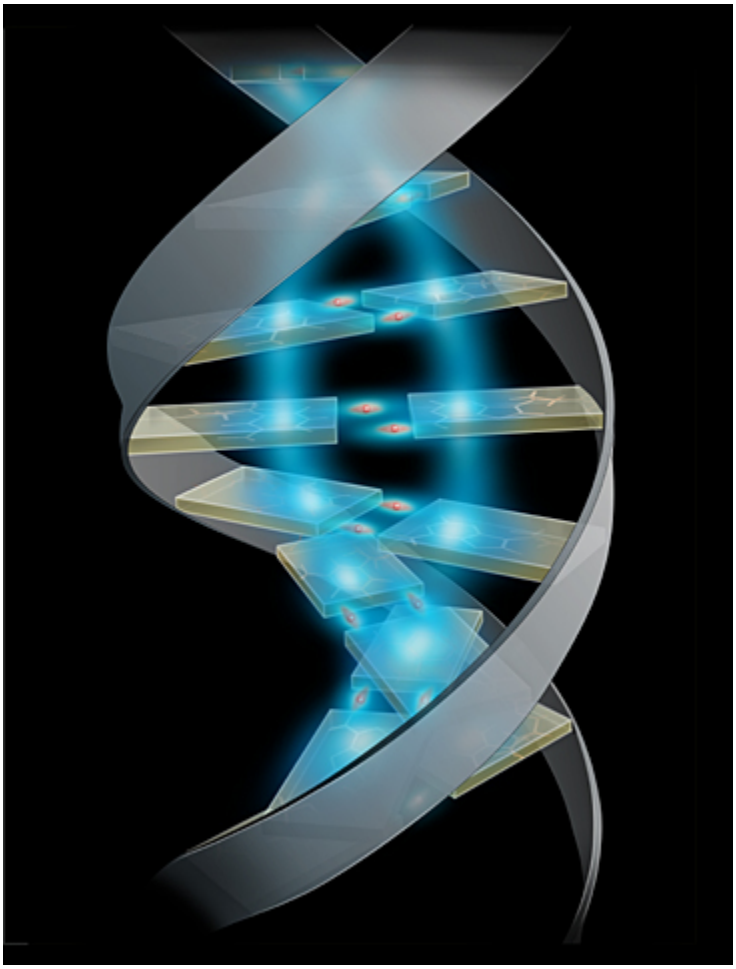
First Trimester
Interlaboratory Comparison Program

Distribution 2009 FT-B



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

INTRODUCTION

Explanation of Data Listing and Analysis

Specimen Options: The ICP offers two choices for specimens. The first specimen set is designed for those participants using hCG in their screening marker combination (**hCG sample set**). The second set is designed for those participants using free beta subunit in their screening marker combination (**free beta sample set**). The specimens may consist of: 1) unmodified patient pools, 2) patient pools diluted with normal human serum and spiked with recombinant hCG or recombinant free beta subunit (but not both) and PAPP-A concentrate (and inhibin), or 3) normal human serum spiked with recombinant hCG or recombinant free beta subunit (but not both) and PAPP-A concentrate.

The PAPP-A target concentrations in both sets are the same for all five specimens, allowing for a unified evaluation for both specimen sets. However, participants should not test or report hCG measurements made in the free beta set, nor free beta measurements in the hCG set. This can result in spurious results for the following reasons:

- *In hCG specimen sets*: specimens spiked with recombinant hCG can yield a very high non-physiologic level of measured free beta subunit (the recombinant hCG preparation has a significant amount of free beta present).
- *In free beta specimen sets*: some free beta specimens are spiked with recombinant free beta subunit and others are made by spiking normal human serum with recombinant free beta subunit and PAPP-A concentrate. However, no recombinant hCG will have been added to either type of sample, so measurement of these specimens for hCG is not appropriate.

A limited number of additional samples sets are available upon request (free of charge) so that laboratories considering switching from hCG or vice versa can receive both the hCG and the free beta sample sets.

