

Quality assessment of routine nuchal translucency measurements: a North American laboratory perspective

Glenn E. Palomaki, BS¹, Louis M. Neveux, BA¹, Alan Donnenfeld, MD², Jo Ellen S. Lee, PhD³, Geraldine McDowell, PhD⁴, Jacob A. Canick, PhD¹, Anne Summers, MD⁵, GERALYN Lambert-Messerlian, PhD¹, Leonard H. Kellner, MS⁶, Arthur Zebelman, PhD⁷, and James E. Haddow, MD¹

Purpose: To assess nuchal translucency measurements that were performed as part of routine prenatal screening for Down syndrome. **Methods:** Collect ultrasound measurements of nuchal translucency and crown rump length provided by individual sonographers over a 6-month period to six North American prenatal screening laboratories, along with the laboratory's nuchal translucency interpretation in multiples of the median. For sonographers with 50 or more observations, compute three nuchal translucency quality measures (medians, standard deviations, and slopes), based on epidemiological monitoring. **Results:** Altogether, 23,462 nuchal translucency measurements were submitted by 850 sonographers. Among the 140 sonographers (16%) who submitted more than 50 observations, 76 (54%) were found to have all three quality measures in the target range. These 140 sonographers collectively accounted for 14,210 nuchal translucency measurements (61%). The most common single measure to be out of range was nuchal translucency multiples of the median, found for 29 of the 140 sonographers (21%). **Conclusion:** Laboratories should routinely monitor the quality of nuchal translucency measurements that are received for incorporation into Down syndrome screening risk calculations and interpretations. When possible, instituting sonographer-specific medians and providing individualized feedback about performance and numbers of women tested offer the potential to yield more consistent and improved performance. **Genet Med 2008;10(2): 131–138.**

Key Words: quality assessment, nuchal translucency, Down syndrome, prenatal screening

Nuchal translucency (NT) is defined as the collection of fluid behind the fetal neck occurring in the first trimester of pregnancy.^{1–3} When measured and interpreted correctly, it is the most discriminatory marker for Down syndrome that has been reported in a routine setting. Univariately, NT measurements can identify about 60% of Down syndrome pregnancies in the first trimester at a 5% false-positive rate.⁴ Detection increases to between 80% and 85%, when NT is combined with first trimester biochemical measurements of pregnancy-associated plasma protein-A (PAPP-A) and human chorionic gonadotropin (the intact, total, or free- β subunit

forms).⁴ When NT measurements are incorporated into the integrated test (NT and PAPP-A measurements in the first trimester) and the quadruple test (α -fetoprotein, unconjugated estriol, human chorionic gonadotropin and dimeric inhibin-A in the second trimester), 85–90% detection is achievable at a false-positive rate of 2% or less.^{5,6}

Acquiring the skill to properly measure NT requires specialized training and oversight.⁷ Recognizing this, several national and international programs have been established to train and qualify sonographers, and also to assess their ability to consistently and accurately quantify the NT thickness.^{8–11} The hands-on performance training is usually evaluated through still images submitted to one, or a panel of, expert sonographers. Several research trials using NT measurements have shown that such training, although necessary, is not sufficient to assure reproducibility of absolute measurements among sonographers. For example, both the Serum, Urine and Ultrasound Screening Study (SURUSS)⁵ and First and Second Trimester Evaluation of Risk (FASTER) study¹² found that use of sonographer-specific reference data (medians) resulted in improved screening performance.

In the current study, information was sought from screening laboratories regarding how they deal with NT measurements

From the ¹Women & Infants Hospital/Warren Alpert Medical School of Brown University, Providence, Rhode Island; ²Genzyme Genetics, Philadelphia, Pennsylvania; ³Genzyme Genetics, Santa Fe, New Mexico; ⁴LabCorp, Research Triangle, North Carolina; ⁵North York General Hospital, Toronto, Canada; ⁶Lenetix Medical Screening Laboratory, Mineola, New York; and ⁷LabCorp, Seattle, Washington.

Glenn E. Palomaki, BS, Department of Pathology and Laboratory Medicine, Women & Infants Hospital/Brown University, 70 Elm Street, 2nd Floor, Providence, RI 02903. E-mail: gpalomaki@ipmms.org.

Disclosure: The authors declare no conflict of interest.

Submitted for publication July 11, 2007.

Accepted for publication October 4, 2007.

DOI: 10.1097/GIM.0b013e3181616bf8

being routinely performed as part of Down syndrome screening. Then, NT and associated ultrasound data were collected and analyzed, with a view to evaluating individual sonographer performance. This was done using the same three epidemiological parameters that have been proven useful in monitoring biochemical markers¹³ as part of routine screening practice. These parameters have already proven successful for monitoring NT measurements as part of research trials.^{5,6} With such data in hand, it should be possible to shape guidelines for assessing NT and, thereby, help prenatal screening laboratories assure that Down syndrome risk estimates that include NT measurements are reliable.

