Fetal Diagnosis and Therapy

Fetal Diagn Ther 2007;22:401–404
DOI: 10.1159/000106342

Received: April 24, 2006
Accepted after revision: August 8, 2006
Published online: July 24, 2007

Nuchal Translucency Measurements for First-Trimester Screening: The 'Price' of Inaccuracy

Mark I. Evans[a] Hilde Van Decruyes[b] Kypros H. Nicolaides[b]

[a]Fetal Medicine Foundation of America, Comprehensive Genetics, Mt. Sinai School of Medicine, New York, N.Y., USA;
[b]Fetal Medicine Foundation, London, United Kingdom

Key Words
NT screening · Ultrasound · Quality assurance

Abstract
Objective: First-trimester aneuploidy screening is in transition from the phase of 'development' to that of 'diffusion.' As with all other technologies, there is risk of diminished performance until newer centers are well experienced. Typically, inexperienced sonographers under-measure nuchal translucency (NT), and experience, training and continual monitoring of data are needed to ensure that measurements do not deviate. Here we assess the impact of systematic under-measurement on abnormality detection. ***Study Design:*** Actual NT measurements from 13,887 normal fetuses, 82 with trisomy 21 (T21) and 61 with other abnormalities (OA) with birth outcome data were mathematically modified to show either a 25% or 0.5-mm decrease in measurement. The impact upon sensitivity and screen-positive rates were assessed. ***Results:*** Reducing the NT measurements of T21 and OA cases lowers the sensitivity from 81.7 and 70.5%, respectively, to 67.1 and 62.3% ($p < 0.01$). If normals are correspondingly lowered, then the screen-positive rates are reduced from 6.9 to 2.0 and 2.5%. To maintain the same screen-positive rates and sensitivity, the risk threshold would have to be increased from 1/300 to 1/556. ***Conclusion:*** Minor inaccuracies in NT measurements as small as 25% or 0.5 mm will have very significant negative impacts upon abnormality detection, reducing detection rates by 18% (81.7 to 67.1%). Just as it is completely accepted that laboratory measurements require standardization and quality assurance, NT measurements, because they are used in an algorithm, need to be treated with the same rigor. That way the published data from centers that have developed such screening can be applied by other operators at other sites when counseling their patients.
Copyright © 2007 S. Karger AG, Basel

Introduction

First-trimester aneuploidy screening using nuchal translucency (NT) measurements, then later in conjunction with maternal blood samples for PAPP-A and free β-hCG, has been developed over the past decade. Its efficacy has been demonstrated on several continents in over 1,000,000 patients [1–8]. Numerous studies, principally coordinated by the Fetal Medicine Foundation (FMF), based in London, have shown sensitivities for Down syndrome detection approaching 90% for a 5% false-positive rate. Other publications have demonstrated less success – in some cases dramatically so [9–11]. Further analysis of the unsupportive data has shown that in many cases, measurements were either not made properly, attempted at gestational ages when NT measurements were not technically achievable, or were only con-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2007 S. Karger AG, Basel
1015–3837/07/0226–0401$23.50/0

Accessible online at:
www.karger.com/fdt

Mark I. Evans, MD
Fetal Medicine Foundation of America, Comprehensive Genetics
Mt. Sinai School of Medicine, 131 East 65th Street
New York, NY 10021 (USA)
Tel. +1 212 744 2590, Fax +1 201 585 0685, E-Mail evans@compregen.com

Table 1. NT modifications: impact upon sensitivity

	Risk cut-off	Nl (screen positive)	T21 (sensitivity)	OA (sensitivity)
Total numbers		13,887	82	61
NT (actual)	1 in 300	6.9%	81.7%	70.5%
ABN <25%*	1 in 300	2.0%	67.1%	62.3%
ABN <0.5 mm	1 in 300	2.5%	71.9%	65.6%
All <25%	1 in 556	6.9%	81.7%	70.5%
All <0.5 mm	1 in 538	6.9%	81.7%	70.5%

* $p < 0.01$.

sidered as an afterthought to getting first-trimester scans completed as fast as possible [9–11].

