STATE OF NEW YORK
DEPARTMENT OF HEALTH

Wadsworth Center The Governor Nelson A. Rockefeller Empire State Plaza P.O. Box 509 Albany, New York 12201-0509

Richard F. Daines, M.D.
Commissioner

Wendy E. Saunders
Executive Deputy Commissioner

March 4, 2009

New York State Tumor Marker Proficiency Test 1/2009 Evaluation [1]

Dear Laboratory Director,

Attached is a summary and evaluation of the New York State Proficiency Test from January 27, 2009 for Tumor Markers AFP, CA125, CA15-3, CA27.29 and CA19-9, CEA, PSA, free PSA and complexed PSA.

Samples:

Laboratories were challenged with five (5) different coded specimens prepared by Wadsworth Center personnel. Purified analyte preparations were added in various amounts to a protein-based matrix, sterile filtered, aseptically dispensed into sample vials and stored at 4°C until mail-out. Analyte levels were pre-assayed and stability tested in our laboratory. All laboratories received the same samples, regardless of whether they tested for one or all of the analytes.

Result evaluation:

Your laboratory's results, scores and grades are printed on a separate page. Also included are the grades from the previous two PT events and your performance status. **Please review and sign your evaluation. Keep the signed result sheet in your files.** You will need it for your next laboratory survey to demonstrate successful participation in the NYS PT program.

For your information, we also included a tabular summary of all the results with high/low cut-off values (mean +/-3SD) for each analyte and a graphical comparison of the results obtained with the different assay methods/kits. In order to compare results between different kits more easily across all five samples, figures for CA125, CA15-3, CA19-9, CA27.29 and CEA were prepared from normalized values that were calculated by dividing the mean values for each method by the median of the means for all kits (all kit median) for each sample. The all kit median is used instead of the all lab mean to eliminate some of the bias toward a method used by a large number of labs. For AFP, PSA, free PSA and % free PSA, the figures show the ratio of the peer group means to the assigned target value (see below), instead of the all kit median. When comparing the results, please keep in mind that for some kits the number of results (i.e. N, the number of labs measuring a particular analyte with a specific kit) was small. However, the fact that the relative performance for almost all kits has been very constant over the last several years indicates that the results shown reflect the true behavior of each method compared to its peers, at least under the conditions of the NYS PT. Note that all means were calculated from results that fell within +/- 3SD of the corresponding mean after exclusion of outliers. The tabular summary and the figures include the results from kits used by at least two labs. Finally, we added a sixth group of bars labeled "average bias" to make it easier to compare the methods across all five samples. The straight lines above each bar represent the standard deviation.

[1] The use of brand and/or trade names in this report does not constitute an endorsement of the products on the part of the Wadsworth Center or the New York State Department of Health.

For grading purposes, all results for **AFP, CA125, CA15-3, CA19-9, CA27.29, CEA, PSA, PSA2, free PSA, % free PSA and complexed PSA** were evaluated based on their respective peer group mean. In order for you to more easily compare your results to those of your peer group, we calculated a D/Dmax value and displayed it directly under your individual results. D/Dmax is a measure of how much your result (x) deviates from your peer group mean, D/Dmax=(x-mean)/3SD, with D being the difference of your result from the mean, and Dmax being the maximal allowable deviation, i.e. 3SD. Thus, D/Dmax needs to be between –1 and +1 for a result to be scored. **<u>Note</u>: If your D/Dmax is <u>not</u> within +/- 0.66 (equivalent to 2SD), especially for more than one or two samples, you should carefully check your assay/result(s) since this indicates that your result(s) are significantly different from the mean(s) of your peer group.** While this could be an isolated incident, it could also potentially indicate that your assay may not be performing as well as it should. Furthermore, we have also added an average D/Dmax for each analyte to help you assess your results. If your **average D/Dmax is greater than +/- 0.5,** then this test exhibited a substantial high or low bias when compared to the rest of your method peer group. This suggests that there might be a potentially significant systematic error with your assay. Possible causes could include a calibration drift, reagents that are close to their expiration date, or subtle malfunction of your instrument. We strongly encourage you to take a close look at the run in question and others performed around that time and/or with the same reagent lots.