The American College of Medical Genetics Subcommittee on Quality Assurance was charged in 2005 with drafting Technical Standards and Guidelines for Down syndrome Screening that included the interpretation of NT measurements. The working group formed to draft that document identified a gap in knowledge relating to the current state of NT measurements being routinely received at screening laboratories in North America. Members of that working group proposed that data be collected from laboratories that represented both the clinical and academic settings, had catchment areas ranging from local to nationwide, and had varying amounts of experience dealing with NT measurements. This report describes the results of examining that dataset.

MATERIALS AND METHODS

In follow-up to this proposal, members of the working group identified six laboratories satisfying the specific criteria. These laboratories were contacted and asked to provide information about their procedures and practices, as well as sonographer-specific information, for a 6-month time period. Each laboratory requisition slip asked for sonographer identification (e.g., name, certification number). The intent was to identify the ultrasound operator who made the NT observation. Several slips also contained space for a “reading MD” or “supervising radiologist,” but that information was not used in the analysis. It is possible that, in some instances, the supervisor ID was reported, rather than the sonographer ID. It is unlikely that the ordering physician was named, as all laboratories required evidence of the sonographer’s successful training and credentialing. A 6-month timeframe was chosen to represent the longest interval over which it would be reasonable to collect sonographer’s NT data before analyses were performed and assessed for quality. Monitoring over a longer time period (e.g., 1 year) would be slower to identify problems and to document whether corrective actions were effective. Differing time periods were chosen by each laboratory, at least in part because of available institutional review board approvals, ease of data retrieval (archived versus current records), and availability of an existing dataset. De-identified data were sent in Excel (Microsoft Corporation, Redmond, WA) files to be analyzed at Women & Infants Hospital. The institutional review board at Women & Infants Hospital considered the study to be exempt, as the data were de-identified. All laboratories pro-

vided sonographer codes (numeric or alphabetic), NT measurement in millimeters, crown rump length (CRL) in millimeters, the date of the measurement, and the NT multiples of the median (MoM) as reported by the laboratory for clinical purposes. One laboratory provided two datasets; one when NT measurements were just beginning to be routinely interpreted and the other, more recent. One laboratory also provided the center at which the sonographer obtained the measurement. Two laboratories included the maternal age.

Epidemiological monitoring involved computing three quality measures for each sonographer who provided at least 50 NT measurements. This number limits random error in the estimates and allows for computation of provisional sonographer-specific medians. The first quality measure, percent increase in NT thickness per week, was computed by linear regression using CRL as the independent variable and the logarithm of NT as the dependent variable. We restricted the analysis to observations with CRL measurements between 32 and 84 mm. Similar limits for NT measurements (on a linear scale) were 0.1–4.0 mm. A published equation relating CRL to gestational age¹⁴ was used to convert the slope to percent increase per week. Median MoM, the second quality measure, was directly derived from each sonographer’s series of NT MoM levels reported by the laboratory. The third quality measure, logarithmic standard deviation of the NT MoM levels, was derived using the intervals between the 5th and 95th centiles, 10th and 90th centiles, and the 25th and 75th centiles. The 5th and 95th interval is more reliable, because it is based on a higher proportion of the data. The 25th and 75th interval, however, is more reliable when the tails of the data do not fit a logarithmic Gaussian distribution.

One of the authors (A.D.) is board-certified in obstetrics and gynecology, clinical genetics, and clinical cytogenetics. He is certified to perform NT examinations by the Fetal Medicine Foundation (FMF) and the Society for Maternal Fetal Medicine and performs 30 to 40 each week as part of practice.

RESULTS

Table 1 summarizes the data submitted to each participating laboratory, totaling 23,462 NT measurements from 850 sonographers. Overall, 98% of the measurements were associated with CRL measurements indicating gestational ages of 11, 12, or 13 completed weeks (21%, 54% and 23%, respectively). An additional 2% of samples were assigned gestational ages of 10 or 14 completed weeks. The use of sonographer-specific medians varied between laboratories, as reflected in the use of 10 or more sets of medians by two laboratories, as opposed to only a single fixed set of medians by two others. The remaining two laboratories (1–10 sets of medians) created sonographer- or center-specific medians for high-volume clients. The average number of NT observations per sonographer ranged from a low of 19 (Laboratory C) to a high of 67 (Laboratory B). The last column shows that the proportion of sonographers reporting 50 or more NT measurements to each laboratory in 6 months ranged from 34% to 80%, with an average of 61%. This

Table 1
Nuchal translucency (NT) information provided by six participating laboratories

Screening laboratory	Sets of medians ^a	6 months ending	NT observations	Number of sonographers	Proportion of NT observations from sonographers with ≥50 observations (%)
A	>10	Oct 2004	4,646	148	75
B	1–10	Feb 2007	3,697	55	80
C	1	Dec 2005	7,745	402	49
D	1–10	Dec 2006	2,206	70	58
E	1	Dec 2006	2,925	103	34
F	>10	Dec 2006	2,243	72	74
All			23,462	850	61

^aIndicates the number of separate reference ranges in clinical use (either sonographer-specific or center-specific).

