to allow for maternal smoking and ethnicity. The levels of both hCG isoforms are reduced on an average in smokers, and there is a reduction of similar magnitude in PAPP-A levels, whereas inhibin-A levels are increased to an even greater extent. Adjustment is achieved by dividing the observed MoM by the average value reported in the literature among smokers or non-smokers, as approapriate. In women of African Caribbean origin or in African American women, levels of hCG isoforms are increased, whereas those of AFP and inhibin-A are decreased; PAPP-A level is markedly increased in African Caribbeans but not to the same extent in African Americans. In women of South Asian origin, uE_3 and total hCG levels seem to be somewhat higher than those in Caucasian women. In ethnically homogeneous populations, there is no need to make adjustments because the normal median reflects the local ethnicity. In an ethnically mixed population with large enough minorities, MoMs can be calculated with ethnic-specific medians or the observed MoM can be divided by a factor derived from the average in published studies for different ethnic groups, taking account of the local ethnic mix.

MOST DISCRIMINATORY SINGLE MARKER

To calculate the Mahalanobis distance, the most reliable estimates of means and standard deviations are from meta-analyses of all the published literature, as are the correlation coefficients between markers. The advantages of meta-analysis are that it produces the most robust estimate of the mean and by combining the results from a wide range of centers, it reflects the average experience likely to be achieved in practice. Parameters from a single study are subject to considerable sampling error because even the largest study to date includes no more than about 100 affected pregnancies. Nonintervention studies produce estimates of the means for cases that present at term. Intervention studies introduce viability bias that will skew the results toward the extreme. This bias arises because a proportion of those with extreme marker levels who have a termination of pregnancy would have been destined to miscarry anyway, whereas nonviable affected pregnancies with normal screening results will not be known to the investigators.

Table 1 shows the Mahalanobis distance for the commonly used markers, according to gestation, based on published meta-analyses.[1] NT is by far the single best individual marker, followed by PAPP-A, which is the most discriminatory serum marker, although the Mahalanobis distance declines with gestation. In contrast, the discriminatory power of free β-hCG increases with gestation; intact hCG is less discriminatory than free β-hCG at all gestations, particularly before 13 weeks when it is a poor marker. Inhibin-A is the best second-trimester marker, whereas AFP and uE_3 are much less discriminatory than the hCG isoforms or inhibin-A.

MULTIMARKER SCREENING

None of the individual markers is discriminatory enough to stand alone; a Mahalanobis distance of at least 3 would be required for that. This consideration has led to the development of several multimarker tests based on the estimation of risk for Down syndrome from the marker profile. This estimation is done by modifying the maternal age-specific risk by likelihood ratio (LR) derived from the marker profile. The LR is the relative height of the theoretical marker distribution in Down syndrome compared with that of unaffected pregnancies. Multivariate log Gaussian distributions seem to fit the data well, but some investigators have proposed other distributions for NT.

Table 1
Commonly used Down syndrome screening markers and their Mahalanobis distance according to gestation

Marker	Gestation (wk)	Mahalanobis Distance
NT	11	2.02
	12	1.87
	13	1.65
PAPP-A	10	1.31
	11	1.14
	12	0.90
	13	0.61
Free β-hCG	10	0.76
	11	0.94
	12	1.05
	13	1.11
	14–18	1.33
hCG	10	0.05
	11	0.32
	12	0.68
	13	1.14
	14–18	1.15
Inhibin-A	14–18	1.12
AFP	14–18	0.79
uE_3	14–18	0.83

The Fetal Medicine Foundation (FMF) has promoted the use of an empiric distribution of NT values,[2] but this has been criticized on the grounds that it is likely to overfit the initial data set on which it was based. Moreover, because empiric distribution does not lend itself to a simple statistical description, non-FMF screeners have not had access to software using it. More recently, the FMF has moved to a Gaussian approach, albeit using 2 sets of distributions for Down syndrome pregnancies in which proportions differ according to gestational age (the so-called mixture model).[3] It remains to be seen if this model improves on a simple Gaussian approach.

MULTIMARKER TESTS INCLUDING NT

Until recent years, most experience with Down syndrome screening was in the second trimester. The best test at that period was the so-called quad test, which uses 4 markers: intact hCG or free β-hCG in combination with AFP, uE_3, and inhibin-A. However, better performance is obtainable using first-trimester marker combinations, and in this period, termination of pregnancy, if required, is safer, is more acceptable to religious minorities, is less traumatic, and provides earlier reassurance.