Reading the Data Listing: The five page data listing (in a separate pdf file) contains a summary of reported results for all participants, with each page summarizing one specimen. Your laboratory identification number (ID) is listed at the beginning of the row with your results. Missing data (blanks) are likely due to participants who are manufacturers rather than screening labs, or to laboratories that are not yet offering screening services. 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. A logarithmic transformation is used for the analysis of Down syndrome risks.

Conversion of reported risks to first trimester risks. Most laboratories report first trimester risks, but some laboratories report second trimester or term risks. If the reported risks are not first trimester, these risks are displayed in the column labeled "**Report**" under the "**Down S Risk (1:n)**" heading. To allow all risks to be evaluated by a single statistic, second trimester risks are converted to first trimester risks using the factor 0.74. This accounts for fetal loss between the first and second trimesters (43% from first trimester to term and 23% from second trimester to term). For example, if the second trimester risk is 1:1000, the first trimester risk is 1:1000 x 0.74, or 1:740. Term risks are converted to first trimester risks by multiplying by 0.57.

Down syndrome risks from participants using the free beta sample set are listed in the data sheets, but are not included in the calculation of summary statistics. When sufficient numbers are available, a separate analysis will be performed.

Maternal Age Reporting: Maternal age can be reported either as a decimal or as completed years (integer). Although the difference in risk is small for most ages, use of decimal age rather than completed years can be more important for an older woman, especially one whose age falls close to a whole year (*e.g.*, 34.1 versus 34.9 years). Each of these women would be considered to be 34 completed years, even though they are almost one full decimal year different. Laboratories commonly calculate risk using a maternal age equation rather than a table of risks, and it is straightforward to use the more accurate risk to obtain better precision. Almost all labs in the ICP report decimal age. Currently, the lab(s) that report integer maternal ages are listed separately on the data summary results, but in the future such results will be listed along with decimal ages, but will not be included in the calculations.

NT MoM Reporting: The ICP provides a target NT MoM for most challenges. Participants need to generate the MoM values provided in the histories by trial and error, usually by entering various combinations of CRL/NT/GA values. Approximate CRL values (in mm) and GA values (in weeks and days) are provided as an aid. Participants are asked to report the MoM value that they actually obtained to serve as a check on how reliably they could reproduce the targeted MoM value. Almost all participants report NT MoM values that closely match the targeted value. If participants are having difficulty generating a reliable MoM, we can provide assistance.

The ICP also includes at least one challenge that provides a patient CRL and NT value (in mm), along with a set of NT and CRL values from the submitting 'hypothetical' sonographer (identified by initials) who provided those measurements. Participants can then use that set of sonographer-specific NT/CRL values to generate NT medians for use in converting the NT values (in mm) to MoM. That NT MoM is then used along with maternal age and the chemistry results to calculate the patient-specific Down syndrome risk. We also provide an Excel spreadsheet that can be used to calculate the CRL/NT median equation with accompanying quality assurance parameters (e.g. slope and log standard deviation).

Labs that do not use the MoM for interpretation of NT will only be evaluated for analyte values.

Greater than and less than risks: Risks that are reported as less than (<) or greater than (>) are displayed in the "**Report**" column under the "**Down S Risk (1:n)**" column. These risks are listed as the actual numeric risk in the "1st trim" column and may be included in the final calculation of the consensus risk.

Free Beta Subunit Results: The data listings include the analyte and MoM values for the free beta measurements for those laboratories using the free beta specimen set. A median is reported, but a comprehensive analysis is not performed due to small numbers. However, each of these participants can review their own results by inspection of the data listing. Currently, all participants receive the free beta sample set report in term risks and these are listed in the "**Report**" column under the "**Down S Risk**" heading. It is important to remember that these term risks would not be comparable to the risks listed in the "1st trim" column, even if converted to 1st trimester risks, because of the use of free beta subunit rather than hCG in the calculation of risk.