Table 2
Numbers of nuchal translucency (NT) measurements provided by contributing sonographers, stratified by laboratory

Screening laboratory	NT measurements by sonographer per 6 months						
	Fewer than 50			50 or more			
	<10	10–29	30–49	50–99	100–199	200–399	≥400
A	74	31	10	24	7	1	1
B	0	23	8	17	5	1	1
C	222	98	41	31	8	1	1
D	37	11	5	15	2	0	0
E	14	57	21	8	2	1	0
F	35	18	5	4	10	0	0
All	382	238	90	99	34	4	3

indicates that, on average, 39% of all NT measurements received by these laboratories are from sonographers who may not have sufficient numbers to perform quality assessment or to validate or compute sonographer-specific reference data within a 6-month time window.

Table 2 shows a further breakdown of the numbers of observations reported by sonographers and stratified by laboratory. All laboratories reported either <10 or 10–29 NT measurements over 6 months as being the most common category. Overall, 710 sonographers (84%) reported fewer than 50 observations and 140 sonographers (16%) reported 50 or more observations. Fifty observations in 6 months would translate into about eight NT measurements per month or about two per week. Only 3 of the 850 sonographers had an average rate of four NT measurements per working day (400 or more in 6 months).

Figure 1 displays the three quality measures for the 140 sonographers who could be reliably assessed. The left-most group shows that the median increase per week for NT measurements is 22%. The target range was taken to be 15–35%, approximately the 10th and 90th centiles. Based on modeling, values that are more than 3 percentage points higher (or lower) than these limits (indicated by dashed lines) would be considered significantly higher (or lower) even with as few as 50 ob-

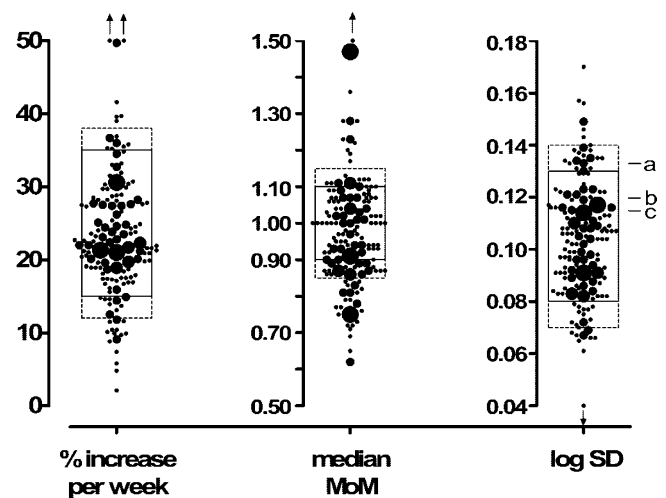


Fig. 1. Summary of quality assessment parameters for sonographers providing 50 or more nuchal translucency (NT) observations in 6 months. The figure shows the percent increase per week (left column), the median NT MoM (center column), and logarithmic standard deviation (right column) for 140 sonographers. The size of the circle relates to the number of NT observations submitted within 6 months (largest circles indicate 400 or more, moderately large ones indicate 200–399, medium ones indicate 100–199, and smallest ones indicate 50–99). The accompanying target ranges (boxes) indicate the approximate 10th to 90th centiles of the observations (percent increase per week and log SD). The target range for the median MoM is based on laboratory experience with serum markers. The dotted extensions indicate the level at which the point estimate for sonographers reporting as few as 50 observations would be statistically “outside” the target range. The small ticks labeled “a,” “b,” and “c” to the right of the log SD plot indicate estimates for this parameter, reported from three large published datasets.

servations. The middle group in Figure 1 shows the median NT MoM for these same sonographers. The target range of 0.90–1.10 is based on extensive knowledge of the impact of inaccuracy on Down syndrome risk estimates using serum markers.¹³ Median MoM levels more than 0.05 higher (or lower) than the respective limits would be significantly outside the range, even with as few as 50 observations. The right-most group in Figure 1 shows the logarithmic standard deviation of the NT MoM. The consensus estimate is 0.105, with a target range of 0.08–0.13, again roughly corresponding to the 10th and 90th centiles. Values deviating by more than 0.10 are significantly outside the target range. The small labels (a, b, and c) to the right of the figure show the logarithmic standard deviation at 12 weeks’

Table 3

Summary of sonographer performance, based on three quality assessment parameters, for sonographers with 50 or more observations in 6 months, stratified by laboratory

Laboratory	Number assessed	Number of sonographers outside the target range for:						
		None	Only slope	Only SD	Slope and SD	Only MoM	Any other	All three
A	33	15	3	4	1	8	2	0
B	24	17	2	1	1	2	1	0
C	41	21	5	1	2	11	1	2
D	17	9	3	2	0	3	0	0
E	11	7	1	0	0	2	0	1
F	14	10	1	0	0	3	0	0
All	140	79	15	8	4	29	4	3

gestation reported from three large studies of NT measurements. These values of 0.133,⁵ 0.120,¹² and 0.115¹⁵ are all based on NT MoM levels using either sonographer-specific medians^{5,12,15} or a single set of medians with strict adherence to a specified methodology.¹⁵ These three estimates are higher than 86%, 76%, and 69% of the computed standard deviations for the 140 assessed sonographers, respectively.

