Impact of Quality of Nuchal Translucency Measurements on Detection Rates of Trisomies 13 and 18

Mark I. Evans\textsuperscript{a} Eugene Pergament\textsuperscript{b}

\textsuperscript{a}Comprehensive Genetics, Fetal Medicine Foundation of America, and Department of Obstetrics and Gynecology Mt. Sinai School of Medicine, New York, N.Y., and \textsuperscript{b}Northwestern Reproductive Genetics, Chicago, Ill., USA

\textbf{Introduction}

Over the past 40 years, prenatal screening for aneuploidy has evolved from using maternal age alone to mid-trimester single and then multiple maternal serum markers to first-trimester biochemical markers and ultrasound measurements \cite{1}. Each step has increased the sensitivity and specificity of aneuploidy screening. As a direct consequence, detection of aneuploid pregnancies markedly increased, while at the same time the number of pregnancies exposed to the obstetrical risk of expensive and potentially complicated procedures has been reduced \cite{2}.

In the laboratory all tests are regularly subjected to quality control to determine their reliability. It has been well established that standardization has led to markedly improved test performance \cite{3}. For example, during the 1970s and 1980s, it was widely held that variability in maternal serum \textit{H}_{9251}-fetoprotein (AFP) in different patient populations required that each laboratory establish their own medians and distributions. In fact, it was the considerable differences in analytical methods, rather than patient variability, that accounted for the disparity in results between laboratories \cite{1, 2}. Once standardized, intra-laboratory variation in the case of maternal serum AFP was actually greater than the variation between patient populations \cite{4}. At the same time, adjustments for certain in-
dependent variables were also required in screening for aneuploidy using AFP and subsequently multiple maternal serum markers. These included gestational age, maternal weight, race, ethnic group, twins, and diabetes [5]. This process was essentially replicated when first-trimester serum markers were introduced into clinical practice.

The development of ultrasound markers identifying pregnancies at risk of aneuploidy presented a different problem. From a historical perspective, ultrasound evaluations initially were based on individual experiences, and many ultrasonographers, accomplished and otherwise, were resistant to the concept that they had to be subjected to objective external evaluation of their ultrasound measurements for purposes of quality assurance [6]. However, repeated experience has shown that if ultrasound measurements were to be used in a laboratory algorithm, quality control was essential. The net effect of failing to obtain standardization in ultrasound measurements was to lower the detection rate or unnecessarily alarm prospective parents by assessing their pregnancy as ‘at increased risk’ for aneuploidy [1]. Thus, the accreditation process of the Fetal Medicine Foundation (FMF) in London was created a decade ago. In one United States program that refused to participate in the accreditation process, 82% of cases were below the national median for nuchal translucency (NT) measurements, and 40% were below the 5th percentile. In response, the skewed distribution was interpreted as reflecting ‘patient population’ differences rather than acknowledging that technical competence was at issue (unpublished).

More than 50,000 physicians and sonographers worldwide have now been trained by the FMF [7]. The process includes didactics, examinations, submission of films, certification, and periodic recertification [6, 7]. In the United States, the Society for Maternal Fetal Medicine created the Nuchal Translucency Quality Review (NTQR) Program to provide education and quality review. These two programs differ in their approach to quality control, although both are based upon underlying didactic training and documentation of proper techniques and reliability of measurements. The FMF, in part by virtue of a longer history, has had a continuing reassessment and recertification process in place for nearly a decade. The NTQR is still developing its program for continual monitoring and assessment. While most of the focus of first-trimester screening has been targeted at trisomy 21, it has long been recognized that trisomies 13 and 18 can be detected by measurement of NT and their pattern of markedly lower free β-human chorionic gonadotropin (free β-hCG) and pregnancy-associated plasma protein (PAPP-A) [1, 2, 6, 7]. Here, we seek to assess any epidemiologic differences in outcomes of biochemical and NT performance between the United Kingdom and United States for the detection of trisomies 13 and 18.

**Methods**

De-identified data from the FMF in the United Kingdom and from the NTD (New York) laboratory databases in the United States were assessed for the performance of NT and biochemical screening in cases known to have resulted in either trisomies 13 or 18. The study periods for both groups began in 1998 and continued to 2002 for the UK and 2007 for the US as NT screening grew much faster in the UK than the US. UK data have been published, and US data were provided to the investigators by the NTD laboratories [8]. The United Kingdom data included 124 cases and the United States 66 biochemical cases, of which 42 also had NT measurements [8]. Performance of the UK and US screening programs was then compared by laboratory analysis of free β-hCG and PAPP-A with and without NT measurements. While all the UK data are from FMF accredited providers, we do not have the break down of US providers who were FMF America certified or NTQR. Data were analyzed by χ² analysis and t tests as appropriate with calculation of detection rates and false-positive rates (FPRs).