Discussion:

Results were reported by 120 labs using 10 methods to measure **<u>CA125</u>** (Fig.1); however, five methods were used by less than 10 labs each. The results from seven of the ten methods used to measure CA125 showed relatively consistent differences from a high of +6% (Siemens ADVIA Centaur, COB/BA1) to a low of -12% (Siemens Immulite 2500, DPF/DP5) relative to the medians, and reasonable consistency across all five samples. The exceptions were results from the Abbott Axsym and Architect (ABB & ABH/AB1) and the TOSOH A1A/ST-A1A (TOM/TO1) methods that were 28% and 57% higher, respectively, and, on the other side, from Fujirebio/Centocor (GAA/CE1) that were on average 40% lower than the medians. However, limited data from only two labs were available for the Fujirebio/Centocor method and these data were thus excluded from the all kit means and medians. In conclusion, consistent with previous results, there were substantial differences in how CA125 was measured.

The MUC1 breast cancer antigen was measured by 105 labs, with slightly more than half (53%) using one of six different **<u>CA15-3</u>** assays (Fig. 2), and the remainder using one of two different **<u>CA27.29</u>** assays (Fig. 3). Interestingly, there was a substantial difference in the performance of the different methods relative to each other between samples TM191-194 on the one hand and sample TM195 on the other. Whereas the three Siemens Immulite methods (grouped together for this analysis) gave results that were twice as high as those of the other methods in samples TM191-194, their result was much closer to the median in sample TM195. At the same time, results from the Beckman Unicel/Access instruments were much lower in sample TM195 compared to the median than in samples TM191-194. Thus, it appears that the Siemens Immulite and Beckman methods measured CA15-3 substantially lower in sample TM195 than in the other four samples. At this time we have no explanation for this observation. Furthermore, none of the other methods displayed a similar difference in how they measured CA15-3 in samples TM191-194 compared to sample TM195. Because of the large positive bias of the Siemens Immulite results, these were not included in the calculations for the all kit means and medians because of the impact that these large differences would have had on those values. It is noteworthy that a similar pattern of high results for the Siemens Immulite methods was also reported in the last two CAP surveys (TM-A 2008 & TM-B 2008), although the discrepancies were not quite as large. For CA27.29, results from only two methods were reported (Fig. 3). Those obtained with TOSOH A1A/ST-A1A (TOM/TO1) were between 6% and 30% (average 19%) higher than those from Siemens ADVIA-Centaur and ACS-180 (COB & COS/BA1). Furthermore, this difference was more pronounced at lower levels of CA27.29.

Results for **<u>CA19-9</u>** (Fig. 4) were reported by only 61 labs. Just under half (30) of these labs used Siemens ADVIA Centaur (COB/BA1), 13 labs (23%) used Beckman Unicel and Access (BCU & BCX/BC1, which were combined for this analysis because their results were similar), 7 labs (12%) used Roche Elecsys/Cobas

e411 and E170/Cobas e601 (BME & BMR/BM1), and 7 labs (12%) used the TOSOH A1A/ST-A1A (TOM/TO1) methods. Two of the methods, Beckman and Roche, gave CA19-9 results that were close and essentially identical to the medians, whereas results from TOSOH were lower than the medians overall by about 20%. In contrast, measurements of CA19-9 by Siemens ADVIA Centaur stood out and were 80-99% higher than those from the other methods. Consequently, the results from this method were not included in the calculations for the all kit means because of the impact that these large differences would have had on those values. However, the results from the Siemens ADVIA-Centaur method were included in the calculation of the all kit medians. For reasons that are unclear at this time, the TOSOH method, whose measurements of CA19-9 were lower by about 20% overall, showed a greater negative bias in their TM195 results compared to those for the other four samples while, on the opposite side, the ADVIA Centaur showed a greater positive bias in their results for this sample. Thus, there seem to be clear differences between how different methods measured CA19-9.

