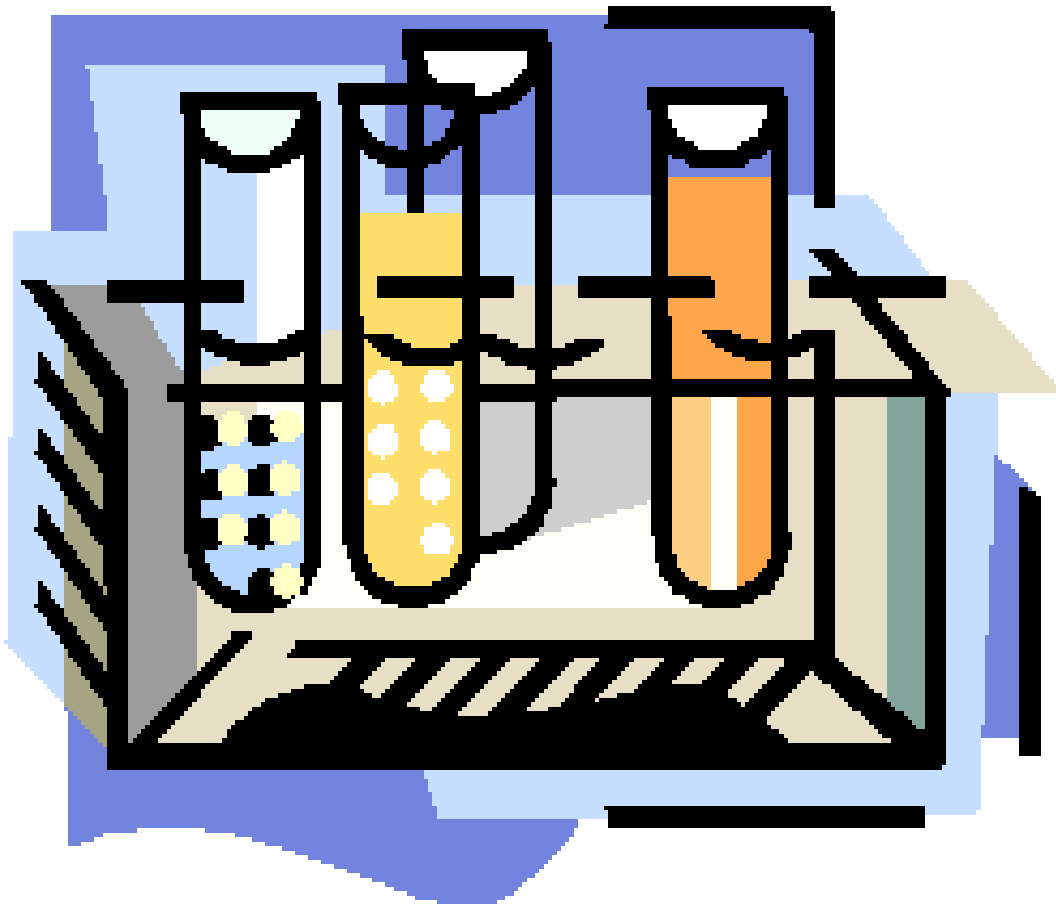


First Trimester Interlaboratory Comparison Program

Distribution 2006 FT-A



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

Discussion:

Data Listing and Analysis

Reading the Data Listing: The following five pages contain 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. In the future, we will implement more sophisticated trimming algorithms. **Comment:** The same two laboratories are having difficulties in reporting the correct maternal age at EDC as found in 2005 FP-C. These laboratories should examine their software/methodology and make corrections.

Specimen Creation: Four of the proficiency testing specimens in this second distribution of the First Trimester Interlaboratory Comparison Program (ICP) were pools of patient sera. FT-01 was an 11 week pool, FT-02 was a 13 week pool spiked with recombinant hCG, FT-03 was a 12 week pool and FT-04 was the same 13 week pool diluted 2:3 with normal human serum. Sample FT-05 was an artificial sample 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). Remember that the gestational ages assigned to each specimen do not necessarily correspond to the pool GA. They were varied to yield a wider range of MoM values. **Comment:** More samples in FT-B will be artificial. The aim is to convert to all artificial samples in 2007.

