

First Trimester Interlaboratory Comparison Program

Distribution 2006 FT-B



Sponsored by:
Department of Pathology and Laboratory Medicine
Women & Infants Hospital
Providence, RI

DISCUSSION

Data Listing and Analysis

Reading the Data Listing: This five page data listing contains a summary of reported results for all participants; one page summarizing each of the five specimens. Your lab ID is listed at the beginning of the row with your results. Missing data (blanks) are likely due to manufacturers who do not screen or laboratories that are not yet offering clinical services. Missing data may also result because some laboratories do not measure 'total or intact hCG' but some other marker. Outliers for gestational age or maternal age are identified as those outside +/- 0.2 weeks (or years) of the correct answer. For the assay results (in mass units or MoM) and Down syndrome risks, outliers are defined as being outside of +/- 2 trimmed standard deviations, after accounting for rounding, and after a logarithmic transformation of risk. In the future, we will implement more sophisticated trimming algorithms. **Comment:** One laboratory reporting integer maternal age consistently is one year lower than the correct answer. This laboratory should examine its software/methodology and make corrections.

Specimen Creation: Three of the proficiency testing specimens in this second distribution of the First Trimester Interlaboratory Comparison Program (ICP) for 2006 were pools of patient sera. FT-07 was a 12 week pool (the same pool as FT-03 in the FT-A distribution). FT-09 was a 13 week pool diluted 2 parts pool to one part normal human serum, and FT-10 was a 15 week pool. FT-06 and FT-08 are artificial samples composed of normal human serum spiked with recombinant hCG, recombinant DIA and a purified native PAPP-A concentrate derived from patient sera (material donated by Diagnostic Systems Laboratories, Webster, TX). FT-06 is the same specimen as FT-05 in the FT-A distribution. The gestational ages assigned to each specimen do not necessarily correspond to the pool gestational age, but were varied to yield a wider range of MoM values.

Variance of Mass Units and MoM Results: The PAPP-A analyte values for the three specimens that primarily consist of patient pools (FT-07, FT-09, FT-10) yield CVs from 9% to 16%. Some of this variability appears to result from systematic differences in calibration between the DSL, DPC, and Delfia methods, but the between-kit differences in values are not large. In contrast, the CVs for the artificial samples (FT-06, FT-08) for PAPP-A are much larger (25% and 44%, respectively). FT-08 was targeted to yield a value of 0.7 mIU/ml (based on assigning a targeted value using the DSL kit), and the consensus value of 0.6 mIU/ml (similar to the DSL consensus value) is close to the target. However, the values for the four PerkinElmer (PE) kit users were all 0.3 mIU/ml, and the PE values for FT-06 were also systematically lower than for the DPC and DSL kits. This suggests that the PAPP-A spike may not accurately reflect what is obtained with real patient specimens, where the PE kit yields values similar to DSL and DPC. If confirmed, this will complicate comparing MoM values with artificial samples, and may require identifying a different source of PAPP-A. In contrast, the hCG CVs for all five specimens are similar (9% to 13%), suggesting that recombinant hCG reflects the specificity observed for patient samples.

The CVs for PAPP-A MoM values for the manufactured specimens (27% for FT-06 and 44% for FT-08) are higher than for patient pools (12 % to 16 %) reflecting the relatively high CVs for the PAPP-A values themselves. The CVs for the hCG MoM values in patient pools versus artificial samples are similar (8% to 15%).

The CVs of the risk values (on a log scale) are relatively low for the real patient samples (3% to 9%), but higher for the artificial samples (14% and 19%). The higher CVs for the artificial samples reflect the variability in both values and Mom for these specimens. Interpretations for the three real patient samples were consistent. All participants called them screen negative (not surprisingly, since the three consensus risks were all relatively low). FT-06 interpretations were also reasonably consistent (12 of 13 screen positive). In contrast, FT-8 received mixed interpretations, with 6 of 13 labs interpreting the sample as screen positive and 7 as screen negative. This may reflect the fact that the consensus risk is 1:250, close to commonly used screening risk cut-off levels.

Long term ICP control (12 week pool): Specimens FT-03 from the 2006 FT-A distribution and FT-07 from the 2006 FT-B distribution were prepared from the same 12 week patient pool. This allows a direct comparison of results both for the overall consensus and for individual laboratories. The consensus PAPP-A value for FT-03 of **2.70 mIU/ml** showed very close agreement with the consensus of **2.80 mIU/ml** for FT-07 (CV of 0.6%). However, 11 of 14 laboratories reported increased PAPP-A values for FT-7, suggesting that the increase in average value, although slight, may be real.

