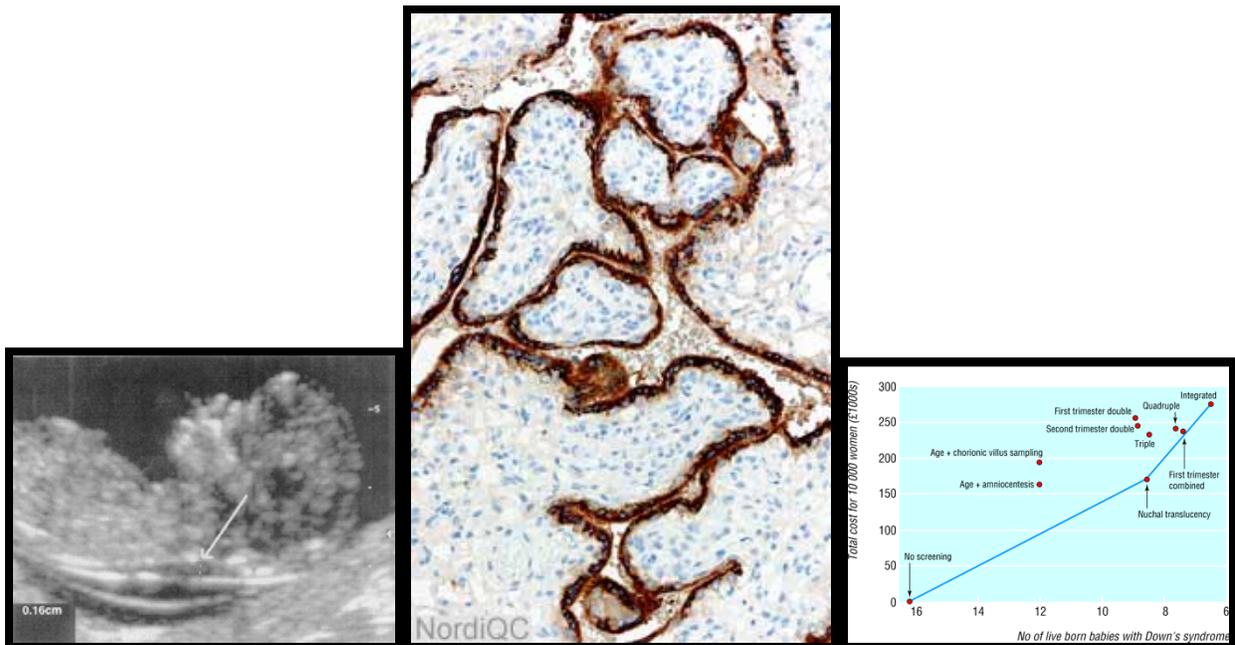


First Trimester Interlaboratory Comparison Program

Distribution 2005 FT-C



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

Participant Summary Report

First Trimester ICP: FT-2005C

Summary of the Laboratory Profiles

Methods: All 12 responding laboratories/manufacturers measured PAPP-A (8 DSL ELISA, 2 DPC Immulite 2000 and 2 PE DELFIA). There was more variability in what constituted the second analyte. Nine laboratories measured hCG (4 DPC Immulite 2000, 2 Beckman Access, and 1 each DPC Immulite 1000, Bayer ACS 180 and Bayer ADVIA Centaur). Three laboratories did not measure intact or total hCG.

Trimester of risk: Among the 10 laboratories reporting risk, seven reported first trimester risks, one reported second trimester risks, one reported term risks, and one reported not knowing the trimester of risk. **Comment:** This last laboratory should make an effort to determine the source (and trimester) of their risks, as their risks reported in Q7 did not closely match any other set of reported risks.

Source of NT medians: Four laboratories reported using a single set of medians (two reportedly based on internal data, and two based on Fetal Medicine Foundation data, such as Nicolaides et al., *Prenat Diagn* 1998;**14**:203-8). Two reported using center-specific medians. Three reported using sonographer-specific medians. Two additional laboratories reported using combinations of these. Two participants do not provide clinical results. **Comment:** One study (SURUSS, Wald et al *J Med Screen* 2003;**10**:56-104) has shown that sonographer-specific medians are better than center-specific ones. The FASTER study confirms this, and found that using center-specific medians results in an improvement over a single set of medians (Malone FD, et al, *Am J Obstet Gynecol*, 189:S232, 2003). Given this, laboratories should, at a minimum, have the capability of using center-specific medians.