The best first-trimester results are obtained using NT in combination with PAPP-A and either total hCG or free β-hCG, the so-called combined test. There is an important practical constraint influencing the design of such policies, namely, the results of a scan can be reported to the patient immediately, whereas a serum test result is not usually available for several days. The reason for the delay is that biochemical assays are normally done in batches, which, to avoid unnecessary expense, include about 50 to 100 samples. However, new techniques that allow single samples to be tested economically and results to be available in an hour have been developed. This means that if the test equipment is installed close to the ultrasound unit,

combined serum test and ultrasonography results can be reported together (sometimes known as OSCAR, one-stop clinic for the assessment of risk). Concurrent screening can also be performed without such equipment, provided a blood sample is obtained a few days before the scheduled scan appointment and arrangements are made to ensure that the serum MoMs are available for risk calculation as soon as the NT is measured (sometimes known as IRA, instant risk assessment).

The combination of first- and second-trimester serum markers yield better results than the combined test. One approach is to measure all markers when they are most discriminatory, that is, to measure NT and PAPP-A in the first trimester but to delay hCG or free β-hCG measurement until the second trimester with other quad markers.[4] This 6-marker combination, known as the integrated test, requires nondisclosure of any intermediate risk based on the levels of NT and PAPP-A. Some regard the nondisclosure to be unethical or at least impractical because of the difficulty for professionals to not act on intermediate findings that would of themselves be abnormal, particularly the NT. Furthermore, any increase in detection is paid for by sacrificing early diagnosis and reassurance. Alternative 2-stage 7-marker strategies have been suggested to overcome these limitations. One approach is the stepwise sequential test in which the first stage is the same as in the combined test, and women with risks less than the cutoff are offered the same second-trimester markers as the in the quad test, with the final risk based on all markers.[5] The first-stage cutoff risk is much higher than usual for the combined test. The contingent test is similar except that only women who are at borderline risk after the first stage are offered the second-stage markers.[6]

MODEL PREDICTIONS

The performance of the different tests in terms of detection rate (DR), which is the proportion of affected pregnancies referred for invasive prenatal diagnosis, and the false-positive rate (FPR), which is the proportion of unaffected pregnancies referred, is predicted from statistical models. Two widely used methods have been adopted: numerical integration and Monte Carlo simulation. Numerical integration uses the theoretical log Gaussian distributions of each marker in Down syndrome pregnancies and unaffected pregnancies. The theoretical range is divided into several equal sections, thus forming a grid in multidimensional space. The Gaussian distributions are then used to calculate for each section (square for 2 markers, cube for 3 markers, and so on) the proportion of Down syndrome pregnancies and unaffected pregnancies and the LR in that section. These values are then applied to a specified maternal population. At each maternal age, the number of Down syndrome pregnancies and unaffected pregnancies is estimated from the age-specific risk curve. The distributions of risks are then calculated from the grid values. Monte Carlo simulation also uses the Gaussian distributions, but instead of rigid summation over a fixed grid, it uses a random sample of points in multidimensional space to simulate the outcome of a population being screened.

The model predictions are highly dependent on the maternal age distribution, and to allow comparison between tests a standard population is used; in this article, it is a Gaussian distribution with mean age 27 years and standard deviation 5.5 years.[7] The relative benefits of different tests can be judged by fixing the FPR (eg, 1% or 5%), and the practical implications of changing test are seen by fixing the risk cutoff (eg, 1 in 250 at term).

Table 2 shows the model predictions for NT alone and for combined tests, according to gestation and the level of hCG isoform. The DR for a fixed FPR declines with

Table 2
Model-predicted performance: NT alone and the Combined test

		DR for FPR		1 in 250 Cutoff Risk[a]	
Test	**Gestation (wk)**	**1% (%)**	**5% (%)**	**DR (%)**	**FPR (%)**
NT alone					
	11	64	77	73	2.9
	12	62	75	70	2.7
	13	57	71	66	2.8
Combined test					
NT, free β-hCG & PAPP-A	11	74	87	81	2.4
	12	72	84	79	2.5
	13	66	80	75	2.8
Combined test					
NT, hCG & PAPP-A	11	71	84	79	2.5
	12	70	83	77	2.5
	13	67	81	76	2.7