Variance of Mass Units and MoM Results: The PAPP-A analyte values for the four specimens that primarily consist of patient pools (FT 01, 02, 03, 04) yield CVs from 11 to 17%. Some of this variability may result from systematic differences in calibration between the DSL, DPC, and Delfia methods. The data are too few to allow calculation of separate statistics for each method for the purposes of comparison. Although the three PerkinElmer and three Diagnostic Products assay results show good agreement the MoM values for PAPP-A are more variable than the analyte values themselves (18-24%), suggesting that median values established by each laboratory may not be optimum, or are outdated because of lot-to-lot shifts in assay values. In contrast, the CVs for hCG values for the four patient pools are, on average, less (9-12%), and the CVs for the MoM values are also less (8-15%). The CVs for PAPP-A analyte and MoM values for the manufactured specimen (FT-05) are much higher than for patient pools (30% and 28%, respectively) although again the PerkinElmer and DPC assays are more consistent within themselves. It will require more data to determine if manufactured samples continue to show more between assay variability than patient pools. The CVs of the risk values (on a log scale) are relatively low (with the exception of FT-01 which is inflated because of two risks that just missed being trimmed as outliers), indicating good agreement between the risks generated between laboratories. Interpretations for the five samples were reasonably consistent: FT-01 and FT-04 (9 of 10 screen positive); FT-02, FT-03, FT-5 (8 of 10; 7 of 10; and 9 of 9 screen, negative, respectively). One laboratory did not provide a risk or interpretation for FT-05 (10.7 weeks), presumably because they do not screen at this time in gestation. It is not possible to make definitive statements about the 'correctness' of the

interpretations because laboratories employ different screening cut-offs and screening parameters,

Diluted Specimen: Since FT-03 and FT-04 were derived from the same 3 week pool, it is possible to compare matched results for each laboratory. The values from FT-04 divided by the result from FT-03 should be approximately 67%. Among the 14 laboratories reporting PAPP-A measurements, the average ratio was 0.71 (95% prediction limits 0.60 to 0.82), after removing one outlying observation (ratio of 0.88). Among the 11 laboratories reporting hCG measurements, the average ratio was 0.67 (95% prediction limits 0.58 to 0.75) with no outliers.

Interpretative Questions: Maternal Weight Adjustment of PAPP-A Levels

Does your laboratory provide clinical test results for Down syndrome screening? Of the 15 participating laboratories answering the interpretative question, 10 provide clinical test results. The subsequent analyses will be limited to these participants.

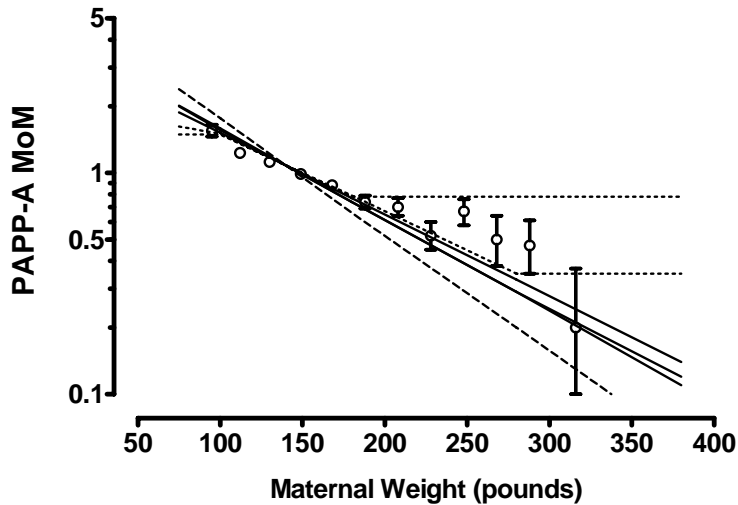
Do you routinely adjust PAPP-A measurement for maternal weight? All 10 participants responded that such adjustments are routine.

Sample FT-02 was from a 140 pound woman. Fill in the following table with weight-adjusted PAPP-A MoM levels for the provided maternal weights in pounds. Table entries ranged from a low of 75 pounds to a high of 380 pounds. This was done to determine if and what truncation limits are being applied. An analysis of these data will be provided later.