The consensus hCG value for FT-03 of **84.6 IU/ml** was also close to the consensus for FT-07 of **86.1 IU/ml** (CV of 0.3%). Five of twelve reported an increased value, suggesting that the observed increase may be due to chance. This excellent agreement suggests that both PAPP-A and hCG are reasonably stable in patient pools.

The median within-laboratory CV for those participants reporting PAPP-A values for both FT-03 and FT-07 was 1.6% (range 0.6% to 2.8%, with one outlier of 9.5%). For hCG values the median within-laboratory CV for the same two samples was 1.2% (range 0.2% to 5.4%). Overall, this shows excellent consistency in the values over a period of three months for individual assays.

Interpretative Questions: Racial adjustment of analyte value

Does your laboratory provide clinical test results for Down syndrome screening? Of the 19 participating laboratories answering the interpretative question (one laboratory did not respond), 15 provide clinical test results. The subsequent analyses will be limited to these participants.

Does your laboratory adjust PAPP-A values if the patient is African-American? Thirteen of the 15 (87%) reported they do not make a racial adjustment of PAPP-A values. Two did adjust by dividing by either 1.38 or 1.34. Each used different software.

Several programs have reported the median PAPP-A MoM level in African-Americans (Afro-Caribbean for European programs). Table 1 summarizes these studies. Little difference in the effect was seen with/without maternal weight adjustment. Overall, the consensus PAPP-A MoM level in African Americans is higher than in Caucasians. However, there is considerable heterogeneity between the two studies from the UK (about 54% higher) and the one from the US(12% higher).

Conclusion: Laboratories should consider adjusting the higher PAPP-A values found in African Americans. The amount of correction is unclear, but the average level found in this racial group is likely to be between 12% and 50% higher than in Caucasians.

Table 1. Summary of published studies of PAPP-A levels in African Americans

Study	PAPP-A MoM ¹ (observations)		Approximate 95% CI
	Caucasians	African Americans	
Spencer <i>et al.</i> , 2000	1.00 (5,422)	1.48 (752)	1.52 to 1.54
Krantz <i>et al.</i> , 2005	1.00 (42,611)	1.12 (2,682)	1.10 to 1.14
Spencer <i>et al.</i> , 2005	1.00 (61,219)	1.55 (2,942)	1.52 to 1.58

¹ After adjusting for maternal weight

References

- Spencer K, Ong CY, Liao AW, Nicolaides KH. The influence of ethnic origin of first trimester biochemical markers of chromosomal abnormalities. *Prenat Diagn* 2000;20:491-494.
- Krantz DA, Hallahan TW, Macri VJ, Macri JN. Maternal weight and ethnic adjustment within a first-trimester Down syndrome and trisomy 18 screening program. *Prenat Diagn* 2005;24:635-640.
- Spencer K, Heath V, El-Sheikhah A, Ong CY, Nicolaides KH. Ethnicity and the need for correction of biochemical and ultrasound markers of chromosomal anomalies in the first trimester: a study of Oriental, Asian and Afro-Caribbean population. *Prenat Diagn* 2005;25:365-369.

Dimeric inhibin-A: DIA measurements were reported by five participants. All reported using the same method (Di-01 - DSL). The following table provides the reported DIA values and MoM levels for the three laboratories for all five samples. The target value for FT-08 was 550 mIU/mL, but the five laboratories all reported values 10 times higher. This was due to an inadvertent error that resulted in a recombinant inhibin spike that was 10X that required to obtain the target value. However, this has provided an interesting test of the accuracy of the dilution of DIA. It is remarkable how close the values are to each other considering the need for dilution to allow the value to fall on the standard curve.

Dimeric inhibin-A measurements for FT-B, 2006

Sample Number	Laboratory	Value ²	MoM	DS Risk (1:n)	DIA LR ¹
FT-06	A	346			
	B	411	1.45	110	1.26
	C	335	1.23	536	1.36
	D	433	1.66	69	1.19
	E	337			
FT-07	A	252			
	B	297	1.20	1000	1.16
	C	213	0.86	4900	2.47
	D	278	1.32	1190	1.50
	E	291			
FT-08	A	5588			
	B	5850	25.5	61	0.24
	C	5533	22.9	19	0.02
	D	5144	26.2	9	0.07
	E	5322			
FT-09	A	162			
	B	188	0.60	4200	1.24
	C	136	0.51	<10,000	-
	D	185	0.71	8590	1.84
	E	163			
FT-10	A	270			
	B	313	1.09	4800	0.76
	C	231	1.05	1170	1.13
	D	304	1.77	387	0.18
	E	260			

¹ The increase/decrease in risk from the combination of NT, PAPP-A and hCG, divided by the risk that includes DIA measurements. In some instances, a capped risk (e.g., <1:10,000) was reported and it was not possible to compute the likelihood ratio.