Table 3 summarizes the number of sonographers who are outside of the target range for one or more of the quality assessment parameters, stratified by laboratory. Overall, 79 of the 140 assessed sonographers (56%) were within target ranges for all three parameters. Only three sonographers (2%) were outside of the target range for all three parameters. The most common single quality assessment parameter outside the target was the median NT MoM, found for 29 sonographers/laboratories (21%). As a way to further explore any relationship between these quality parameters, we examined all three pairwise correlation coefficients. Only the correlation between percent increase per week and logarithmic standard deviation was significant ($r = 0.46$, 95% confidence interval 0.32–0.58). The remaining two correlations were nonsignificant (0.19 for logarithmic standard deviation versus median MoM, and 0.04 for percent increase per week versus median MoM). The relationships between the number of NT measurements and the three quality assessment parameters were also evaluated, by dividing the sonographers into two approximately equal groups (69 with between 50 and 75 observations and 71 with more than 75 observations). There was a strong association ($\chi^2 = 4.3$, $P = 0.04$) between numbers of observations and the proportion of percent increases per week within the target range (68% inside the range for smaller numbers and 83% inside the range for larger numbers). There were no associations between numbers of observations and the other two quality assessment parameters ($\chi^2 = 0.2$, $P = 0.7$; $\chi^2 = 0.01$, $P = 0.9$; for numbers of observations, versus median and logarithmic standard deviation, respectively).

Figure 2 shows the sonographer-specific median NT MoM for one laboratory at two intervals over an 18-month time period. The data for the left hand side of the graphic were collected over 6 months in early 2005 (these data are not in-

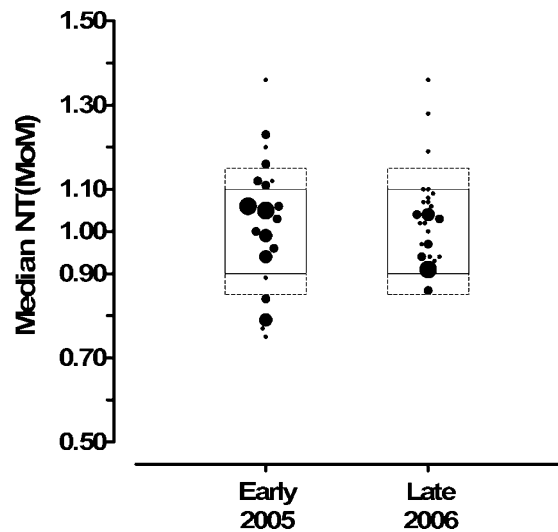


Fig. 2. Documentation of improvement in nuchal translucency (NT) median MoM levels over an 18-month period by sonographers submitting measurements to one screening laboratory. In early 2005, sonographers were required to be trained and credentialed, but there was no feedback concerning their median NT MoM levels and the laboratory used a single set of medians. By late 2006 routine quality assessment was being performed, sonographers with outlying values were being contacted, and sonographic-specific medians were in place.

cluded in any of the other analyses). At that time, the laboratory required that sonographers be certified, but the laboratory was not routinely computing and using sonographer-specific reference data. Overall, 8 of the 20 assessed sonographers (40%) were within the target range. By late 2006 (right hand side) routine quality assessment was being performed, sonographers with outlying values were being contacted, and sonographer-specific medians were in place. At this point, 21 of the 24 assessed sonographers (88%) were within range. This rate is higher than the overall rate of 23% achieved by the other five laboratories (Table 3), indicating that this may represent a real improvement. Active exchanges between laboratory and sonographer and implementation of sonographer-specific medians for those sonographers with sufficient data materially improved the interpretation of NT measurements.

Table 4

Observed and predicted number of NT MoM values at or above 1.5 and 2.0 MoM, by quintile of sonographer-specific NT MoM variance

Quintile	SD ^a (range)	Number of sonographers	Number of NT observations	Percent ≥1.5 MoM		Percent ≥2.0 MoM	
				Observed	Predicted	Observed	Predicted
1	<0.084	29	2,734	2.5	0.9	0.5	<0.1
2	0.084–0.097	28	3,115	3.8	2.6	0.7	<0.1
3	0.098–0.108	29	2,548	5.5	4.6	1.3	0.2
4	0.109–0.122	26	3,442	6.2	6.3	1.1	0.5
5	>0.122	28	2,371	9.4	9.7	2.2	1.3
All	0.105	140	14,210	5.3	4.6	1.1	0.2

^aLogarithmic standard deviation computed using the interval between the 5th and 95th centiles, divided by 3.2.

Table 4 examines the relationship between logarithmic standard deviation and proportion of NT measurements at or above 1.5 and 2.0 MoM. Sonographers were divided into five approximately equal groups, based on the logarithmic standard deviation (second column). Between 2371 and 3442 NT observations are contained in each quintile. Before this analysis was performed, each of the 140 sonographer-specific datasets was adjusted (by dividing by the sonographer’s overall median MoM) so that the resulting median MoM would be exactly 1.00. This ensures that the proportion of NT measurements above a selected MoM level is dependent only on the spread of the data and not on the median value. As expected, the proportions of NT measurements at or above 1.5 MoM are clearly dependent on the quintile of logarithmic standard deviation, ranging from a low of 2.5% to a high of 9.4%. Using the median log SD for each interval (0.075, 0.081, 0.105, 0.116, and 0.135, respectively), the expected percentage at or above 1.5 MoM can also be calculated. Especially at the lower quintiles, the predicted percentages are much lower (e.g., 2.5% observed and 0.9% predicted for the first quintile). The same analysis was then performed using the cutoff of 2.0 MoM. The percentage of NT MoM levels above this cutoff are lower, but still show a strong association with quintile of logarithmic standard deviation. Again, the predicted percentages, based on the fitted logarithmic standard deviations, are lower than the observed percentages.