Results for **CEA** (Fig. 5) were reported by 176 labs using 11 different methods. Changes to the assay configuration and/or reagents for both the Beckman Access and Unicel, and the Siemens Centaur and ACS-180 instruments were introduced by those companies prior to the September 2008 PT and these required the participants to carefully choose their reagent codes, as both the old and new reagents were still in use. According to the results submitted for this PT event, there were only a few labs that still used the original Beckman CEA assay instead of the new enhanced CEA2, whereas a substantial number of labs each reported using either the original Siemens Centaur CEA antibody (COB/BA2) or the new CEA antibody (COB/BA3). Although results for the Siemens Centaur using either the original antibody or the new CEA antibody were similar, these reagent groups were relatively large and thus were kept separate for analysis. In contrast, the majority of labs reporting results for the Beckman Unicel and Access used the new enhanced CEA2 assay with only a few labs reporting to still using the original Beckman CEA assay. Thus, their results were grouped together with those from the new CEA2 assay according to their respective instruments because the results for both were similar. Overall, there was poor concordance between the different methods, with the results spanning an approximately two-fold range. The highest measurements were from the TOSOH A1A/ST-A1A (TOM/T01) and Ortho Clinical Vitros Eci (JJC/JJ1) methods ranging from 15 to 66% higher, whereas, on the other side, the two Roche methods, Elecsys/Cobas e411 and E170/Cobas e601 (BME & BMR/BM1) gave measurements that were between 19% and 28% lower than the medians. Consequently**,** the TOSOH and Vitros measurements were almost twice as high as those from Roche. With the exception of the Ortho Clinical Vitros Eci, measurements were relatively consistent across all five samples within each peer group. Overall, the CEA results from this PT, shown in Figure 5, as well as those from previous PTs, suggest that large differences exist among the methods used to measure CEA.

As in the last several PT events, target values were assigned using traceable International Standards for **AFP, free PSA and PSA**. Although results for AFP, PSA and free PSA were evaluated based on their respective peer group means for grading purposes, information on the performance of individual methods relative to the target values for these analytes is provided in the discussion below, as well as in the summary tables and graphs.

Absolute target values for AFP, PSA and free PSA were established based on the following International Standard preparations that were obtained from NIBSC (National Institute for Biological Standards and Control, A WHO International Laboratory for Biological Standards, Blanche Lane, South Mimms, Poters Bar, Hertfordshire EN6 3QG, UK, http://www.nibsc.ac.uk): PSA (free), 96/668, 1 µg per vial; PSA (90:10), 96/670, 1 µg per vial; and AFP, 72/225, 100,000 IU per vial with a conversion factor of 1.21 ng/IU. Each vial was resuspended as recommended by NIBSC, followed by serial dilution to obtain six different concentrations. Each dilution was measured twice in duplicate on a Beckman Access and a Roche Cobas e411 instrument (only AFP and total PSA on Cobas), and in collaboration with Siemens Diagnostics, on an ADVIA Centaur (only AFP and total PSA). Additionally, each dilution of PSA was measured on the Beckman Access using both the Hybritech standard calibration, as well as, the new WHO standard calibration. The raw data from each measurement were used to construct separate standard curves, which were then used to assign the respective analyte concentrations (assigned target values) to the TM191-195 samples that had been measured in the same run as the standards. Thus, two sets of target values were obtained from the Beckman Access for AFP and free

PSA and 4 sets for total PSA (2 sets for each different set of calibration standards), and from the Roche Cobas e411 instrument for AFP and total PSA only, as free PSA was not yet available. Additionally, one set of target values was obtained for AFP and total PSA from the Siemens/Bayer ADVIA Centaur. These were then averaged to obtain the target values for each sample and analyte. The respective target values with their standard deviations can be found in the summary tables.

Results for **AFP** (Fig. 6) were reported from 105 labs using 7 different methods. For this PT analysis, the Roche Elecsys/Cobas e411 and the E170/Cobas e601 methods were combined, as were the Siemens Immulite 1000, 2000 and 2500, because the results within those method groups were essentially the same. All results were evaluated according to traditional peer group statistics and received a passing score if they fell within the mean +/-3SD. In addition to the peer group statistics, the ratio of the group mean/target value is given for each sample to compare measurement biases between the different methods. Most results showed a small positive bias of 5 to 15% except for the Ortho Clinical Vitros Eci (JJC/JJ1), which gave results that were on average 9% lower than the target. Overall, the AFP results suggest that most methods are reasonably well standardized.