Software: Five laboratories used LMS α .lpha, three used in-house software, and one each used Benetech PRA and Maciel Prenatal Interpretive Software. **Risk cut-off level:** Among the 10 laboratories reporting risk, we converted the results reported by those laboratories using second trimester and term risk cut-off levels to first trimester cut-offs using the conversion factors 0.77 (survival from 2nd trimester to term) and 0.55 (survival from 1st trimester to term). The first trimester risk cut-off levels ranged from 1:165 to 1:270, with a median of 1:220.

Maternal age-associated risks: The figure shows the age-associated Down syndrome risks reported by the eight laboratories that screen using first trimester risks. They used three sources [Five used Cuckle et al., *Br J Obstet Gynaecol* 1987;**94**:387-402 (solid lines); two used Hecht & Hook *Am J Med Genet* 1996;**62**:376-85 (dashed lines); and one used Morris et al., *Prenat Diagn* 2003;**23**:252-8, dotted line)]. Although difficult to see, all but three of the laboratories using the Cuckle equations agree closely on the age-associated risks.

Comment: Two labs appear to report second, rather than first, trimester risks (1:281 and 1:295 @ 35.5 years) and one reports term risks (1:390 @ 3.5 years).

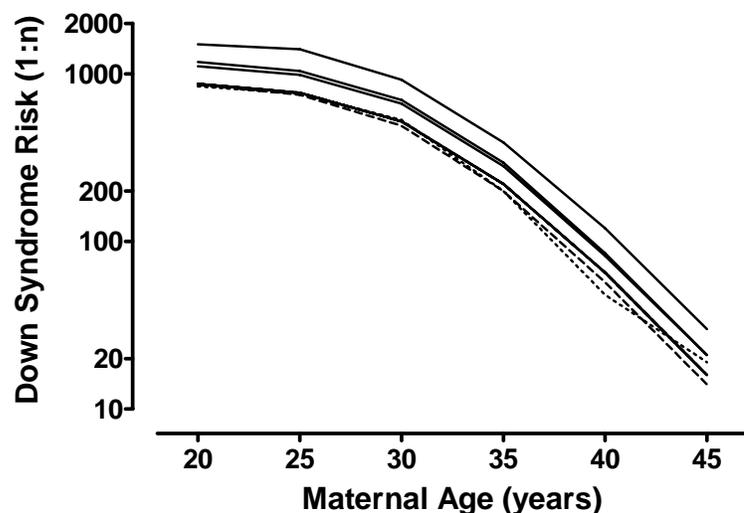


Figure 1

Other biochemical markers: Three clinical laboratories asked that dimeric inhibin-A (DIA) be included in the ICP. One each asked for the inclusion of ITA and free β . **Comment:** We will consider the inclusion of DIA in the next distribution.