As a way to further investigate the discrepancy between the observed and expected proportion of NT observations above selected cutoff levels, Figure 3 shows probability plots for two sonographers, one from the fifth quintile (logarithmic standard deviation of 0.133) and one from the first quintile (logarithmic standard deviation of 0.083). The plots show that the data fit well between the 5th and 95th centiles, indicating that the method of computing standard deviations is reasonable. There is significant discrepancy in observations below 0.8 MoM, but with the truncation limit usually set at 0.7 MoM, the likelihood ratios assigned to these two groups will be similar. Above 1.3 MoM (about the 75th centile and associated with a likelihood ratio of about 1.0), the two plots again diverge, with much lower values being recorded at a given centile for the sonographer with the smaller standard deviation. This will

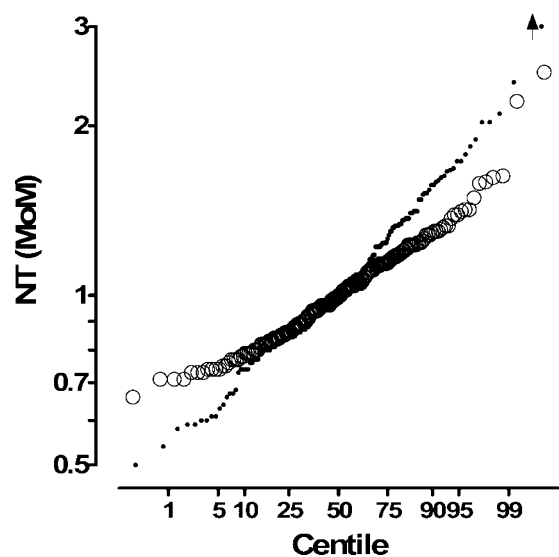


Fig. 3. Probability plots for NT measurements from two sonographers. One sonographer (small filled circles) reported 181 NT measurements, with a logarithmic standard deviation of 0.133. Another sonographer (large open circles) reported 203 measurements with a logarithmic standard deviation of 0.083. The proportions of observations above 2.0 MoM are 2.8% and 1.0%, respectively.

have the effect of increasing the assigned Down syndrome risk by smaller amounts than the increases associated with the measurements obtained by the sonographer having the larger standard deviation. At the 97th or 98th centiles, the observed NT measurements for both sonographers are higher than expected, based on the computed standard deviation (straight line through the majority of data). It is this effect that results in the discrepancy between the observed and expected proportion above 2.0 MoM.

Two laboratories also provided the maternal age of the women being tested. To examine whether sonographers objectively measured NT in light of the woman’s age, sonographer-specific NT MoM levels were stratified by maternal age (<35 years, ≥35 years). A total of 113 sonographers reported 50 or more NT measurements, with at least 10 measurements occurring in each of the two age groupings. If the two median MoM levels were within 10% of each other (e.g., 0.96 MoM in younger women and 1.03 MoM in older women), the results

were considered to be consistent. Thirteen sonographers reported higher levels in older women, whereas eight sonographers reported lower levels. This difference is not statistically significant (binomial test of the 21 sonographers whose differences fell outside of $\pm 10\%$, $P = 0.2$).

One laboratory provided both a sonographer code and a code for each sonographer's center. To explore within-center differences, we computed a new set of medians for one large center composed of 24 sonographers, providing results for 764 women. Thus, this group will, by definition, have an overall median NT MoM of 1.00. The sonographer-specific median NT MoM levels were then compared for the eight who reported NT results for 50 or more women. Based on an analysis of variance ($F = 5.32$, $P < 0.001$), one had a significantly low median (0.88 MoM), one had a significantly high median (1.19 MoM), and the remaining six were consistent (0.96–1.05 MoM, $F = 1.4$, $P = 0.2$). The median MoM for the 105 women receiving NT measurements from the 16 additional sonographers who reported fewer than 50 observations was 0.90, but this difference was not significantly lower ($F = 0.8$, $P = 0.4$).

DISCUSSION

The three main findings from this study are as follows: (1) many sonographers submit NT measurements for only a few women over a 6-month time period, and, therefore, their performance cannot be reliably assessed by the laboratory; (2) among those sonographers whose performance can be assessed, there are important differences in each of the three quality assessment parameters; and (3) laboratories need to improve their interpretation of NT measurements by implementing sonographer-specific medians, when possible. Each of these is discussed in more detail, below.