As a result, the FMF established a training and certification process to ensure appropriate quality control. Sister organizations such as the Fetal Medicine Foundation of America were established to aid the FMF in training across the globe. Other organizations such as the Society for Maternal Fetal Medicine have also made it their policy to endorse the concepts of quality review of such measurements.

The concept of considering NT measurements to be merely another lab analyte value when used in an algorithm has required a major culture adjustment to those who considered ultrasound to be an 'art' rather than a lab value [12]. As United States National Institute for Child Health and Human Development funded studies such as BUN and FASTER trials have confirmed the efficacy of first-trimester screening, there has been a rapid expansion of utilization of screening in the United States [13, 14].

The incorporation of new technologies follows a pattern of development (in which a small number of centers develop and perfect a new technique) that then moves to a phase of diffusion, when more and more centers begin to employ the new methods [15]. NT and first-trimester screening is in transition from the phase of 'development' to that of 'diffusion'. As seen with other technologies, there is great risk of diminished performance until newer centers are well experienced. Training and continual monitoring are needed to ensure accurate measurements. Typically, inexperienced sonographers under-measure NT. Some have suggested that the way to counteract such biases is to use center-specific and sonographer-specific medians rather than national and international standards. Such underestimation was shown in the BUN study to gradually disappear with more experience over the three years of that study [13]. Here we assess the impact of systematic under-measurement on abnormality detection, i.e. garbage in/garbage out.

Methods

Actual NT measurements by certified sonographers from 13,887 clinically documented normals, 82 fetuses with trisomy 21 (T21) and 61 with other abnormalities (OA) with documented birth outcome data from the FMF in London were used for this study. The patients were a random subset taken from the FMF database of over 200,000 cases, and the data follow the distribution of cases seen in the overall FMF database. The data were then mathematically modified to show either a 25% or 0.5-mm decrease in measurement. These alterations were chosen as they represent the types of variations commonly seen from physicians and sonographers in the training process prior to their achieving certification from the FMF. The resultant impacts upon sensitivity and screen-positive rates were then assessed using age and NT alone, i.e. without biochemistry in order to not compound the issues. Methods and protocols for NT measurements, incorporation into an algorithm for risk computation and sensitivities and specificities for the detection of aneuploidy and other genetic abnormalities have been previously published and will not be repeated here [1–8, 13, 14]. χ^2 analyses were performed as appropriate.

Results

Reducing the NT measurements of T21 and OA cases, as would be seen if inaccurate measurements of the abnormals were placed against the world-wide standardized database, lowers the sensitivity of finding abnormals from 81.7 and 70.5%, respectively, to 67.1 and 62.3% ($\chi^2 = 78.3$, $p < 0.01$) (table 1). Under these circumstances, the screen-positive rate is correspondingly reduced from 6.9 to 2.0% and 2.5%. If one uses a center-specific database that systematically lowers all the measurements, then the 'normals' are correspondingly also lowered. However, to maintain the same screen-positive rates and therefore sensitivity, the risk threshold cutoff would then have to be increased from 1/300 to 1/556, which would render the risk rates quoted per patient to be significantly inaccurate.

Discussion

Records from the FMF and others using certified sonographers' NT measurements and data have shown a very high efficacy of first-trimester screening. In combination with first-trimester blood specimens for free

β-hCG and PAPP-A, detection statistics for Down syndrome, other aneuploidies and multiple genetic syndromes have consistently approached 80–90% when performed in the rigorous, quality-controlled methods pioneered by the FMF [7]. The training process pioneered by the FMF has certified thousands of physicians and sonographers whose data are monitored by the FMF to ensure continual quality control. In the United States, the Society for Maternal Fetal Medicine decided to create its own process to train and quality review American physicians and sonographers.