[a] At term.

advancing gestation, but even with NT alone at 13 weeks, the rate is comparable with that of the quad test, which has a predicted DR of 71% for a 5% FPR using free β-hCG and 67% for intact hCG. The combined test performs considerably better than NT alone at all gestations. The use of free β-hCG improves detection compared with total hCG when a combined test is performed before 13 weeks. Despite this, another modeling exercise claims that there is no material difference in the DR of combined test according to hCG isoform.[8] This model used parameters from the First- and Second-Trimester Evaluation of Risk (FaSTER) trial together with hCG levels based on the retrospective assaying of stored serum samples from only 79 Down syndrome pregnancies and 395 unaffected pregnancies. Larger data sets are needed before concluding that there is no difference.

Table 3 shows the predicted rates for the integrated, stepwise sequential, and contingent tests. The integrated test is predicted to increase detection for a fixed 5% FPR by more than 10%. However, the stepwise sequential and contingent tests have a predicted rate comparable with the integrated test. A retrospective analysis of data from the FaSTER trial has reached the same conclusion.[9] Marker levels from women who completed the first and second stages of the trial—intervention was in the second stage—were used to calculate risks of Down syndrome. For the contingent test, DR was 91% and FPR was 4.5%; the initial DR was 60%, and the initial FPR was 1.2%, and 23% had borderline risks. Stepwise testing had a DR of 92% and an FPR of 5.1%; integrated screening had a DR of 88% and an FPR of 4.9%. These DRs are lower than expected from **Table 3** because some early detected cases, particularly those with cystic hygromas, were excluded. From modeling and the FaSTER results, the practical conclusion is that given the human and practical benefits and lower costs, the contingent test should be the across-trimester strategy of choice.

CONSEQUENCES OF NT ERRORS

The models assume that the parameters in the risk calculator correspond to the distribution of marker levels in the population being screened. Specifically, the mean NT level in the unaffected pregnancies being screened is assumed to be 1.00 MoM,

Table 3
Model-predicted performance: integrated, stepwise sequential, and contingent tests

Test[b]	Gestation (wk)	DR for FPR 1% (%)	DR for FPR 5% (%)	1 in 250 Final Cutoff[a] DR (%)	1 in 250 Final Cutoff[a] FPR (%)
Integrated					
NT, PAPP-A & Quad[a]	11	85	93	87	1.6
	12	83	92	86	1.7
	13	79	89	84	2.0
Stepwise sequential					
NT, free β-hCG & PAPP-A; Quad if negative	11	85	94	89	1.7
	12	84	93	88	1.9
	13	80	91	86	2.1
NT, hCG & PAPP-A; Quad if negative	11	86	94	89	1.6
	12	83	92	87	1.8
	13	80	91	85	2.0
Contingent					
NT, free β-hCG & PAPP-A; Quad if borderline	11	85	92	88	1.6
	12	83	91	86	1.7
	13	79	88	84	1.9
NT, hCG & PAPP-A; Quad if borderline	11	84	90	86	1.4
	12	82	89	85	1.6
	13	79	88	83	1.8

[a] At term.
[b] Quad: AFP, uE_3, free β-hCG, and inhibin-A; initial cutoff for stepwise sequential and contingent, 1 in 50 at term; borderline cutoffs for contingent, 1 in 50–1500 at term.

and in the earlier mentioned modeling, the $\log_{10}$ MoM standard deviations are 0.132, 0.116, and 0.112 at 11, 12, and 13 weeks, respectively; in Down syndrome pregnancies the means are 2.30, 2.10, and 1.92 MoM at each week, and the standard deviation is 0.229.

Table 4 shows what would happen to the Down syndrome risk based on NT alone in a 25-year-old woman, if the accuracy of the NT at 12 weeks' gestation in the screened population is altered by systematically shifting the mean up or down by 10%. At this age, the practical consequences are in NT values that are greater than about 1.50 MoM. For example, using the risk calculator, which assumes complete accuracy and average precision, a value of 1.80 MoM would correspond to a term risk of 1 in 250, exactly on the cutoff used in many countries. But if the operator is overmeasuring by 10%, the true risk would only be 1 in 550, whereas if there is a 10% undermeasurement the risk will be 1 in 95.