PAPP-A Units: Until recently, all laboratories reported PAPP-A analyte values in mIU/mL, allowing for a unified analysis of all kit methods. Recently, Beckman has introduced a PAPP-A methodology for their Access instrument that reports values in ng/mL. A preliminary analysis relating mIU/mL to ng/mL has been performed (see Supplemental Analysis below). Some of these laboratories report in µg/mL, and these have been converted to ng/mL for analysis. In the summary reports for each specimen the values for the Beckman methodology (assigned the new code of Be-01) are surrounded by light borders but are not included in the all lab consensus.

Values in Boxes: The ICP uses two types of boxes in the individual data listings.

- Thin lined boxes are used to identify values that call attention to values that are significantly different from the consensus but are not considered outliers (e.g., 1.07). For example, a group of laboratories appear to use only a single set of median NT values (rather than sonographer-specific reference ranges) for calculating MoM values. These differ significantly from the results reported by participants using sonographer-specific medians, but are not considered outliers.
- Thick lined boxes are used to identify values that are outliers as compared to the group consensus (e.g., 25.0).

RESULTS

FT-06 and FT-06fb:

Participants were asked to calculate an NT MoM value, given a CRL of 70 mm (~ 13.1 weeks' gestation) and an NT value of 1.8 mm submitted by sonographer "FST". Participants were provided with a set of 150 NT/CRL measurements for FST and may have already calculated a sonographer-specific median equation (sent again in this distribution for those who may need to recompute the median equation). However, participants may or may not have used those medians to calculate their MoM value, depending on their own laboratory protocols. The expectation is that the resulting MoM values reported by laboratories that use sonographer-specific medians should be similar, while those using a single fixed set of NT medians might be different. We calculated the median equation for sonographer FST to be: median NT= $10^{(-0.373+0.00613*CRL)}$ using the Excel calculator supplied to participants. This equation yields an expected median NT value of 1.13 mm for a CRL of 75 mm, which results in an NT MoM value of 1.59 (1.8/1.13). The consensus NT MoM (calculated as the trimmed mean value) value is 1.58, in agreement with expectation. Four results (light boxes) differ from the consensus (1.02, 1.04, 1.11, and 1.16 MoM). Three of these participants indicate that they do not use sonographer specific medians. Rather, they use a single set of medians, which likely accounts for the difference. The fourth indicates that their source for NT medians is unknown. Overall, all laboratories can derive a median equation given sonographer NT/CRL values, and most can (or do) use those medians when providing clinical interpretations.

A CRL of 70 mm was provided for this sample, requiring each participant to calculate gestational age. Assigned gestational ages for FT-01 ranged from 12.9 to 13.3 weeks. As discussed in the 2009 FT-A report, differences most likely reflect the 'CRL to decimal weeks' equation selected for use by each laboratory.

The CVs of the PAPP-A value and MoM were unremarkable as compared to that typically seen for other specimens (12% and 22%, respectively), while the corresponding CVs for the hCG values and MoM were tighter (10% and 11%, respectively). In general, both hCG analyte and MoM values consistently have lower CVs than do PAPP-A results. The low CVs for hCG values are noteworthy, given that all methodologies are analyzed as a group without stratifying by kit manufacturer. The CV of log risk was low (9%). This trimmed mean risk was 1:930 and all labs considered it screen negative.

FT-07 and FT-07fb:

This specimen is a pool of sera from 12 week pregnancies, and should, therefore, more accurately reflect actual between-lab and between-kit differences than manufactured samples. The CVs for the PAPP-A values is low (9%), while the CV for MoM is distinctly higher (20%). The CVs for the hCG value and MoM values are typically low (11% and 9%, respectively). It is noteworthy that the CV for the hCG MoM is as low as for the values themselves, suggesting that laboratories

have reliable hCG median values. The trimmed mean risk was very low (1:4800), and all participants reported it to be screen negative.