Results were reported by 263 labs using 12 different methods to measure total **PSA** (Fig. 7). The samples were prepared as mixtures of free and ACT-complexed PSA mixed in different proportions as indicated by the % free PSA in Figure 7. All results were evaluated according to traditional peer group statistics and received a passing score if they fell within the mean +/-3SD. In addition to the peer group statistics, the ratio of the group mean/target value is given for each sample to compare measurement biases between the different methods. The average bias for all methods in this PT was 12.3%. Expectedly, results for the lowest PSA sample in TM195 showed the greatest variability and, in particular, Ortho Clinical Vitro Eci showed a large positive bias in the measurement of this sample. As for the last PT, the results showed a more or less continuous gradient from +28% bias to 2% bias, without the clear grouping into "high" and "low" methods seen in some of the previous events. Sample pair TM193 and 194 were targeted for identical total PSA but different % PSA values. All methods measured the total PSA levels in this sample pair fairly equally, indicating that all methods are essentially equimolar, as was observed for similar sample pairs in the last PT event.

Eighty-two labs measured **free PSA** (Fig. 8) with the majority (39%) of the results reported with the Beckman Hybritech Access or Unicel (BCX or BCU/BC1) methods. All results were evaluated according to traditional peer group statistics and received a passing score if they fell within the mean +/-3SD. However, in addition to the peer group statistics, the ratio of the group mean/target value is given for each sample to compare measurement biases between the different methods. Not surprisingly, the very low levels of free PSA in sample TM195 resulted in the largest variability of the measurements. However, in practice such a sample would probably not have been assayed. Overall, the Beckman Unicel and Access and Siemens Dimension results were consistently higher than the target values, on average by 64%, 59% and 29%, respectively, whereas most of the results from the other methods were within 15% of the targets.

As in prior surveys, the figure for **% free PSA** (Fig. 9) is meant as a qualitative comparison only since there is a large number of method combinations used for its determination. The figure shows the method mean % free PSA/target % free PSA (generated from the ratios of the free PSA target values to the total PSA target values). As usual, since the % free PSA is derived from the ratio of free to total PSA, the differences in free PSA and total PSA measurements are reflected, or possibly even exaggerated, in the ratio. In general, the % free PSA calculated for the lowest PSA sample, TM195, showed a greater positive bias because the free PSA measurements for this sample by all methods showed a large positive bias. As could be expected, the higher free PSA values measured with the Beckman Hybritech Unicel and Access assays resulted in % free PSA values that were between 31-57% higher than the targets. Measurements of the free and total PSA by Abbott Axsym, Imx, and Architect and Roche Elecsys and E170 were close to target for TM191-194, while those by Siemens Dimension were about equally above target. Consequently, the % free PSA calculated for these three methods was close to target. In contrast, the lower than target measurement of free PSA coupled with the higher than target measurement of total PSA by Siemens Immulite resulted in a lower than target % free PSA for that method combination.

Note: Several labs measured free PSA even though the total PSA was outside the range for measuring free PSA given by the lab. This appears a violation of these labs' respective policies, and indicates that they did not treat the PT specimen exactly like a patient sample. Labs are expected to calculate the % free PSA if they perform the free and total PSA assays and would do so for a similar patient sample. However, if a lab's policy is not to measure and calculate %free PSA outside a certain range of total PSA, then this rule should also be applied to the PT samples. In that case, please indicate this on the result sheet, so we know that the failure to provide a result was deliberate, or the absence of the % free PSA calculation without an acceptable explanation for its omission will be counted as a failure. Furthermore, some labs did not follow their policy and calculated % free PSA when they didn't need to. This is also against their lab policy even though there would presumably not be any negative consequence for the extra calculation. Please note that results must be given as percent free PSA, and not as a fraction.

Only 12 labs measured **complexed PSA**, and all of these used the Siemens ADVIA Centaur method, with good agreement between the labs. Furthermore, the mean % complexed PSA calculated from these values of 87.3% for TM191, 193 and 195 compared well with the mean of 10.3% free PSA for these three samples and, that of 76.7% for TM192 and 194 compared well with 21.3% for these two samples.

Cut-off values: As explained previously, the result we intended to get for cut-off values was the upper limit of your normal or reference range for each analyte, above which you (or your computer) would flag a result as elevated or abnormal. We also asked you to classify each result as either normal, i.e. within the normal or reference range, or abnormal or elevated, i.e. above or outside of the reference range. We will continue to ask for this and expect it to be filled in on the result form. As recommended in the instructions included with the samples, where there is a range of reference values (for example, age-specific reference values, or smokers versus non-smokers), please enter that information in the comments on the form. Also, if there are two or more reference values, e.g. smoking versus non-smoking populations, please use the non-smoking reference for your normal versus abnormal evaluation, but enter a note in the comment section that there are two or more reference values and list the other values if possible.