The current cross-sectional study used observations from six experienced prenatal screening laboratories in North America that currently receive NT measurements for clinical interpretation of Down syndrome risk. This setting differs from most other reports on this topic, in that the NT measurements are being performed as part of routine practice and not as part of a research protocol.^{5,6,16,17} Given the widely varying characteristics of our participating laboratories, the findings are likely to be representative of current practice in North America. Overall, 39% of the screened women had NT measured by a sonographer who submitted too few samples to that laboratory for adequate and timely quality assurance monitoring. It is possible that these sonographers provided testing to more women, but sent the results to another laboratory for reasons such as health insurance requirements. A program such as that proposed by Nuchal Translucency Quality Review (NTQR) program could help resolve this situation,^{8,18} but only if all laboratories and sonographers involved participated. It is also possible that the sonographer would practice at several centers, and it has not yet been shown that these data can be combined. One possibility is that there is a concerted effort to credential as many sonographers as possible at some centers offering NT measurement, rather than concentrate such test-

ing among a few highly experienced sonographers. At one participating laboratory for example, the number of sonographers submitting samples more than doubled over two years while the proportion of sonographers with sufficient observations for monitoring (i.e., at least 50 every 6 months) remained below 40% (A Summers, Unpublished data). For now, laboratories can monitor the proportion of NT observations from sonographers who can be assessed (Table 1) and, through education, attempt to increase this proportion over time. Our data do not provide clear guidance on whether the performance of sonographers reporting relatively few NT measurements is similar to, worse, or better, than for sonographers reporting at least 50 observations in 6 months. To deal with this problem, laboratories might consider alternative ways to collect at least 50 observations, such as asking the sonographer (or NTQR) for additional measurements submitted elsewhere, or performing the evaluation over a longer time period (e.g., 1 year). It might also be possible to compare the NT measurements from these sonographers with other more experienced sonographers at the same center, by using appropriate center-specific medians. As a last resort, laboratories could decide to not accept samples for interpretation. A future study covering a longer time period may be able to determine whether sonographers performing relatively few NT measurements have better, worse, or equivalent performance to sonographers who can be routinely monitored.

Epidemiological monitoring of NT measurement shows that there are important differences in each of the three quality measures chosen (Fig. 1). The percent increase per week in NT measurements has been reported to be about 20%. We found the 10th to 90th centile for this parameter to be from about 15% to 35%, with several sonographers having very small (<10%) or very large (>40%) increases in NT measurements with increasing gestational age. There is no obvious explanation for this variability. However, larger increases may be due to sonographers inadvertently adjusting their NT measurement in light of their knowledge of the CRL measurement. Smaller-than-expected increases may be due to low magnification (or low resolution equipment), where little variability in NT measurements can be detected. This might also explain why there is an important positive correlation between percent increase per week and logarithmic standard deviation ($r = 0.46$). Variability in the NT MoM is easier to explain. There are many sources of variability in measuring the NT (in millimeters), including technique, magnification, machine quality, and measurement protocol (e.g., largest or mean measurement). Several studies have shown important differences between sonographers, even after training and feedback.^{5–7,16–20} At least two studies have shown that the use of sonographer-specific medians improves screening performance.^{5,12} Most of the laboratories involved in the current study do not routinely use and update sonographer-specific medians. These two factors likely explain the high variability seen for the median NT MoM. We were surprised at the wide range of logarithmic standard deviation of the NT MoM. Three major studies reported this quality measure (at 12 completed weeks' gestation)

to be between 0.115 and 0.133. Our study found that between 69% (using a cutoff of 0.115) and 85% (using a cutoff of 0.133) of the sonographers evaluated had logarithmic standard deviations below these reported values. Such a “tightening” of the distribution cannot be accomplished by more careful measuring. Although its cause is not known, it may be due to the “routine” nature of NT measurements in the general population. Sonographers are best at identifying qualitative rather than quantitative changes. Thus, a sonographer who assigns values that vary little (resulting in a reduced variance) might still recognize a large NT as abnormal and assign that NT a high value (Fig. 3).

Laboratories are experienced in adjusting reference data as serum assays change over time, or when a change in manufacturer occurs. The two main differences between this historical experience and implementing sonographer-specific medians are that there are far more sonographers (hundreds) than serum assays (four or five), and the amount of data available for analysis is much smaller for sonographers (16% of sonographers with 50 or more NT measurements reported in 6 months) than for serum assays (often 50 or more samples per day). Implementing and monitoring sonographer-specific medians will be challenging, but probably necessary. Preliminary evidence shows that it can be done as part of routine practice (Fig. 2). Two large studies involving routine measurement of NT converted those results to MoM using a single set of medians, center-specific medians, and sonographer-specific medians. In both studies, the matched analyses indicated that sonographer-specific medians resulted in the best screening performance (highest detection for a given false-positive rate).^{5,6} Given the variability of NT measurements documented in the current dataset, it is reasonable to conclude that implementation of sonographer-specific medians would improve overall performance in practice, as well. Our data also suggest that center-specific medians may have some advantage but can hide important differences among sonographers, if individual monitoring is not also performed. The value of using center-specific medians for sonographers who submit relatively few observations has not yet been established, but their use seems a reasonable alternative.