FT-08 and FT-08fb:

The CVs for PAPP-A analyte and MoM values are 9% and 20%, respectively. The consensus value for hCG is 226 IU/mL, and the CV is 17%. This CV is somewhat higher than that calculated for specimens with values around 100 IU/mL, indicating that hCG methods may be more variable at high values (see FT-10) than at low values. The consensus risk was 1:120, with three labs reporting risks considerably higher than the consensus (*i.e.*, 1:415 to 1:430). These three labs reported the specimen as screen negative while all but one other participant considered it to be screen positive.

FT-09 and FT-09fb:

The CV for PAPP-A values was 11%, which is similar to the CV of 9% for specimen FT-07, both of which are patient pools. The CVs for both hCG values and MoM levels were typical (11% and 14%, respectively). The consensus risk for this 26.1 year old woman was very low (1:6400), and all participants considered it screen negative. One outlying risk at 1:937 was identified.

FT-10 and FT-10fb:

This specimen was targeted to have a low PAPP-A and a very high hCG value. The trimmed mean values of 0.61 mIU/mL and 372 IU/mL, respectively, were consistent with these targets. The CV of the PAPP-A values was reasonable (17%), particularly considering the low target. The CV of log risk (14%) was distinctly higher than is typical for other samples. This may reflects the fact that the PAPP-A trimmed mean MoM of 0.39 is very low and corresponding hCG MoM of 4.27 is very high, falling at the extreme of the population distribution of values. Small differences in MoM values can yield relatively large differences in the likelihood ratios that are used in the risk calculation. The trimmed mean risk was 1:68 and all but two participants called the specimen screen positive.

Dimeric inhibin-A (DIA)

First trimester DIA measurements were reported by four participants. Table 1 lists the reported DIA values and MoM levels for each of the five samples. DIA values show reasonable between-method and between-lab agreement. MoM levels are more variable, but show good between-lab agreement with a few exceptions. Included in the table are the DIA likelihood ratios (LR) in the context of the other markers.

Table 1. Dimeric Inhibin-A results for FT-B 2008

Sample No.	Lab	Method	Value[1]	MoM	DS Risk (1:n)	DIA LR[2]
FT-06	A	DSL Elisa	303	1.38	450	0.93
	B	Beckman Dxl	279	1.24	7200	0.83
	C	Beckman Dxl	280	1.17	681	0.58
	D	Beckman Access	269	1.02	1690	0.48
FT-07	A	DSL Elisa	320	1.37	4300	1.00
	B	Beckman Dxl	301	1.25	10000	0.89
	C	Beckman Dxl	269	0.97	12800	0.45
	D	Beckman Access	289	0.99	9200	0.49
FT-08	A	DSL Elisa	485	2.15	87	0.85
	B	Beckman Dxl	542	1.84	530	0.81
	C	Beckman Dxl	559	1.61	363	0.81
	D	Beckman Access	529	1.48	123	0.78
FT-09	A	DSL Elisa	382	1.07	7800	0.88
	B	Beckman Dxl	343	1.14	9000	0.79
	C	Beckman Dxl	343	0.83	9660	0.67
	D	Beckman Access	345	0.94	11900	0.75
FT-10	A	DSL Elisa	941	3.10	60	1.33
	B	Beckman Dxl	1012	3.53	31	2.03
	C	Beckman Dxl	1030	2.76	204	1.18
	D	Beckman Access	1018	2.92	39	1.44

[1] Rounded value

[2] For each participant, the DIA LR is computed by dividing the reported risk for NT, PAPP-A and hCG by the risk that also includes DIA measurements. If blank, the likelihood ratio cannot be reliably determined, usually because one (or both) of the risks are very high (e.g., >1:10) or very low (e.g., <1:10,000).