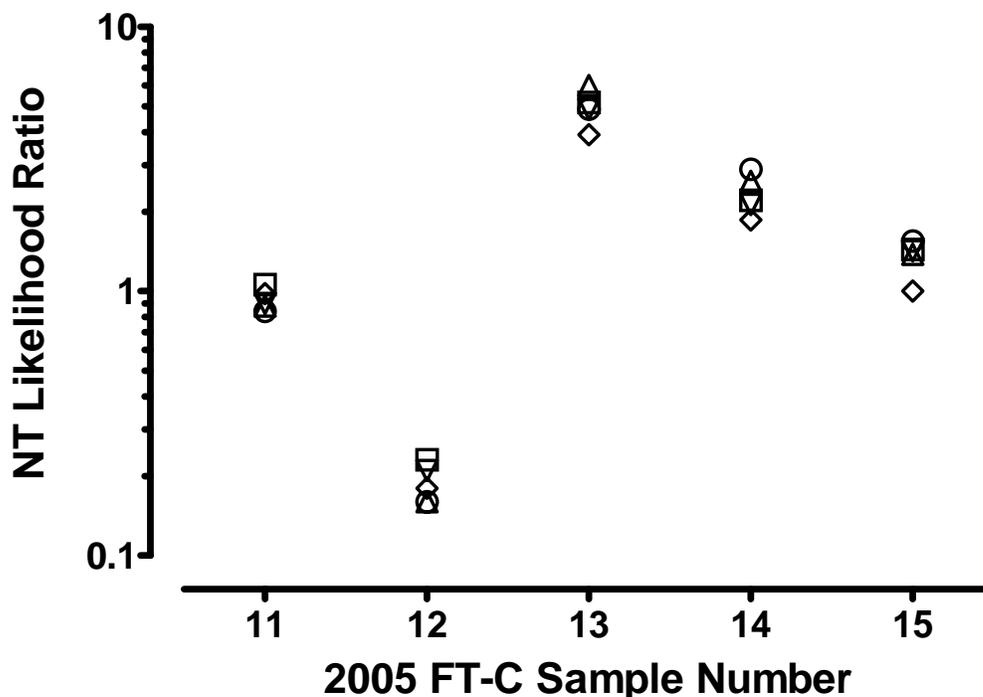
Interpretative Questions: Down Syndrome Risks with NT Results

Do you offer first trimester screening based only on biochemistry: Of the 10 participating clinical laboratories that report first trimester screening results, six do not routinely offer screening based on biochemistry only (NO). Four laboratories do make such screening available (YES). One of this latter group reports that 10% of testing is biochemistry alone, one reports <1% and two report they have the ability, but have not yet reported such results. **Comment:** Given the relatively poor performance of biochemistry alone in the first trimester, its routine use should be discouraged.

In the rare instance that NT is not available, do you provide a risk: Among the six laboratories not offering such testing, three (50%) say they are able to provide Down syndrome risks based on biochemistry alone. **Comment:** Laboratories should consider having the ability to provide a risk in situations where the sample has been collected, the NT was attempted, but never completed. However, this is likely to be a relatively rare event.

Down syndrome risk with NT excluded: Five laboratories were able to provide Down syndrome risks based only on biochemistry. Figure 2 shows the relationship between their initial Down syndrome risk (including NT) and the revised Down syndrome risk (excluding NT). This is done by calculating the NT likelihood ratio which can be computed by dividing the biochemistry only risk by the initial risk that included NT (e.g., $LR_{NT} = 1:500 / 1:100 = 5$). The NT measurement (in MoM) was reported in the clinical history for each sample. Examining the NT likelihood ratios for each laboratory allows an evaluation of how much variability is introduced into the final risk calculation by differences in the NT MoM-likelihood relationship. Each laboratory is represented by a different symbol. The likelihood ratios for each of the samples agree closely, indicating that relatively consistent methods and parameter sets were used for computation of the NT likelihood ratio. The median likelihood ratios for samples FT-11 to FT-15 are 0.90, 0.18, 4.89, 2.21, and 1.42, respectively (the corresponding NT MoMs are 1.41, 0.80, 1.84, 1.55 and 1.52, respectively). **Comment:** This exercise does not test the laboratory's ability to compute an NT MoM. This activity will be included in later distributions.