Table 4 also shows what would happen in the same circumstances, if the precision was to be changed by a 0.020 reduction or an increase in the standard deviation of $\log_{10}$ MoM in unaffected pregnancies. The corresponding parameter for Down syndrome pregnancies is obtained by reducing or increasing the variance by the same amount it has been changed in unaffected pregnancies. The same observed value of 1.80 MoM with an apparent 1 in 250 term risk would have a much higher 1 in 95 risk, if the operator had greater-than-average precision and a much lower risk, 1 in 430, if the precision was low.

Table 5 shows the consequences of these changes on the model-predicted DR and FPR at 12 weeks' gestation for NT alone and the combined test. As might be expected, a change in accuracy shifts the DR and FPR in the same direction. A change

Table 4
Risk for Down syndrome (1 in x at term) in a 25-year-old woman according to NT level at 12 weeks and the quality of the local NT distribution

NT Level (MoM)	Average Accuracy & Precision	Accuracy[a]		Precision[b]	
		−10%	+10%	−0.02	+0.02
0.50	3700	5700	2300	1300	6000
0.60	7200	9000	5300	4600	8200
0.70	9600	10,000	8300	8900	9000
0.80	10,000	9100	9900	11,000	8500
0.90	9000	7200	9900	11,000	7300
1.00	7200	5100	8800	9100	5900
1.10	5300	3400	7200	6400	4500
1.20	3700	2200	5500	4100	3400
1.30	2500	1300	4000	2400	2400
1.40	1600	800	2800	1300	1800
1.50	1000	470	1900	720	1200
1.60	650	280	1300	370	870
1.70	400	160	840	190	610
1.80	250	95	550	95	430
1.90	150	55	350	45	300
2.00	95	30	230	25	210
2.10	55	19	150	12	140
2.20	35	11	95	10	100
2.30	20	10	60	10	70
2.40	14	10	40	10	50
2.50	10	10	25	10	35

[a] Change in the median NT MoM.
[b] Change in the $\log_{10}$ standard deviation.

in precision mainly affects the FPR; a tighter distribution reduces the FPR, whereas a broader distribution increases it. The effect of changes in accuracy or precision is less for the combined test than for NT alone because any loss of performance is cushioned to some extent by the other markers in the combined test. This effect is even more marked for the integrated, stepwise sequential, and contingent tests because of more markers, and hence, this is another argument for adopting one of these 2-stage approaches.

EVIDENCE FOR SUBOPTIMAL PERFORMANCE

Examination of the results from prospective intervention studies of Down syndrome screening is a means of determining whether or not there is substantial reason for concern about the quality of NT measurement. However, the observed DR in such studies is necessarily an overestimation of the true rate because of the nonviability bias described earlier. To overcome this, an unbiased estimate can be derived from the observed numbers of Down syndrome cases using the formula $(n1 \times p + n2)/(n1 \times p + n2 + n3 \times p + n4)$, where n1, n2, n3, and n4 are the observed numbers of screen detected and terminated, screen detected but not terminated, missed by screening but terminated subsequently, and missed by screening and born cases of

Table 5
Model-predicted performance at 12 weeks according to the quality of the local NT distribution

	Accuracy[a]		Precision[b]	
Average Accuracy & Precision (%)	−10% (%)	+10% (%)	−0.02 (%)	+0.02 (%)
NT alone				
70 & 2.7	63 & 1.2	76 & 5.4	71 & 1.3	70 & 4.7
Combined: NT, free β-hCG, & PAPP-A				
79 & 2.5	73 & 1.4	83 & 4.3	80 & 1.6	78 & 3.8
Combined: NT, hCG, & PAPP-A				
77 & 2.5	72 & 1.3	82 & 4.5	78 & 1.5	77 & 3.9

[a] Change in the median NT MoM.
[b] Change in the $\log_{10}$ standard deviation.

Down syndrome, respectively, and p is the intrauterine survival rate for Down syndrome at the time of prenatal diagnosis.