Interpretative questions: Weight adjustment of PAPP-A MoM values

A total of 25 participants reported providing clinical services and responded to the interpretative questions. One of these (a screening laboratory) reported not adjusting PAPP-A results for maternal weight. This is not considered good practice, and that laboratory should considered taking weight into account in their interpretations as soon as possible. Among the remaining 24 participants, 12 reported using the log-linear model (or its alternative exponential form). Another 11 reported using the reciprocal model, and one used a fourth order polynomial. Sample FT-07 was from a 160 pound woman, and the consensus PAPP-A MoM (with weight adjustment) was 1.51. In order to determine how well each laboratory's weight adjustment equation fitted actual data, we mathematically adjusted each laboratory's reported FT-07 PAPP-A MoM to read 0.92, about equal to the expected value in a general pregnancy population with an average weight of about 150 pounds. Each of the other results from that laboratory was then adjusted according to reported result so that the ICP's 'computed' weight adjustment line would go through 1.00 MoM at 150 pounds. Figure 1 shows the resulting weight adjustment curves for the 12 laboratories using log-linear (Figure 1A) along with the remaining 12 results, usually the reciprocal (Figure 1B) models. In both figures, a set of observed PAPP-A MoMs from several thousand women in various weight groupings are shown (filled circles with 95% confidence intervals). The maternal weight adjustment relationship for each laboratory is represented as a line/curve.

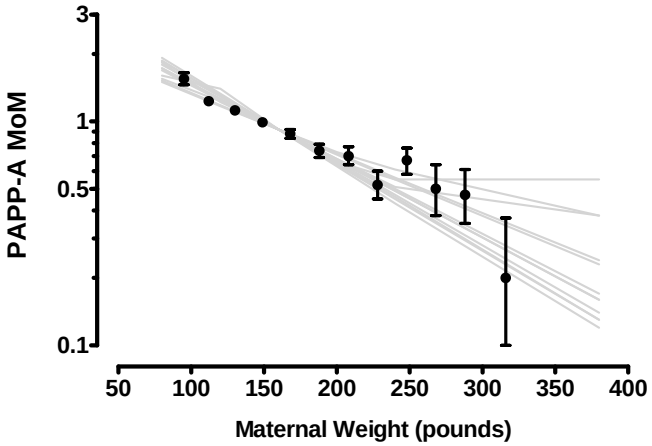


Figure 1A

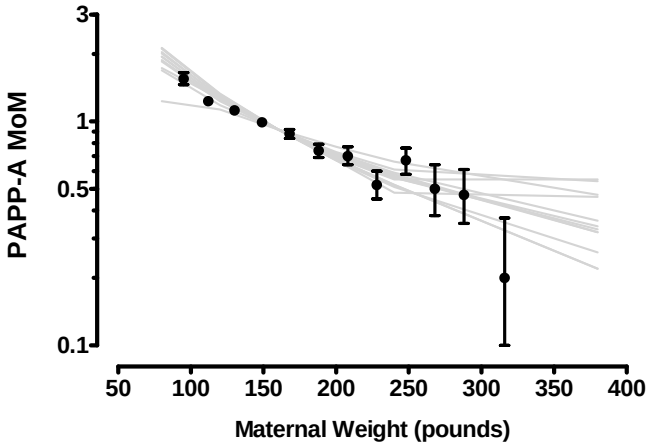


Figure 1B

In Figure 1A, the linear nature of the model (on a log scale) is apparent over a broad range of maternal weights. In practice, truncation limits are used to ensure that reliable adjustments are made at the extremes of weight. Only three of the 12 laboratories using this model appear to be utilizing any truncation limits. These can be identified by the 'break' in the straight line at higher weights. A reasonable upper truncation limit might be 240 pounds. Nine participants should consider implementation of such a limit (they can easily be identified, as they each reported a weight adjusted PAPP-A MoM of 5 or higher for the 380 pound women).