**Results**

For both the United Kingdom and United States data, biochemical detection of trisomies 13 and 18 was virtually identical (table 1). In the United States, the sensitivity was 82.3% for a 1.7% FPR, and in the United Kingdom, it was 84.5% for a 1.9% FPR [8]. With the addition of NT, the detection rate in the United Kingdom increased to 94% with a 0.3% FPR but only to 88 for 0.7% FPR in the United States (χ² = 4.4, p < 0.05). Multiples of median (MoM) for free β-hCG (US 0.22 vs. 0.38 MoM UK) and PAPP-A (US 0.32 vs. 0.20 MoM UK) were comparable between groups, but median NT measurements were sig-

<table>
<thead>
<tr>
<th></th>
<th>Free β-hCG and PAPP-A</th>
<th>Free β-hCG and PAPP-A plus nuchal translucency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>detection rate, %</td>
<td>false-positive rate, %</td>
</tr>
<tr>
<td>USA</td>
<td>82.3</td>
<td>1.7</td>
</tr>
<tr>
<td>UK</td>
<td>84.5</td>
<td>1.9</td>
</tr>
</tbody>
</table>

* Table 1. Comparison of detection and false-positive rates in screening for trisomies 13 and 18 in the USA and UK"
have higher performance than that of a clinically deter-

quality of a ‘laboratory measurement per se’ is likely to

of our study.

rigor of the grading of the scans. This is a limitation

ed, and the NTQR has required 10. It has not been pos-

addition the submission of 50 pictures of scans to be grad-

established in the US. For example, both programs re-

data collection and feedback to providers as it has become

variable, and the NTQR has gone through evolutions of

Model of the FMF has been published many times

performance as well as low coefficients of variation. The train-

standardized with training and credentialing, as has been

in the United Kingdom, the detection rate was increased to 94% with a FPR of 0.3%, whereas in the Unit-
ed States, with the addition of NT measurements, the de-

tection rate was lower, 88.4%, with a higher FPR of 0.7%.

These data are consistent with the well-established

First-trimester detection of trisomies 13 and 18 based

on biochemistry alone demonstrated sensitivities of over

80% for a screen positive rate of <2% and contributed a

higher proportion than NT to the diagnoses of trisomies

13 and 18 when compared to that for trisomy 21 or Turn-
er syndrome. By comparison, the ‘biochemical only’

Down syndrome detection rate is approximately 63% for

a 95% specificity [1, 6]. When NT was added and stan-

dardized with in the United Kingdom, the detection rate was

increased to 94% with a FPR of 0.3%, whereas in the Unit-
ed States, with the addition of NT measurements, the de-

tection rate was lower, 88.4%, with a higher FPR of 0.7%.

These results speak to the general proposition that the

quality of a ‘laboratory measurement per se’ is likely to

have higher performance than that of a clinically deter-

mined measurement such as ultrasound [9]. We have

made a similar observation in the increased birth rate of

Down syndrome infants in Colorado over the last 20

years as amniocentesis and chorionic villus sampling

procedures have decreased, and use of the ‘genetic ultra-

sound’ has increased [10]. The variability of the perfor-

mance of ultrasound measurements parallels the experi-

ence with second-trimester estriol screening in the 1990s.

Whereas the coefficients of variation (CV) for both AFP

and hCG were in the range of about 7%, many estriol as-

says had CVs of 15–20%. Not surprisingly, in retrospect,

those programs with tight estriol CVs showed increased

performance in triple versus double screening, and those

programs with poor estriol assays showed diminished

performance [1]. The same principle applies to NT mea-

surements in that tight distributions of normal data in-

crease the likelihood that abnormal cases will be identi-

fied as such as opposed to being considered within the

normal range.