Twenty-five large second-trimester intervention studies have been analyzed in this way, and the results have been found to be consistent with model predictions.[1] But when the same was done for studies using NT, the results seemed to be suboptimal. For the 6 studies of NT alone that expressed the results in terms of risk, there were a total of 142,000 screened women of whom 643 were observed to have a fetus with Down syndrome. This finding yielded an observed DR of 84%, equivalent to 72% after allowance for bias, and an FPR of 8.4%. When 15 studies of the combined test were analyzed, with a total of 145,000 women including 638 with Down syndrome pregnancies, the observed DR was 89% and the unbiased DR was 81% with an FPR of 5.9%. There have been 4 prospective intervention studies of the integrated test,[10–13] totaling 50,000 pregnancies, 135 with Down syndrome; observed and unbiased DRs of 88% and 85%; and FPR of 2.8%. Some of the shortfall in detection was because of a failure of all women to complete both stages of the screening protocol. The completion rates ranged from 75% to 92%. On the other hand, the largest study also acted on a high NT alone,[11] a type of stepwise sequential protocol. There has so far been only 1 published intervention study of the stepwise sequential test and it was small.[14] The test was performed on 1528 women, and there were only 3 Down syndrome cases, all of which were identified with an FPR of 6.9%. No contingent test results have been published yet.

QUALITY CONTROL METHODS

NT is visualized in the midsagittal section used for crown-rump length (CRL) measurement, and the FMF has published a standardized technique to be adopted for CRL measurement.[15] This technique relates to the position of the fetus, the ultrasound section chosen, the separation of the fetus from the amnion, the placement of the calipers, and the magnification of the image. Various methods have been described for scoring the quality of the image per se.[16–18]

In addition to ensuring that the aforementioned guidelines on measurement are understood by all sonographers taking part in a particular screening program and, if possible, to having senior staff oversee new trainees, it is necessary to carry out epidemiologic monitoring of results.

A direct approach is to compare the observed positive rate, excluding any known cases of aneuploidy, with the expected rate for the maternal age distribution in the

population being screened. The incidence of Down syndrome is not high enough for the observed DR to be a practical indicator of performance.

Using NT alone, a particularly high positive rate may indicate an upward shift in values, a broader spread of results, or both. A low rate could relate to a downward shift in values but could also mean that NT is being measured more precisely than expected. For a combined test or the 2 trimester policies, an excess or a deficit may be contributed to by the biochemical marker distributions. There may be a problem with NT even when the positive rate is consistent with the expected rate, if the biochemical markers are performing particularly well. In these circumstances, indirect indicators of performance are preferable, specifically the median MoM and the standard deviation, on a logarithmic scale, of the MoM values.

This concept is no different from quality assessment of biochemical screening markers. The median MoM value should be calculated on a regular basis for the overall program and for each sonographer. The observed median, excluding only known cases of fetal aneuploidy, is the best estimator of the unaffected mean because it is not subject to distortion by occasional outliers. Similarly, the overall and operator-specific standard deviations of $\log_{10}$ MoM should be calculated. The nonparametric estimator based on the difference between the 90th and 10th percentiles, in $\log_{10}$ MoM, divided by 2.563 is relatively unaffected by outliers.

A value for the median, which is outside the range 0.90 to 1.10 MoM, is a matter of concern. Depending on the number of scans included in the calculation, it is possible to exceed these limits by chance alone, and a statistically significant deviation would be a more compelling evidence of a problem. If deviant results are obtained for an individual or with all operators, some form of retraining will be required. But if this has no effect, one possibility is to use operator- or center-specific normal median curves for calculating MoMs.[19] Such adjustment was used by the Serum, Urine and Ultrasound Screening Study and the FaSTER trial.

For the model prediction in this article, the NT standard deviations were obtained from 4 large prospective studies combined.[2] Because this involved several different centers and operators, the values are necessarily wider than should be obtained for a single operator or even center. From the author's personal experience, the target for an individual operator should be a $\log_{10}$ standard deviation of 0.09 with an acceptable range of 0.07 to 0.11; for a whole center a realistic value might be 0.10, with range 0.08 to 0.12. An individual with an NT value less than the range might have a particularly precise technique, but another possibility is that the different NTs may not be discriminated sufficiently, which could be as serious as measuring imprecisely. The use of MoMs implies multiplicative accuracy, which for biochemical markers is equivalent to having a good recovery in doubling dilutions. One way of assessing accuracy for NT is to observe the rate of change in median NT according to CRL. As the MoM equation is curved, the rate of change is not uniform, but as a guide, the median NT should increase by about one-third for a 20-mm span of CRL at 11 to 12 weeks and by about 10% at 12 to 13 weeks. A shallower increase in an operator with a low standard deviation would be a concern.