Figure 2 Comparison of NT Likelihood Ratios



Supplemental Questions: CRL and Gestational Age

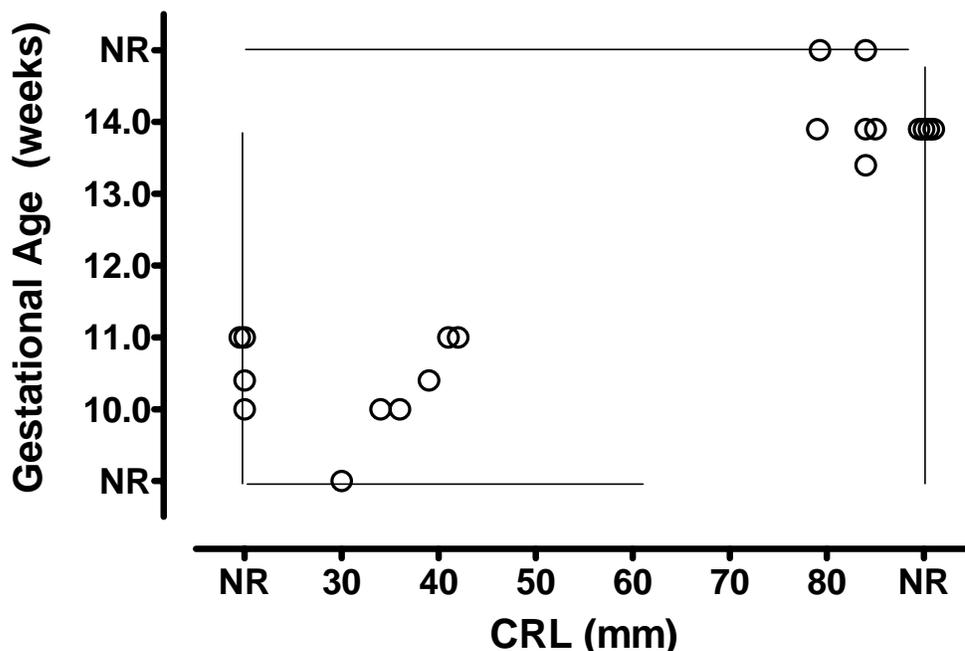
A total of 10 participating laboratories reported that they routinely provide first trimester screening results on a clinical basis. The responses from these laboratories form the basis of the following analyses.

Proportion of NT measurements with an accompanying CRL: Eight laboratories report that virtually all (100%) NT measurements have an accompanying CRL. The other two laboratories report 98% and 90%. **Comment:** It is assumed that if there is no CRL, the sonographer will have provided a gestational age.

Gestational age is defined by: Nine laboratories respond with the method(s) used to define the gestational age window for screening. Five use the CRL exclusively (including one laboratory that previously indicated that only 90% of NT measurements were accompanied with a CRL). One reports using the gestational age provided by the physician, and three report using a combination of CRL and gestational age.

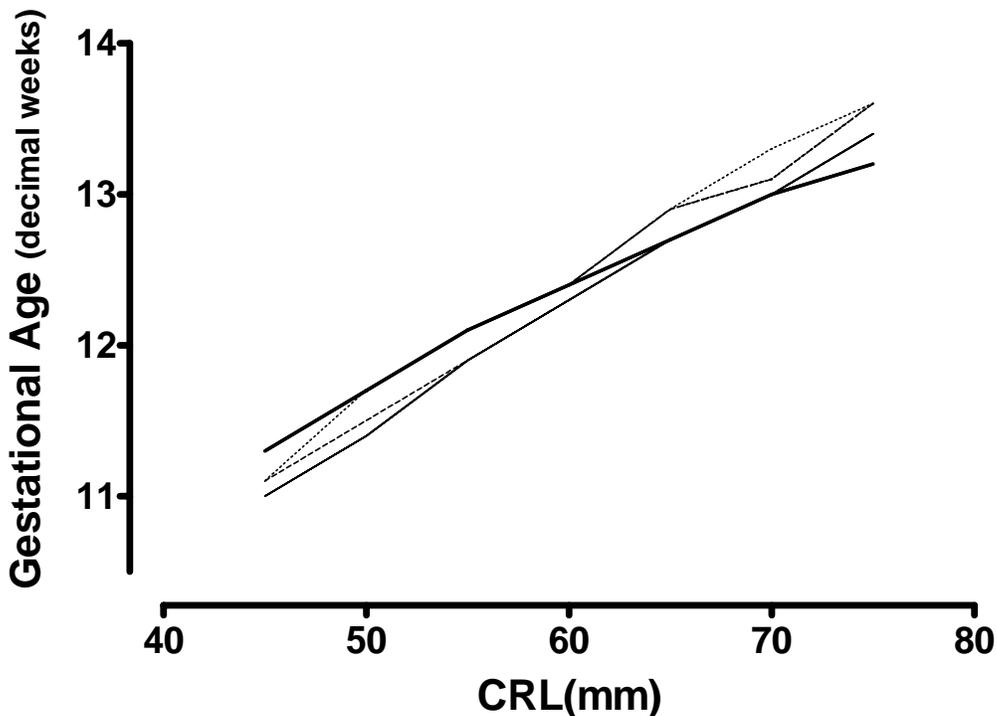
Earliest and latest CRL/gestational age for accepting an NT measurement: Figure 3 shows responses from the 10 laboratories. Some report both the CRL and gestational age defining the lower and upper limit of the screening window. Others report only a CRL cut-off or a gestational age cut-off (indicated in the figure by the circles having a line drawn through the measurement that was not reported – NR). All early cut-off levels fall between 30 and 40 mm and 10 to 11 weeks' gestation. For the lower limit of the screening window, there is a clear relationship between CRL and gestational age for those laboratories reporting both (5 data points, circles without lines). At the upper end, most laboratories use 13.9 weeks as the upper cut-off level, but there is more variability in determining what CRL corresponds to this gestational age (range 70 to 85 mm)[4 data points]. **Comment:** Sonographers experienced in NT measurements suggest the window to be from about 10.5 (or 11) to 13.9 weeks. They disagree on the CRL (in mm) defining this window. A lower limit of 11.0 weeks corresponds to CRL measurements of between 38 and 41 mm. The upper limit of 13.9 weeks corresponds to CRL measurements of between 79 and 84 mm, depending on the conversion equation used. Laboratories with limits outside of this range might want to review this decision in light of recent publications and recommendations.