Provide the mathematical model that your laboratory uses for maternal weight adjustments for PAPP-A. Five participants reported using the log-linear model (log MoM vs. maternal weight). One additional laboratory reported using an 'other' method, but we confirmed that this laboratory also relied on the log-linear model. Among the remaining four laboratories, three used the reciprocal-linear model (MoM versus 1/maternal weight). The 10th laboratory reported using an 'other' method that was a log-log model. We were able to use the reported data from each laboratory to confirm the regression model used, with the exception of the laboratory using a log-log model.

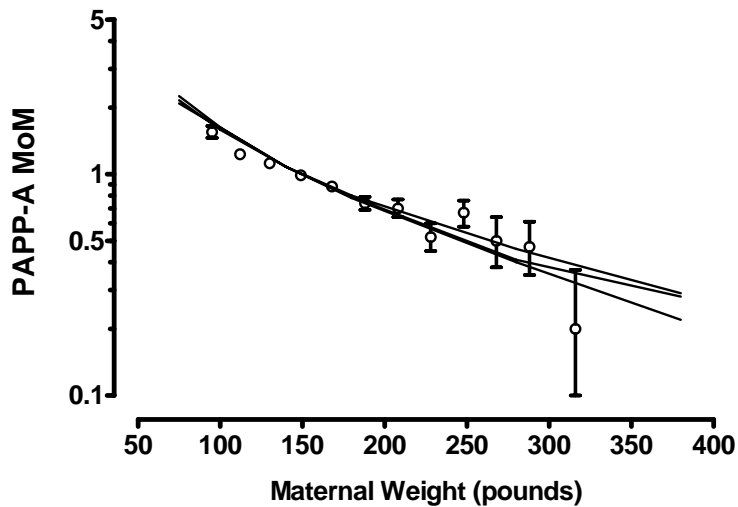
What is the source of the maternal weight adjustment equation for PAPP-A? Three of the five laboratories using the log-linear model, and two of the three using the reciprocal model, derived their equation using 'in-house' data. The remaining three participants used published parameters or did not know the source of the data.

Statistical analysis: In order to analyze the maternal weight adjustments, we made three assumptions. The average maternal weight was 150 pounds, the maternal weight adjusted value for each laboratory at 150 pounds was 7% higher than that reported for FP-02 (the clinical history reported a weight of 140 pounds), and that maternal weight/PAPP-A data from 5,000 patients collected from throughout the U.S. can be used to calculate an accurate representation of the relationship.



This first figure shows the relationship between maternal weight on the horizontal axis and PAPP-A MoM values on the vertical logarithmic axis. Open circles correspond to the paired PAPP-A/maternal weight values obtained in the 5000 patients, along with 95% confidence intervals (reference population). The maternal weight equation used by the six laboratories utilizing the log-linear methodology is shown by the six lines. The three solid lines indicate adjustment equations

that have no truncation limits, but fit the reference population data well between 100 and 200 pounds. The dashed line indicates a laboratory that fits the reference data less well and that also does not utilize truncation limits. The two dotted lines show the adjustment equations for the two remaining laboratories that use truncation limits. The truncation limits are indicated by the horizontal dotted lines. All of the maternal weight adjustments appear reasonable over the range where most of the data lie (between 100 and 200 pounds). However, there are important differences at the higher weights. Log-linear adjustments are known to over-estimate the effect at high weights, and this is why truncation limits have been set at between 250 and 280 pounds for this model.



This figure again compares the MoM/weight relationship for the reference population (open circles) with the maternal weight equations for the three laboratories utilizing the reciprocal-linear methodology. These laboratories fit the observed data more closely and over a wider range than the log-linear model making the use of truncation limits less important.