² Four of these laboratories also reported DIA values for FT-03 (the same pool as FT-07) in the A distribution of 2006 (CVs ranged from 2.3% to 15.4%).

Supplemental Topic: Computing sonographer-specific NT medians

Background

As NT measurements in the first trimester become more widespread, screening laboratories are faced with the issue of interpreting the results from the sonographer (NT in mm and CRL in mm) with biochemical test results to create a pattern-specific Down syndrome risk. Although some effort has been expended to try to have all sonographers measure NT translucency in exactly the same way, these efforts are not completely successful. Below are three figures from two different first trimester screening programs in Europe. The first, published in 2002 by Crossley and her colleagues, is from the Glasgow area. Figure 1 shows NT median MoM values from 13 hospitals in which all sonographers were trained by the Fetal Medicine Foundation - FMF (K Nicolaides). The median NT MoM (short horizontal line) is usually within 10% of 1.00 (the expected value), but several hospitals vary widely (range 0.68 to 1.4 MoM). Figure 2, from the same group, shows NT measurements from four sonographers at a single hospital, demonstrating that there are systematic differences between sonographers even within a group.

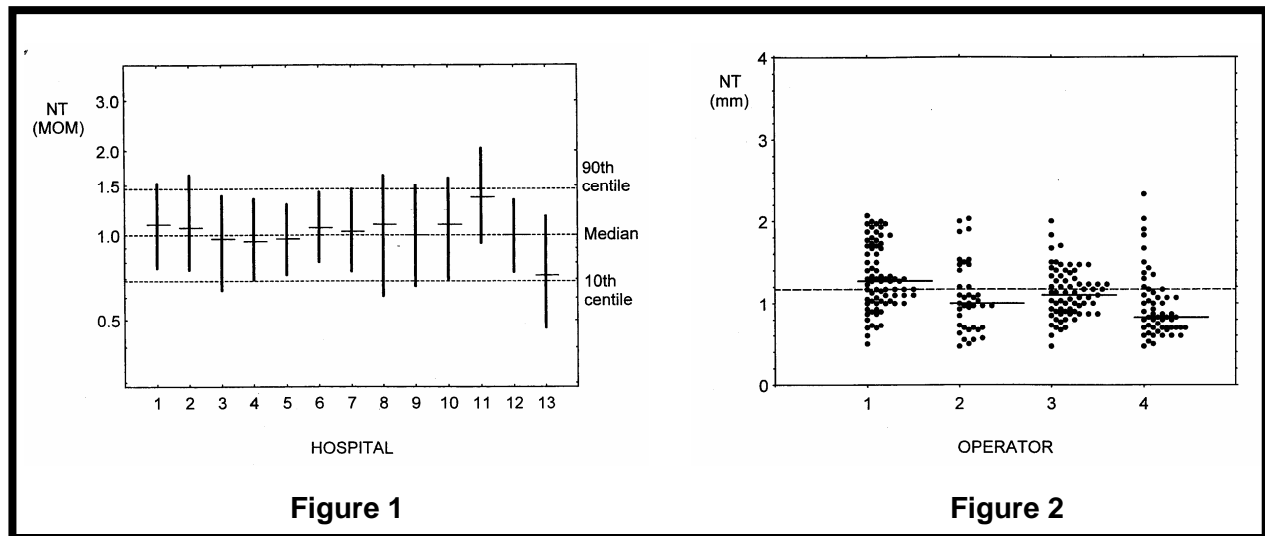
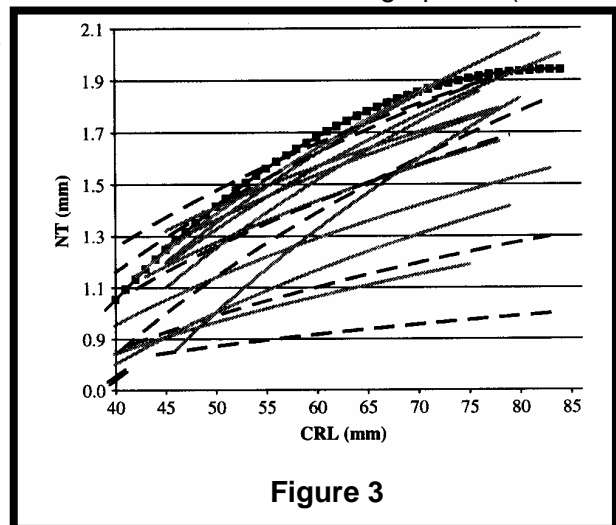