Acceptance of the need for quality review has been high, but not uniform. One center, which refused to participate saying that they were trained maternal fetal medicine specialists and did not need any 'interference', showed data with 82% of their cases measured below the median and 40% below the 5th percentile ($p < 0.001$; unpubl. data). Such results applied to patient care produce falsely reassuring screening results with too few cases called 'at risk'. Such data would produce individual cases being quoted inaccurately low risks, with a number of patients who should have been considered 'at risk' falsely being called 'normal'. The net effect is an increased risk of missing affected cases (false negatives). As modeled in this study, even minor inaccuracies in NT measurements as small as 25% or 0.5 mm can have a very significant negative impact upon abnormality detection. In this mathematical experiment, the under-measurements of NTs reduced detection rates by 18% from 81.7 to 67.1%. Gyselaers et al. [19] investigated the accuracy of new sonographers in Belgium and concluded, as did the BUN study, that new sonographers routinely underestimate NTs, particularly at the low end of the spectrum. They then modeled a 0.1-mm change in NT and concluded that even that would have a significant impact upon the percentage of patients called 'at risk' by the FMF algorithm.

Just as it is completely accepted that laboratory measurements require standardization and quality assurance, NT measurements, because they are used in an algorithm, need to be treated with the same rigor. That way the published data from centers that have developed such screening can be applied by other operators at other sites when counseling their patients. Data from the BUN study showed systematic under-reporting of data in the first year of the study; over the next two years the data approached that of the larger database of the FMF [13]. Trying to add 'crutches' to mathematically combat systematic bias starts a cascade of events that either require alterations of risk-based cutoffs to maintain the screen-positive rate or significant changes in the screen-positive rate to keep the accuracy of the risk quotations accurate.

The standardization and certification concepts have met with varying degrees of resistance in the fetal medicine community. Much of the efforts of the FMF and the Society for Maternal Fetal Medicine and its Maternal Fetal Medicine Foundation and Nuchal Translucency Quality Review program are likely to be centered on the need to change the culture of American clinicians to accept the fact that further standardization of clinical functions is necessary to achieve optimal results beyond that of routine board certification.

Some have suggested that instead of using 'national' or world-wide standards of measurement, center-specific medians could counteract the effects of under- or-over reporting of data [14]. Inaccuracies would phase shift the local curve and produce either a disproportionate amount of abnormals in one direction and under-counting of abnormals in the other. Most publications have data showing that there are not significant differences in curves among centers around the world either by location, race, or ethnic group. As such when comparing a group's systematic under- or over-reporting of NT measurements, the bias has to have deleterious implications for screening efficiency which can play out as lowered sensitivity or specificity, or at the very least having to ignore the published correlations of screening test predictions and actual rates of abnormalities that have been seen which have attested to the accuracy of those predictions [1].

Additional US markers including appreciation of the nasal bone and tricuspid regurgitation have been recently proposed [16–18]. These markers have high sensitivity and specificity but require much more expertise than NT. It is likely that these will be adjunctive to initial screening with NT, free β-hCG and PAPP-A and used for those cases with risks intermediate between high enough to warrant immediate diagnostic testing with CVS and low enough to not need any further evaluation. Nasal bone and tricuspid regurgitation will likely be used principally by centers of excellence. The same concerns for quality control and reliability will apply to those markers as to NT.

In sum, the data in this mathematical modeling experiment show that truncating data by mathematically normalizing wide individual variation in measurements can only decrease the efficacy of screening. All measurements have minor coefficients of variation among operators. Normalizing the ranges of systematic errors, however, can only reduce the quality of care provided.

References

1 Hyett J, Nicolaides KH: First trimester ultrasound screening with nuchal translucency; in Evans MI, Johnson MP, Yaron Y, Drugan A (eds): Prenatal Diagnosis. New York, McGraw-Hill, 2006, pp 289–308.

2 Snijders RJ, Thom EA, Zachary JM, Platt LD, Greene N, Jackson LG, 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.

3 Snijders RJM, Noble P, Sebire N, Souka A, Nicolaides KH: UK multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal translucency thickness at 10–14 weeks of gestation. Lancet 1998;351: 343–346.

4 Monni G, Zoppi MA, Ibba RM, Floris M: Results of measurement of nuchal translucency before and after training. Lancet 1997;350: 1631.