In Figure 1B, the curvilinear nature of the reciprocal model is apparent, along with the 'leveling off' of the effect at higher maternal weights. This makes the choice of truncation limits less important, as the model simply won't create extremely large adjustments. However, one can see that some laboratories do indeed employ truncation limits at the higher weight (indicated by the break in the curve). A reasonable upper limit might be 320 pounds. The one laboratory using a polynomial model can be seen with an underestimate at higher MoM levels. Given that model was reportedly based on small numbers of observations, this laboratory should consider recomputing the coefficients with a larger sample set and, possibly, converting to the reciprocal model.

We encourage laboratories to compute their own population-specific PAPP-A (and hCG/free beta) weight equation. However, several laboratories reported doing this with insufficient numbers of samples (i.e., 120, 237, 200, and 352). If sample sizes less than 1,000 are used, they should be carefully compared to existing equations to ensure reliability.

Supplemental analysis: PAPP-A units comparison

The PAPP-A results from the new Beckman assay (Be-01) appear to be quite different from the other assays currently in use. For example, in the data listing for specimen FT-06 in this distribution the consensus estimate is 0.65 mIU/mL, but the three Be-01 users report results between 204 and 220 ng/mL (Table 2). In some instances, the Be-01 laboratories actually reported in µg/mL, but these were multiplied by 1,000 to yield the equivalent ng/mL value to avoid further complicating the report.

No PAPP-A reference material is available to manufacturers of new PAPP-A kits and Beckman has chosen to use an internal mass standard calibrated in ng/mL. In this distribution, we are not grading the Be-01 users (the thin lined boxes that indicate items for your attention, while the thick lined boxes indicate outliers). Table 2 makes a comparison between the results in ng/mL with the consensus of the remaining laboratories reporting in mIU/mL, stratified from lowest to highest value. The ratios range from 330 to 378 ng/IU, with some indication that the ratio is lower below 1 mIU/mL. A Be-01 user might consider converting PAPP-A results in ng/mL to mIU/mL by dividing by their results in ng/mL by 362 ng/IU. This is not recommended for clinical practice, as we do not have long term experience with this ratio. It is important to note that the Be-01 user MoM levels are consistent with the other laboratories reporting results in mIU/mL – again supporting the practice of internal generated reference ranges (medians).

We will continue to monitor the ratio and if it remains constant, the Be-01 users will have their values converted to mIU/mL so they can be included in the analysis. We stress, however, that this conversion should not be part of clinical practice. Be-01 users should compute new reference ranges in ng/mL (or µg/mL). If properly done, the PAPP-A MoM levels using the Beckman calibration are appropriate for clinical use.

Table 2 Comparison of PAPP-A results in mIU/mL and ng/mL

FT-2009	FTB-10	FTB-08	FTB-06	FTB-07	FTB-09
Lab ID – Be-01					
1109	220	351	628	1001	2667
1113	204	332	619	930	2536
1120	220	350	610	980	2670
Average (ng/mL) A	215	344	619	970	2624
Consensus (mIU/mL) B	0.65	1.00	1.60	2.60	6.95
Ratio A/B	330	344	387	373	378
362 ng/mIU = A/B					

Interpretive Questions – Integrated Screening for Down Syndrome.

Q6. Does your laboratory perform integrated risk interpretations? Among the 25 participants that offer Down syndrome screening services, 18 reported offering integrated screening (or at least an integrated risk) as part of a formal program, five others do not, and two did not respond. Of the 18, all use the quadruple test as their standard second trimester test and all use quadruple markers for the serum integrated test. However, four participants chose to use the triple markers when performing a full integrated test. In addition, two of these four laboratories use the free beta subunit instead of hCG. Consequently, the integrated risks from these participants may not be directly comparable with the 14 laboratories that use the quad test and hCG. However, we are trying to choose samples in which the hCG and free beta subunit MoM levels are similar, and where the likelihood ratio for DIA is relatively close to one. This would minimize these differences. Down syndrome risks for these 18 participants are displayed in Table 3.