Figure 3 comes from a recent study in the Netherlands. It shows the FMF expected median NT values (connected boxes) with regressed medians from FMF trained sonographers (solid lines) and non-FMF trained sonographers (dashed lines). Nearly all of the sonographers provided NT measurements systematically lower than the levels expected by FMF. This means that NT MoM values will be significantly lower than if sonographer-specific median NT values were used. This would result in lower detection rates for Down syndrome (and lower false positive rates than would be projected from published studies (e.g, BUN, SURUSS, and FASTER). Together with the unpublished experience of other screening laboratories in the U.S., it is clear that laboratories need to account for systematic differences between sonographers when interpreting NT measurements for Down syndrome screening.



In effect, using published NT median values (from, for example, the FASTER study or the Fetal Medicine Foundation) or a single set of NT median values for all centers or sonographers is equivalent to using one set of 'package insert' median values for all manufactured kits for a serum marker. From the inception of serum-based screening, it was evident that the goal of having all kits give the same values was unrealistic, and kit-specific median values (and sometimes lot-specific median values) are required for quality screening performance. Monitoring NT values and calculating center- and sonographer-specific NT medians will be a challenge for the laboratory, but it is a responsibility that may have to be accepted, if screening protocols using NT measurements are to achieve their potential. The aim of the current exercise is to help laboratories create the tools and expertise to allow the use of sonographer (or center-specific) medians.

Supplemental Questions

A total of 15 laboratories report providing clinical services for Down syndrome screening. All answered one or more of the supplemental questions.

Question 1. Does your laboratory convert NT measurements (in mm) to multiples of the median (MoM) as part of clinical service? Fourteen of the 15 laboratories (93%) reported using MoM as the interpretive unit. The one laboratory not using MoM units did not answer any of the additional questions.

Question 2. To convert NT measurements to MoM levels, are your medians

- | | |
|---|---------------|
| a. always based on a single median equation / table of median values | 4 of 14 (29%) |
| b. sometimes based on sonographer – or center-specific median values | 1 of 14 (7%) |
| c. nearly always based on sonographer- or center-specific median values | 9 of 14 (60%) |
| d. not sure | 0 of 14 (0%) |

Question 3. Given that you can use sonographer- or center-specific medians, are these

- | | |
|---|---------------|
| a. computed by my clinical software and automatically available for use | 5 of 10 (50%) |
| b. computed off-line and entered into my clinical software | 5 of 10 (50%) |
| c. other | 0 of 10 (0%) |
| d. not sure how this is handled | 0 of 10 (0%) |

Question 4. Perform a regression analysis of the 150 data points in the attached EXCEL spreadsheet that provide reference data for NT measurement from sonographer (or center specific medians) "FSR".

- the regression equation is: Of the 14 laboratories eligible for this analysis, eight reported coefficients for equations of the form "median = A*B^{cr1}", two reported coefficients for equations of the form "log (median) = M*weeks + B" and one reported coefficients for a quadratic equation for GA. The remaining three laboratories did not report coefficients.
- the slope per week is: Of the 14 laboratories eligible for this analysis, 12 (86%) reported a slope per week. Slopes of the NT vs GA equations ranged from 17.3% per week to 24.1% per week. The median slope was 19.0%; the average slope was 19.9%.
- the median NT MoM of the 150 samples is: Of the 14 laboratories eligible for this analysis, 10 (71%) responded. Median NT MoM levels ranged from 0.96 to 1.10. The grand median MoM was 1.00; the average of the 10 median MoMs was 1.01.
- the logarithmic standard deviation (SD) of the 150 NT MoM levels is: Of the 14 laboratories eligible for this analysis, 8 (57%) responded. The log SDs reported ranged from 0.1198 to 0.3000. The median log SD was 0.12; the average of the log SDs was 0.15.