5 Cicero S, Bindra R, Rembouskos G, Spencer K, Nicolaides KH: Integrated ultrasound and biochemical screening for trisomy 21 at 11 to 14 weeks. Prenat Diagn 2003;23:306–310.

6 Avgidou K, Papageorghiou A, Bindra R, Spencer K, Nicolaides KH: Prospective first trimester screening for trisomy 21 in 30,564 pregnancies. Am J Obstet Gynecol 2005;192: 1761–1767.

7 Nicolaides KH, Spencer K, Avgidou K, Faiola S, Falcon O: Multicenter study of first-trimester screening for trisomy 21 in 75,821 pregnancies: results and estimation of the potential impact of individual risk-orientated two-stage first-trimester screening. Ultrasound Obstet Gynecol 2005;25:221–226.

8 Evans MI, Hallahan TW, Krantz D, Galen RS: Meta-analysis of first trimester Down Syndrome screening studies: free beta hCG significantly outperforms intact hCG in a multi-marker protocol. Am J Obstet Gynecol (in press).

9 Bewley S, Roberts LJ, Mackinson AM, Rodeck C: First trimester fetal nuchal translucency: problems with screening the general population. 2. Br J Obstet Gynecol 1995;102: 386–388.

10 Kornman LH, Morssink LP, Beekhuis JR, DeWolf BTHM, Heringa MP, Mantingh A: Nuchal translucency cannot be used as a screening test for chromosomal abnormalities in the first trimester of pregnancy in a routine ultrasound practice. Prenat Diagn 1996;16:797–805.

11 Haddow JE, Palomaki GE, Knight GJ, Williams J, Miller WA, Johnson A: Screening of maternal serum for fetal Down's syndrome in the first trimester. N Engl J Med 1998;338: 955–961.

12 Evans MI, Galen RS, Drugan A: Biochemical screening; in Evans MI, Johnson MP, Yaron Y, Drugan A (eds): Prenatal Diagnosis. New York, McGraw-Hill, 2006, pp 277–288.

13 Wapner R, Thom E, Simpson JL, Pergament E, Silver R, Filkins K, et al: First trimester maternal serum biochemistry and fetal nuchal translucency screening (BUN) study group. First trimester screening for trisomies 21 and 18. N Engl J Med 2003;349:1405–1413.

14 Malone FD, Canick JA, Ball RH, Nyberg DA, Comstock CH, Bukowski R, Berkowitz RL, Gross SJ, Dugoff L, Craigo SD, Timor-Tritsch IE, Carr SR, Wolfe HM, Dukes K, Bianchi DW, Rudnicka AR, Hackshaw AK, Lambert-Messerlian G, Wald NJ, D'Alton ME: First-trimester or second-trimester screening, or both, for Down's syndrome. N Engl J Med 2005;353:2001–2011.

15 Cohen AB, Hanft RS: Technology in American Health Care. Ann Arbor, University of Michigan Press, 2004.

16 Cicero S, Curcio P, Papageorghious A, Sonek J, Nicolaides KH: Absence of nasal bone in fetuses with Trisomy 21 at 11–14 weeks of gestation: an observational study. Lancet 2001;358:1665–1667.

17 Cicero S, Spencer K, Avgidou K, Faiola S, Nicolaides KH: Maternal serum biochemistry at 11–13 + 6 weeks in relation to the presence or absence of the fetal nasal bone on ultrasonography in chromosomally abnormal fetuses: an updated analysis of integrated ultrasound and biochemical screening. Prenat Diagn 2005;25:977–983.

18 Falcon O, Auer M, Gerovassili A, Spencer K, Nicolaides KH: Screening for trisomy 21 tricuspid regurgitation, nuchal translucency and maternal serum free β hCG and PAPP-A at 11 + 0 to 13 + 6 weeks. Ultrasound Obstet Gynecol 2006;27:151–155.

19 Gyselaers WJ, Vereecken AJ, Van Herck EJ, Straetmans DP, de Jonge ET, Ombelet WU, Nijhuis JG: Audit on nuchal translucency thickness measurements in Flanders, Belgium: a plea for methodological standardization. Ultrasound Obstet Gynecol 2004; 24:511–515.