Comment: We suggest that all laboratories utilize the reciprocal-linear methodology. The log-linear model must ultimately break down at higher maternal weights because extending the equation of the line would yield unrealistic adjustment factors. The reciprocal-linear model is intuitively more consistent with expectation, i.e., the rate of change in the MoM adjustment should decrease as weight increases. In effect, maternal weight can be viewed as a diluent for the analyte in question, and can conceptually be thought of as a titer. If the log-linear model is employed it is essential to use truncation limits at both the upper and lower end of the maternal weight scale. Reasonable limits are 90 or 100 pounds at the lower end, and 250 or 280 pounds at the upper end.

Dimeric inhibin-A: DIA measurements were reported by three participants. All reported using the same method (DSL). The following table provides the reported DIA values and MoM levels for the three laboratories for all five samples.

Dimeric inhibin-A measurements for FT-A, 2006

Sample Number	Laboratory	Value	MoM	DS Risk (1:n)	DIA LR ¹
FT-01	A	335			
	B	227	1.01	18	0.56
	C	296	1.36	394	0.72
FT-02	A	272			
	B	267	1.00	460	0.41
	C	261	1.11	1080	0.33
FT-03	A	296			
	B	307	1.19	440	0.89
	C	265	0.97	1140	0.66
FT-04	A	185			
	B	195	0.75	15	-
	C	174	0.75	147	0.29
FT-05	A	390			
	B	442	1.23	8,200	0.83
	C	417	1.44	10,000	-

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

Supplemental Topic: Sequential and Contingent Screening

First trimester screening using the combined test (NT plus PAPP-A, and free beta hCG or hCG, with or without inhibin A) has the perceived advantage of an earlier diagnosis and, if warranted, earlier termination of an affected pregnancy (detection rate of 85% at a 5% false positive rate). The integrated test, where first and second trimester screening measurements are made (NT and PAPP-A in the first trimester and quadruple test results in the second trimester) but are only used to calculate a risk in the second trimester using all of the data, provides better screening performance (DR of 95% for 5% FPR). However, one issue with this strategy is that some women would have relatively high risks using the first trimester information, particularly if the NT measurement is very high, and would choose a diagnostic test based on that high risk. To address this issue, modifications of the basic integrated test have been proposed. Sequential and contingent screening protocols intervene on women with high risks in the first trimester to allow early diagnosis for women while referring some or all of lower risk women on to complete the full integrated test.

The following section provides an overview of these three protocols. Integrated screening holds the first trimester information until the second trimester results are also available. A single risk is then provided to the woman, and a single risk cut-off level is used to define screen positive results (e.g., $\geq 1:200$). Sequential screening initially offers counseling and diagnostic testing to all women with a high first trimester risk (using the results from the combined test) at or above an initial risk cut-off level (e.g., $\geq 1:50$), and refers the remaining lower risk women for a quadruple test (AFP, uE3, hCG and DIA) in the second trimester to allow an 'integrated' interpretation using information from both trimesters. Those with a Down syndrome risk above a final second trimester risk cut-off level (e.g., $\geq 1:270$) are also offered counseling and second trimester diagnostic testing. With the sequential protocol, the low risk women complete the integrated test. Contingent screening differs from sequential screening by having not only a high risk, but also a low risk, cut-off level defined in the first trimester (e.g., $\geq 1:50$ and $< 1:1500$). Women with Down syndrome risks lower than the lower risk cut-off level are informed that they do not require further testing, as they would be unlikely to become screen positive if they completed the integrated test.

Given that integrated screening uses all informative markers prior to assigning a risk and determining who should be offered diagnostic testing, the other two strategies will, of necessity, be less efficient, as defined by detection and false positive rates. This is because both sequential and contingent screening assign an interim risk and make the offer of diagnostic testing in the first trimester based on only a subset of informative markers. Thus, the early detection of some affected pregnancies and the reduced need for second trimester screening (contingent testing) must logically be 'paid for' by having somewhat less efficient screening.

Laboratories that perform first trimester screening have been (or will be) confronted with questions about sequential and contingent screening. Performance estimates for selected combinations of cut-off levels, as well as other issues in considering whether these protocols might be used in your laboratory setting, are presented in our recent publication (enclosed). The reference list in the publication contains additional publications that may also be of interest. Note that serum integrated screening (the integrated test without NT measurements) is usually not considered when discussing sequential and contingent testing as its performance in the first trimester is not sufficiently discriminatory.