Laboratories need to be aware of the inherent variability of NT measurements, as well as reasons why NT measurements might vary from sonographer to sonographer, not because the laboratory is a source of training, credentialing or image review, but as a way to interact knowledgably with the ultrasound community. The within-sonographer short-term standard deviation around a repeated measurement is between 0.3 and 0.5 mm,^{21–24} and seems to be essentially constant across the range of NT measurements. This results in widely varying coefficients of variation, ranging from up to 50% (a 0.5 mm standard deviation at 1.0 mm NT measurement) to as low as 10% (a 0.3 mm standard deviation at 3.0 mm NT measurement). Sonographers will usually obtain multiple technically acceptable measurements of the NT during a single examination. The FMF recommends using the largest NT measurement,^{9–11} whereas SURUSS and the FASTER trial^{5,6} used a central mea-

sure (mean). SURUSS found that multiple measurements provided better performance than a single measurement, and that using the mean value was slightly better than using the largest value.⁵ However, the median levels would be expected to be lower for sonographers reporting the central measure, compared with those reporting the largest value.

The magnification, or image size, is known to impact NT measurements.^{25,26} The original description of NT measurements made no mention of magnification, but pictures show the fetus occupying 40–60% of the image.²² Later publications suggest that the fetus occupy 75% of the image. The latest recommendations suggest that the fetus occupy the entire image (100%). Some sonographers use even higher magnifications, so that the image includes only the head and neck (200%). Studies indicate that increasing magnification (from 100% to 200%) will reduce the NT measurement by 0.1–0.2 mm.^{25,26} Although based on small numbers, one study found the screen-positive rate (for the combination of NT and maternal age) to decrease from 5.1% to 2.3% ($P = 0.048$) when magnification was increased from 100% to 200%.²⁶ The ultrasound machine itself may influence not only the NT measurement, but how often the NT measurement is obtained.⁵ Placement of the calipers is critical to a reliable NT measurement, and specific guidelines have been promulgated. However, these are not easy guidelines to either implement or assess. Two studies found poor intraobserver reliability (κ values <0.4 , for scoring systems ranging from 1 to 4 and 0 to 2) when experts were asked to grade caliper placement on still images.^{27,28} Another found that, after training, 25% of experienced sonographers still placed the calipers incorrectly.²⁸

Image scoring methods for quality assessment have been suggested.^{27–29} The most common criteria include the view, caliper placement, head position, visualization of the amnion, and magnification. One difficulty with scoring systems is the weighting factors. For example, should the quality score be high if all criteria but caliper placement were to be met, even though the result might be poor?⁷ These systems have been found too cumbersome to be used for routine quality assessment. They might be of more value during the initial training period, or when quantitative assessment indicates the need for further scrutiny.

Although the scope of this study is to document the current state of NT measurements and to suggest methods for monitoring, an important ancillary consideration is the impact of between-sonographer differences on Down syndrome screening performance. A recent article³⁰ modeled the effect of a 0.5-mm absolute difference between a sonographer’s NT measurements and expected reference values (at a given CRL measurement). Using only maternal age and NT as the screening test, this difference would result in detection being reduced from 82% to 72% with a corresponding reduction in the false-positive rate from 6.9% to 2.5%. This finding underscores the need for consistency in NT interpretation to have reliable and effective Down syndrome risk estimates.

NT measurements have been shown to be a useful marker for Down syndrome in the late first trimester in both high-risk

and low-risk populations, but only when accompanied by targeted training and ongoing quality assessment. Quality assurance was accomplished within research trials by the threat of removing sonographers, if performance was not acceptable. Outside of trials, quality assurance has been implemented in several ways. The FMF uses certification and has the option of deactivating the associated software, if long-term performance is not adequate. This would impact only those who rely on that software to compute risks. NTQR provides credentialing services and offers retraining for sonographers who perform outside of the expected norm. Laboratories have the additional option of not accepting samples from a specific sonographer, if his/her quality measures are consistently outside of a reasonable target range and retraining is either not accepted or not successful. It must also be appreciated that monitoring laboratory tests is not the same as monitoring individual sonographers. Changes in individual performance can occur because of implementation of new measurement guidelines at their center, or by changing technique after attending a conference or lecture. Ongoing quality assessment by the laboratory should be seen as only one of several ongoing efforts by the screening community in general to help assure that quality Down syndrome screening services are widely available.

ACKNOWLEDGMENTS

Partial funding for this project was provided by the American College of Medical Genetics Foundation. We thank Devereux N. Saller, MD, University of Virginia, for his comments regarding the use of NT measurements in clinical practice. Dr. Saller is board certified in obstetrics and gynecology, maternal-fetal medicine and clinical genetics and credentialed by the Fetal Medicine Foundation and the Society for Maternal Fetal Medicine to perform NT measurements.