Figure 3. CRL/Gestational Age Defining the First Trimester Screening Window



Conversion from CRL to gestational age: Figure 4 shows the relationship between CRL and gestational age reported by the 10 laboratories. Results that were reported in weeks+days have been converted to decimal weeks according to the table provided in the FT-C 2005 instructions. There are two interesting findings. All but one of the relationships (indicated by the thick solid line) have essentially the same slope. Two laboratories report using *Daya et al.*, (*Am J Obstet Gynecol* 1993;168:903-8) either alone, or with another reference. This source assigns one of the highest gestational ages (11.3 weeks) to 45 mm, along with one of the lowest gestational ages (13.2 weeks) to 75 mm, indicated by the solid thick line. The lower thin line is the data provided by the Fetal Medicine Foundation (*Nicolaides et al.*, *Prenat Diagn* 1998;14:203-8). Slightly higher, but parallel, lines appear to come from some other sources (*Robinson et al.*, *Br J Obstet Gynaecol* 1975;82:702-10) indicated by the dashed line and *Hadlock et al.*, (*Radiology* 1992;182:501-5) indicated by the dotted line. **Comment:** The Hadlock reference might be most widely used by sonographers, and we would recommend its use in screening laboratories. However, sonographers trained to perform NT measurements by the Fetal Medicine Foundation are more likely to use data from Nicolaides. In future educational enhancements, we will formally assess these issues and try to make firm recommendations from the laboratory viewpoint.

Figure 4 Reported Relationships Between CRL Measurements and Gestational Age



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 clinically active. Also, missing data in the 'total/intact hCG' column are from laboratories that did not report such a value. Results that are 'boxed' have been determined to be outliers. Outliers for gestational age or maternal age are outside of +/- 0.1 week (or year) 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 untrimmed standard deviations. In the future, we will implement more sophisticated trimming algorithms. **Comment:** At least two laboratories are having difficulties in reporting the correct maternal age at EDC. These laboratories should examine their software/methodology and make corrections. Our aim was to provide a decimal gestational age, and have each laboratory use that value in reporting. If laboratories rely on a CRL instead, they were to repeatedly 'guess' at a CRL that would provide the appropriate gestational age. Some laboratories may not have understood the directions. If questions about this persist, please contact us, using the number provided in the instructions.

Specimen Creation: The proficiency testing specimens in this first distribution of the First Trimester Interlaboratory Comparison Program (ICP) were pools of patient sera collected from women at gestational ages 11 (FT-11), 12 (FT-12 and FT-14), and 13 (FT-13) weeks. An admixture pool was made from a 50:50 mixture of the 11 week pool, and a separate 15 week pool to yield a high PAPP-A and lower hCG value (FT-15). The values for PAPP-A and hCG are expected to be close to the average value for any given gestational week, because approximately 20 sera were used for each pool (with the exception of the admixture pool FT-15). The range of values is thus not very wide. Note, however, that the gestational ages assigned to each specimen did not always correspond to the pool GA and were varied to yield a wider range of MoM values. **Comment:** During the next year, some of the samples will contain spikes, and others will be artificial. The aim is to convert to all artificial samples for 2007.