The data presented here indicated better overall per-

formance of NT measurements in the United Kingdom

than in the United States. Two recent studies, one from

London and another from Denmark, have corroborated a

greater than 90% detection of trisomy 18 for FPRs of

about 0.3% on top of the trisomy 21 FPR [11, 12]. While

is it possible to deliberate about the specific causes of the

difference, the unmistakable conclusion is that better

quality control appears to be in place in the United King-

dom and Denmark than in the United States. Therefore,

ultrasound performance in the United States will likely

continue to lag behind the United Kingdom. These data

suggest that as components of first-trimester screening

programs for detection of trisomies 13 and 18 are eval-

uated: (1) analyses of maternal serum analytes have strict-

ter controls than ultrasound measurements and therefore

more accurate assessments as to their sensitivity and

specificity, and (2) programmatic differences in false-

positive and false-negative rates are likely to be a reflec-

tion of the skill and training of those performing mea-

surements of NT. Given the much larger variability in the

quality of ultrasound measurement among centers in the

United States, mass screening strategies should focus on

optimizing input of the highest quality variables. We

have previously published that even minor underestima-

tions of NT measurements can reduce the sensitivity of

trisomy 21 detection by 18% for the same FPR [13]. Recent

analysis of US NT data shows that the US distribution of

measurements is still skewed to the left with considerable

overrepresentation of cases <5th percentile [14].

---

**Table 2.** Multiples of the median (MoM) by gestational age (GA) in the United States

<table>
<thead>
<tr>
<th>GA, weeks</th>
<th>n</th>
<th>MoM</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>26</td>
<td>2.14</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>1.34</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Discussion

These data are consistent with the well-established laboratory algorithm model in which rigid standardiza-

tion of measurements produces the best possible perfor-

mance as well as low coefficients of variation. The train-

ing model of the FMF has been published many times [6]. The applicability in the US over the last decade has been

variable, and the NTQR has gone through evolutions of

data collection and feedback to providers as it has become

established in the US. For example, both programs re-

quired a didactic session and exam. The FMF required in

addition the submission of 50 pictures of scans to be grad-

ed, and the NTQR has required 10. It has not been pos-

sible to assess the impact of differences in the number of

scans submitted, or whether there are any differences in

the rigor of the grading of the scans. This is a limitation

of our study.

These results speak to the general proposition that the

quality of a ‘laboratory measurement per se’ is likely to

have higher performance than that of a clinically deter-

mined measurement such as ultrasound [9]. We have

made a similar observation in the increased birth rate of

Down syndrome infants in Colorado over the last 20

years as amniocentesis and chorionic villus sampling

procedures have decreased, and use of the ‘genetic ultra-

sound’ has increased [10]. The variability of the perfor-

mance of ultrasound measurements parallels the experi-

ence with second-trimester estriol screening in the 1990s.

Whereas the coefficients of variation (CV) for both AFP

and hCG were in the range of about 7%, many estriol as-

says had CVs of 15–20%. Not surprisingly, in retrospect,

those programs with tight estriol CVs showed increased

performance in triple versus double screening, and those

programs with poor estriol assays showed diminished

performance [1]. The same principle applies to NT mea-

surements in that tight distributions of normal data in-

crease the likelihood that abnormal cases will be identi-

fied as such as opposed to being considered within the

normal range.

The data presented here indicated better overall per-

formance of NT measurements in the United Kingdom

than in the United States. Two recent studies, one from

London and another from Denmark, have corroborated a

greater than 90% detection of trisomy 18 for FPRs of

about 0.3% on top of the trisomy 21 FPR [11, 12]. While

is it possible to deliberate about the specific causes of the

difference, the unmistakable conclusion is that better

quality control appears to be in place in the United King-

dom and Denmark than in the United States. Therefore,

ultrasound performance in the United States will likely

continue to lag behind the United Kingdom. These data

suggest that as components of first-trimester screening

programs for detection of trisomies 13 and 18 are eval-

uated: (1) analyses of maternal serum analytes have strict-

ter controls than ultrasound measurements and therefore

more accurate assessments as to their sensitivity and

specificity, and (2) programmatic differences in false-

positive and false-negative rates are likely to be a reflec-

tion of the skill and training of those performing mea-

surements of NT. Given the much larger variability in the

quality of ultrasound measurement among centers in the

United States, mass screening strategies should focus on

optimizing input of the highest quality variables. We

have previously published that even minor underestima-

tions of NT measurements can reduce the sensitivity of

trisomy 21 detection by 18% for the same FPR [13]. Recent

analysis of US NT data shows that the US distribution of

measurements is still skewed to the left with considerable

overrepresentation of cases <5th percentile [14].
Currently, the more effective quality control lies in the biochemical screening parameters of free β-hCG and PAPP-A as compared to ultrasound measurement of NT. In the case of combined screening for trisomies 13 and 18, rigid oversight of NT measurements with regular feedback, as practiced by the UK system, has demonstrated an effectiveness superior to that currently in practice in the US system. The process of continual assessment with feedback, a well-established principle in business and technology assessments, has considerable applicability with regard to aneuploid screening in the first trimester [15].

References