References

- ACOG Practice Bulletin No. 77: screening for fetal chromosomal abnormalities. *Obstet Gynecol* 2007;109:217–227.
- Nicolaides KH, Azar G, Byrne D, Mansur C, et al. Fetal nuchal translucency: ultrasound screening for chromosomal defects in first trimester of pregnancy. *BMJ* 1992;304:867–869.
- Nicolaides KH, Brizot ML, Snijders RJ. Fetal nuchal translucency: ultrasound screening for fetal trisomy in the first trimester of pregnancy. *Br J Obstet Gynaecol* 1994;101:782–786.
- Palomaki GE, Lambert-Messerlian GM, Canick JA. A summary analysis of Down syndrome markers in the late first trimester. *Adv Clin Chem* 2007;43:177–210.
- Wald NJ, Rodeck C, Hackshaw AK, Walters J, et al. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *J Med Screen* 2003;10:56–104.
- Malone FD, Canick JA, Ball RH, Nyberg DA, et al. First-trimester or second-trimester screening, or both, for Down's syndrome. *N Engl J Med* 2005;353:2001–2011.
- Snijders RJ, Thom EA, Zachary JM, Platt LD, 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.
- Nuchal Translucency Quality Review Program. Available at: <https://www.ntqr.org/SM/default.aspx> Accessed May 30, 2007.
- Fetal Medicine Foundation United Kingdom. Available at: <http://fetalmedicine.com/fmf.htm> Accessed May 30, 2007.
- Fetal Medicine Foundation United States. Available at: <http://fetalmedicine.com/usa/> Accessed May 30, 2007.
- Fetal Medicine International. Available at: <http://mfmedicine.com/> Accessed May 30, 2007.
- Malone FD, Wald NJ, Canick JA, Ball RH, et al. Use of overall population, center-specific, and sonographer-specific nuchal translucency medians in Down syndrome screening: which is best? *Am J Obstet Gynecol* 2003;189:S232.
- Knight GJ, Palomaki GE. Epidemiologic monitoring of prenatal screening for neural tube defects and Down syndrome. *Clin Lab Med* 2003;23:531–551, xi.
- Hadlock FP, Shah YP, Kanon DJ, Lindsey JV. Fetal crown-rump length: reevaluation of relation to menstrual age (5–18 weeks) with high-resolution real-time US. *Radiology* 1992;182:501–505.
- Spencer K, Bindra R, Nix AB, Heath V, 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;22:142–148.
- Crossley JA, Aitken DA, Cameron AD, McBride E, et al. Combined ultrasound and biochemical screening for Down's syndrome in the first trimester: a Scottish multi-centre study. *BJOG* 2002;109:667–676.
- Schielen PC, van Leeuwen-Spruijt M, Belmouden I, Elvers LH, et al. Multi-centre first-trimester screening for Down syndrome in the Netherlands in routine clinical practice. *Prenat Diagn* 2006;26:711–718.
- Gyselaers WJ, Vereecken AJ, Van Herck EJ, Straetmans DP, et al. Audit on nuchal translucency thickness measurements in Flanders, Belgium: a plea for methodological standardization. *Ultrasound Obstet Gynecol* 2004;24:511–515.
- Malone FD. Nuchal translucency-based Down syndrome screening: barriers to implementation. *Semin Perinatol* 2005;29:272–276.
- Wapner R, Thom E, Simpson JL, Pergament E, et al. First-trimester screening for trisomies 21 and 18. *N Engl J Med* 2003;349:1405–1413.
- Pajkrt E, Mol BW, Boer K, Drogtop AP, et al. Intra- and interoperator repeatability of the nuchal translucency measurement. *Ultrasound Obstet Gynecol* 2000;15:297–301.
- Pandya PP, Altman DG, Brizot ML, Pettersen H, et al. Repeatability of measurement of fetal nuchal translucency thickness. *Ultrasound Obstet Gynecol* 1995;5:334–337.
- Braithwaite JM, Kadir RA, Pepera TA, Morris RW, et al. Nuchal translucency measurement: training of potential examiners. *Ultrasound Obstet Gynecol* 1996;8:192–195.
- Newey VR, Nassiri DK, Bhide A, Thilaganathan B. Nuchal translucency thickness measurement: repeatability using a virtual ultrasound scanner. *Ultrasound Obstet Gynecol* 2003;21:596–601.
- Edwards A, Mulvey S, Wallace EM. The effect of image size on nuchal translucency measurement. *Prenat Diagn* 2003;23:284–286.
- Teoh M, Meagher SE, Choong S, Shekleton P, et al. The effect of image size on screen-positive rates for nuchal translucency screening. *BJOG* 2006;113:479–481.
- Fries N, Althuser M, Fontanges M, Talmant C, et al. Quality control of an image-scoring method for nuchal translucency ultrasonography. *Am J Obstet Gynecol* 2007;196:272, e271–e275.
- Herman A, Dreazen E, Maymon R, Tovbin Y, et al. Implementation of nuchal translucency image-scoring method during ongoing audit. *Ultrasound Obstet Gynecol* 1999;14:388–392.
- Wojdemann KR, Christiansen M, Sundberg K, Larsen SO, et al. Quality assessment in prospective nuchal translucency screening for Down syndrome. *Ultrasound Obstet Gynecol* 2001;18:641–644.
- Evans MI, Van Decruyels H, Nicolaides KH. Nuchal translucency measurements for first-trimester screening: the 'price' of inaccuracy. *Fetal Diagn Ther* 2007;22:401–404.