Variance of Mass Units and MoM Results: Both the PAPP-A analyte and MoM values yield CVs of 11 to 16%. 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. However, the fact that the MoM values are as variable as the analyte values themselves suggests that median values established by each laboratory may not be optimum. In contrast, the CVs for hCG values are on average less (7-13%), and the CVs for the MoM values are even less (3-9%). The reduction in variance for the MoM values suggests that some of the differences between values is systematic, either due to differences in calibration or between lab variability in values obtained using the same kit. The CVs of the risk values (on a log scale) are relatively low, indicating good agreement between the risks generated between laboratories. All of the five samples were called screen negative by the laboratories that provided an interpretation, with the exception of FT-13 which one lab called screen positive (risk of 1:200).

Duplicate Specimen: Given that FT-12 and FT-14 were duplicates of the 12 week pool, it was possible to compare matched results from each laboratory (Figure 5). This provides an estimate of the within-lab within-run CV (you might want to perform your own analyses as well, especially if several technicians ran the sample, or multiple tests were performed). The figure displays the PAPP-A (13 laboratories) and hCG values (10 laboratories) reported, along with mean values for the two samples and range of laboratory CVs for the paired duplicate samples.

PAPP-A results are slightly more variable than hCG, but overall the agreement between the two measurements for any given laboratory is very good. It is interesting that the hCG values appeared systematically higher, in FT-14, compared to FT-12 (8 of 10 laboratories reported higher levels) but this did not reach statistical significance. **Comment:** In future distributions duplicate specimens will be included in separate shipments to assess between-assay variability.

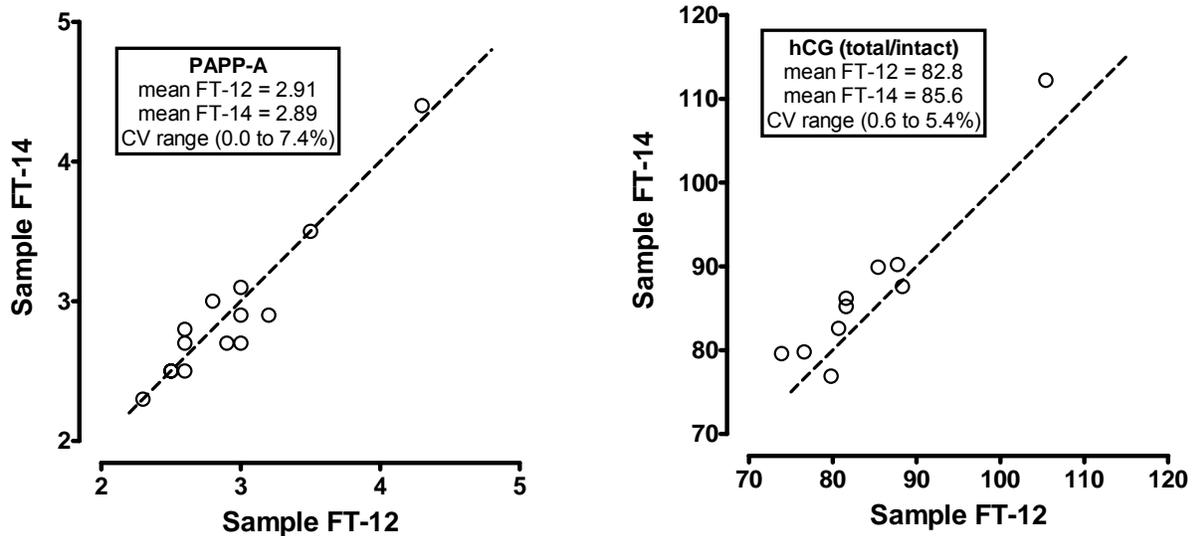


Figure 5. Comparison of PAPP-A and hCG Measurements in Duplicate Samples

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

Glenn E. Palomaki, B.S.
 GERALYN M. MESSERLIAN, Ph.D.

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

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

December 12, 2005