FMF has an external quality assessment scheme for NT, which has branches in different countries. Sonographers receive training, initial credentialing, and remediation by sending images and data to the scheme. The Nuchal Translucency Quality Review scheme in the United States also performs a similar function.

Whatever the external scheme involved, the import step is for those who are out of target to be assessed by experienced colleagues in the same center. When biochemical screening was the norm, it was possible for an individual laboratory to manage its own quality, but with the newer tests incorporating NT, the laboratory needs to have

good working relationships with those in each participating ultrasound unit who can take responsibility for NT results. In some settings, this can be logistically difficult.

QUALITY MONITORING RESULTS

It is common for inexperienced sonographers to underestimate NT. In one study of NT quality in 19 trainees, the criterion was the proportion of results less than the normal median derived by experienced sonographers. Only after a minimum of 50 scans were half the results less than the median, and on an average, it took 131 scans.[20] Similar results were found in a Danish training study.[17] Among the inexperienced, an excess of low values is particularly seen at the low end of the NT range,[21] and in these circumstances, a moderately reduced median MoM together with a moderately increased standard deviation may be the grounds for reviewing the individual's technique.

In the initial half year of the BUN (**B**iochemical, **U**ltrasound, **N**uchal translucency) trial, NT quality was assessed using an FMF protocol.[15] Of the 5 sonographers fulfilling training requirements, 4 had NT values on average significantly less than the mean expected by FMF and 1 was significantly greater than the mean based on 23 to 136 images reviewed. In the next half year, the situation had not materially changed based on a further 24 to 153 images reviewed, but when feedback was given on the quality of their images, the next half year saw a considerable convergence of results (41–370 new images).

Undermeasurement is not just a problem for the inexperienced. An audit of 264 sonographers providing results for a large laboratory in Belgium found widespread undermeasurement.[22] One of the sonographers was FMF trained and had a median delta-NT of 0.03 mm but the rest had a median of −0.14 mm.

An insight into the extent of poor performance in routine practice can be found in a study of 14,210 NT scans by 140 sonographers providing at least 50 results for 6 laboratories in the United States.[23] Three epidemiologic indicators were used: median, 0.9 to 1.1 MoM; standard deviation, 0.08 to 0.13; and slope, 15% to 35% per week. Only 56% of operators were within all 3 targets.

In the FaSTER trial, high NT quality was maintained in a 3-level approach. A total of 102 participating sonographers received training, and a minimum of 50 images were assessed by a single external reviewer before active screening began. Thereafter, each sonographer used a checklist to confirm adherence to the protocol and a within-center assessor reviewed all that was imaged. Using center-specific medians, epidemiologic monitoring was performed with median MoM, standard deviation, and slope as the indicators. A recent analysis of these results has shown that despite this intensive review, some 7% of NT measurements were inadequate and changes in the NT measurements occurred over time.[24]

In the Netherlands, a retrospective analysis was performed on 27,738 NT measurements recorded centrally in the National Institute for Public Health and the Environment.[25] A single published MoM curve was used.[26] The 42 sonographers credentialed by FMF got a mean NT value of 0.98 MoM, whereas the remaining 64 got a mean of 0.92 MoM. Of even greater concern was the upward trend in values over the 2-year study period from a mean of 0.86 MoM increasing to 0.96 MoM.

SUMMARY

Current best practice for Down syndrome screening involves the use of NT measurement in combination with maternal serum markers. Differences in the distribution of NT levels between the theoretical values in the risk calculator and the actual practice lead

to changes in performance. The observed positive rate is a direct indicator of performance, but the essential indicators are the median MoM and standard deviation of log MoM. Operators need to be credentialed and monitored by external schemes using this approach. Laboratories should also monitor the sonographers performing NTs as part of tests for which they are responsible.