In conclusion, there can be significant differences between results obtained with various methods, especially for CA125, CA15-3, CA19-9 and CEA, as observed previously. While some of these may be due to the artificial nature of the PT samples, others are probably due to inherent differences in the assays themselves. We will continue to try to minimize the differences that can be attributed to the sample composition. Nevertheless, despite the admittedly somewhat artificial nature of the PT samples, we would like to suggest that the differences between results obtained by various methods might also be reflected in patient serum samples. Therefore, caution needs to be used when comparing the results from the same patient obtained with different methods, since clearly not all methods are equal. For this reason, we require that the method used must be clearly indicated on the patient report (Oncology Standard OC 1b). We would also like to encourage you to educate your physician clients about this potential problem. Furthermore, the comparison of method means to target values set by traceable International Standards for PSA and free PSA clearly shows that not all methods are calibrated equally, as discussed in the respective paragraphs.

Finally, we would like to raise the usual cautionary notes when interpreting these results which are 1) since some of the assays were done by a small number of labs, the results might be skewed due to a lack of statistical power; 2) it is difficult to make an accurate comparison of results when the % CVs are large; and 3) the analyses for PT purposes are done with artificially prepared mixtures of proteins which may or may not accurately reflect patient derived samples.

Important Reminder regarding the data submission process: Be sure your results are submitted. If results are saved but **not submitted**, they will be graded as an administrative **fail**.

Note: Please be aware that in each subsequent event, fields will be pre-populated based on what you entered this time or a previous time. **Therefore, make sure that the selected instruments and reagents are correct, whether this is pre-populated from the last event or newly entered information.** This is important and in your interest since we need this information to properly evaluate your results and compare them to those of your peers. **You are at risk** of receiving a technical failure for results evaluated outside of the correct peer

group or an administrative failure for incorrect methodology. **No changes can be made for incorrect or missing information once the submission deadline has passed**.

We would like to comment again on some difficulties that were encountered with electronic submission of the PT results. Some required fields that continued to cause problems were those for the range of total PSA for measuring free PSA and calculating the free/total PSA ratio. Values for a quantitative range or text, such as "all levels", "NA" (N/A with a slash is not accepted), "not applicable" or "see comments" could be entered here. If the test was performed, then something had to be entered in the range field to go forward to the results page. One cautionary note: please be sure to apply the stated ranges to all of your PT samples, as a failure to apply the range correctly to all can result in sample failure.

Additionally, the information regarding the PSA2 line in the event menu still applies. The PSA2 line was added to allow entry of results from a second PSA assay **only** for those labs that use a different method for total PSA for the determination of the free/total PSA ratio. If only one PSA test was done, then these results should have been entered in the first PSA line. Most labs should have selected "test not performed" for PSA2 since only a few actually do perform a second assay. For labs that entered two PSA tests, the primary PSA test should have been entered on the first PSA line and the secondary assay for determination of the free/total PSA ratio on the PSA2 line.

Finally, on the results pages, the absence of data in the required fields for upper limit of the normal reference range and sample interpretation led to problems. Furthermore, some labs appear to be confusing the limits of the normal reference range for the test interpretation with the assay's lower or upper limits of detection.

Please note that questions regarding the electronic proficiency testing reporting system (EPTRS) account application process and the entry and submission of proficiency test results can be directed to clepeptrs@health.state.ny.us, or directly to Kathi Wagner at (518) 402-4266 or by e-mail at klw05@health.state.ny.us.

If you have any questions or wish to discuss some of the issues alluded to in the PT discussion, you may contact us at the address below. Also, this discussion with the tables and figures (in color) will eventually be posted on our website at http://www.wadsworth.org/labcert/clep/PT/oncology/index.htm.

For your information, the tentative schedule for the 2009 Tumor Marker Proficiency Test mail-outs is:

Mail-out date:	**Due date:**
May 5, 2009	May 20, 2009
September 15, 2009	September 30, 2009

Erasmus Schneider, Ph.D.
Director, Oncology Section
Clinical Laboratory Evaluation Program
Wadsworth Center
Empire State Plaza
Albany, NY 12201-0509
Ph: (518) 474-2088
FAX: (518) 474-1850
email: schneid@wadsworth.org