Comments:

Forms of the equations: There are two forms of the log-linear equation that can be used to fit log NT measurements (y-axis) versus gestational age (on the x-axis either in CRL-mm, GA-days or GA-decimal weeks). These are essentially the same equations that fit AFP and uE3 measurements in the second trimester. Below is some algebra that shows the relationship between the two forms. For the example provided, we have computed $A=0.103459$ and $B=1.197714$ from the exponential equation for CRL, and $b=-0.98447$, $m=0.078353$, when NT is regressed versus gestational age in decimal weeks. Figure 4 shows the 10 equations superimposed on the 150 NT measurements (and associated CRL and GA measurements) sent to participants as part of the FT-B distribution. One laboratory (line above all 150 data points) apparently transferred its coefficients incorrectly to the reporting form, as its answers to parts b, c and d were correct. When using the form of the equations AB^{CRL} , make sure that the exponentiation (B^{CRL}) occurs prior to the multiplication $A * B^{CRL}$). Also, notice the linear (or near linear) relationship between CRL and gestational age in this range (13 mm per week between 11 and 13 weeks' gestation). The corresponding CRL measurements in the figure are derived from Hadlock et al., 1992.

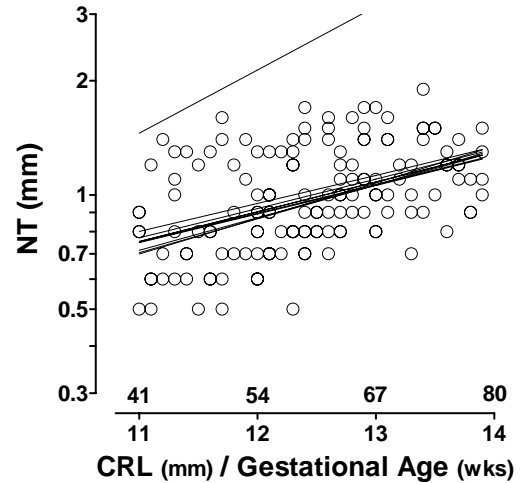


Figure 4

Table 3 Showing the equivalence of two forms of the equation relating NT to gestational age

$\text{median NT} = AB^{GA}$ $\log(\text{median NT}) = \log(AB^{GA})$ $\log(\text{median NT}) = \log(A) + GA(\log(B))$ $\text{median NT} = 10^{(\log(A) + GA(\log(B)))}$ $\text{median NT} = 10^{(b + mGA)}$	<p>A and B are coefficients, GA is gestational age take the logarithm of both sides do the math do the math b = intercept, m = slope, GA is gestational age $b = \log(A)$ and $m = \log(B)$</p>
--	--

Slope (increase per week): The 150 data points that formed this exercise were created using a random number generator that relied on the grand NT median equation from the FASTER trial (hence the code 'FSR'). The actual percent increase per week for that equation is about 20% per week (question b), but the relatively small number of observations (150) means that there will be some difference in the observed increase per week. In order to convert CRL to GA (as we had done for you in the spreadsheet), the laboratory can choose a conversion equation (we used Hadlock et al., 1992, Table 3 – the most common equation used by laboratories reporting in FT-C 2005). Figure 4 shows the data plotted, with the slopes being quite similar for all but one.

Median NT MoM levels: The expectation is that when a new set of medians is derived from a dataset, then the resulting median MoM will be close to 1.00. In the accompanying worked out example, our median NT MoM was 0.98. The consensus grand median MoM for NT of 1.00 among the 10 responding laboratories confirms this expectation. The laboratory reporting a 1.10 as the median should probably check its calculations, as this is relatively far from the expected value.

Logarithmic standard deviation: According to two published sources (Wald *et al.*, J Med Screen 2004;10:56; Spencer *et al.*, Ultra Obstet Gynecol 2003;22:142), the log SD of NT measurements in unaffected pregnancies varies by gestational age. We used a log SD of 0.13 at 11.0 weeks and varied it to 0.11 at 13.9 weeks. Thus, the overall log SD is expected to be about 0.12. The median of the SD values reported by the eight laboratories is 0.12. The laboratory with a value of 0.3 may not have transformed the data prior to calculating the standard deviation. Standard deviations are the square root of the variance (the sum of the squared differences between the mean and the observed value) and, therefore, cannot be negative (as one laboratory reported).

George J. Knight, Ph.D.
Jacob A. Canick, Ph.D.

Glenn E. Palomaki, B.S.
Geraldyn M. Messerlian, Ph.D.

(207) 657-7888
(401) 453-7654

Department of Pathology and Laboratory Medicine
Women and Infants Hospital, Providence, Rhode Island