Q7/8. Using the Down syndrome risks from FP-08 (CAP FP-B 2009 Survey) to report integrated risks using FT-08 results (after appropriate modifications to the draw date). Analyses will focus on the ratio of the risks or likelihood ratios (last three columns) rather than the reported risks. This makes comparison between the laboratories more meaningful. Although there are a few outlying likelihood ratios, and the distributions are rather broad, the LR's, in general make sense. For example:

- The PAPP-A LR comparison (quadruple risk/serum integrated risk) is 0.53, or nearly a two-fold reduction in risk. Given that the median PAPP-A MoM for FT-08 is 1.18, this is reasonable (remember, even a 1.00 MoM will reduce the risk).
- The NT LR comparison in the last column (serum integrated/full integrated) is 0.39, or about a 2.6 fold reduction in risk. Given that the median NT is 1.20, this again is reasonable (remember, the LR = 1 will occur roughly mid-way between the unaffected and Down syndrome medians, or at about 1.5 MoM).
- The NT & PAPP-A LR in the middle column (quadruple/full integrated) is 0.17, or about a 5.9 fold reduction in risk. As expected, the product of the two individual LR (2.6*2 = 5.2) is a reasonable approximation.

Q11. Do you use the same parameter sets for both the quadruple and integrated test? Eighteen participants responded. Twelve (67%) reported that they do use the same parameter set (the preferred methodology); five reported that they did not (28%) and one did not know (5%). Several of the 'outlying', or more extreme likelihood ratios, come from participants where the parameter sets are not matched. This is likely to be the cause of much of the variability in likelihood ratio comparisons involving quadruple/triple and integrated risks.

Table 3. Down syndrome integrated risks and associated likelihood ratios

Second Trimester Down syndrome risk (1:n)			Likelihood ratio for[1]		
Quad (Q) (FP-08)	**Full Integrated (FI)** (FP-08 & FT-08)	**Serum Integrated (SI)[2]** (FP-08 & FT-08)	**PAPP-A (Q/SI**	**NT+PA (Q/FI)**	**NT (SI/FI)**
43	1,570 Q	674	0.06	0.03	0.43
140	85 Q	55	2.55	1.65	0.65
140	460 Q	230	0.61	0.30	0.50
160	600 Q	330	0.48	0.27	0.55
224	27,067 Q	9,179	0.02	0.01	0.34
270	1,300 Q	650	0.42	0.21	0.50
270	340 Q	170	1.59	0.79	0.50
390	4,300 Q	1,500	0.26	0.09	0.35
440	415 Q	714	0.62	1.06	1.72
480	2,900 Q	1,400	0.34	0.17	0.48
689	3,600 Q	1,300	0.53	0.19	0.36
899	510 Q	220	4.09	1.76	0.43
1,800	22,000 T	5,900	0.31	0.08	0.27
2,000	20,000 T	8,400	0.24	0.10	0.42
2,400	29,000 Q	7,80	0.31	0.08	0.27
3,000	20,000 T	1,200	2.50	0.15	0.06
3,100	32,000 T	5,800	0.53	0.10	0.18
4,600	81,000 Q	29,000	0.16	0.06	0.36
460	**1,250**	**3,250**			
		outlier	0.02	none	1.72, 0.06
		Trimmed Mean LR	**0.53**	**0.17**	**0.39**
		log SD of LR	0.46	0.60	0.14
		Mean - 2sd	0.06	0.01	0.21
		low	0.06	0.01	0.18
		high	4.09	1.76	0.64
		Mean + 2sd	4.26	2.58	0.74

[1] Derived by dividing the associated Down syndrome risks
[2] Q indicates that the quadruple test is included with NT and PAPP-A; T indicates triple used instead

George J. Knight, Ph.D. Glenn E. Palomaki, B.S. (207) 894-6610
Jacob A. Canick, Ph.D. Geralyn M. Messerlian, Ph.D. (401) 453-7650

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