REFERENCES

1. Cuckle H, Benn P. Multianalyte maternal serum screening for chromosomal defects. In: Milunsky A, editor. Genetic disorders and the fetus: diagnosis, prevention and treatment. 6th edition. Baltimore (MD): Johns Hopkins University Press; 2009. p. 771–818.
2. Spencer K, Bindra R, Nix AB, et al. Delta-NT or NT MoM, which is the most appropriate method for calculating accurate patient-specific risks for trisomy 21 in the first trimester? Ultrasound Obstet Gynecol 2003;22:142–8.
3. 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(4):376–83.
4. Wald NJ, Rodeck C, Hackshaw AK, et al. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). Health Technol Assess 2003;7(11):1–88.
5. Malone FD, Canick JA, Ball RH, et al, First- and Second-Trimester Evaluation of Risk (FASTER) Research Consortium. First trimester or second-trimester screening, or both, for Down's syndrome. N Engl J Med 2005;353(19):2001–11.
6. Wright D, Bradbury I, Benn P, et al. Contingent screening for Down's syndrome is an efficient alternative to non-disclosure sequential screening. Prenat Diagn 2004;24(10):762–6.
7. Cuckle H, Aitken D, Goodburn S, et al, UK National Down's Syndrome Screening Programme, Laboratory Advisory Group. Age-standardisation when target setting and auditing performance of Down syndrome screening programmes. Prenat Diagn 2004;24(11):851–6.
8. Canick JA, Lambert-Messerlian GM, Palomaki GE, et al, First and Second Trimester Evaluation of Risk (FASTER) Trial Research Consortium. Comparison of serum markers in first-trimester Down syndrome screening. Obstet Gynecol 2006;108:1192–9.
9. Cuckle HS, Malone FD, Wright D, et al, FaSTER Research Consortium. Contingent screening for Down syndrome—results from the FaSTER trial. Prenat Diagn 2008;28:89–94.
10. Weisz B, Pandya PP, David AL, et al. Ultrasound findings after screening for Down syndrome using the integrated test. Obstet Gynecol 2007;109:1046–52.
11. Okun N, Summers A, Hoffman B, et al. Prospective experience with integrated prenatal screening and first trimester combined screening for trisomy 21 in a large Canadian urban center. Prenat Diagn 2008;8:987–92.
12. Wald NJ, Huttly WJ, Murphy KW, et al. Antenatal screening for Down's syndrome using the Integrated test at two London hospitals. J Med Screen 2009;16(1):7–10.
13. Rodrigues LC, Ramos-Dias AM, Carvalho V, et al. Evaluation of four years of prenatal screening for aneuploidies in Hospital S. Francisco Xavier using the integrated test. J Med Screen 2009;16(1):46–7.
14. Benn PA, Campbell WA, Zelop CM, et al. Stepwise sequential screening for fetal aneuploidy. Am J Obstet Gynecol 2007;197(3):312, e1–5.

15. 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(4):353–9.
16. Herman A, Maymon R, Dreazen E, et al. Nuchal translucency audit: a novel image-scoring method. Ultrasound Obstet Gynecol 1998;12:398–403.
17. Wøjdemann KR, Christiansen M, Sundberg K, et al. Quality assessment in prospective nuchal translucency screening for Down syndrome. Ultrasound Obstet Gynecol 2001;18(6):641–4.
18. Fries N, Althuser M, Fontanges M, et al. Quality control of an image-scoring method for nuchal translucency ultrasonography. Am J Obstet Gynecol 2007; 196(3):272, e1–5.
19. Logghe H, Cuckle H, Sehmi I. Centre-specific ultrasound nuchal translucency medians needed for Down's syndrome screening. Prenat Diagn 1995;23(5): 389–92.
20. Frey Tirri B, Troeger C, Holzgreve W, et al. Quality management of nuchal translucency measurement in residents. Ultraschall Med 2007;28(5):484–8.
21. Evans MI, Van Decruyes H, Nicolaides KH. Nuchal translucency measurements for first-trimester screening: the 'price' of inaccuracy. Fetal Diagn Ther 2007;22: 401–4.
22. Gyselaers WJ, Vereecken AJ, Van Herck EJ, et al. Audit on nuchal translucency thickness measurements in Flanders, Belgium: a plea for methodological standardization. Ultrasound Obstet Gynecol 2004;24(5):511–5.
23. Palomaki GE, Neveux LM, Donnenfeld A, et al. Quality assessment of routine nuchal translucency measurements: a North American laboratory perspective. Genet Med 2008;10(2):131–8.
24. 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(2):142–6.
25. Koster MP, Wortelboer EJ, Engels MA, et al. Quality of nuchal translucency measurements in The Netherlands: a quantitative analysis. Ultrasound Obstet Gynecol 2009;34(2):136–41.
26. Nicolaides KH, Snijders RJ, Cuckle HS. Correct estimation of parameters for ultrasound nuchal translucency screening. Prenat Diagn 1998;18(5):519–23.