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

James W. Clyne, Jr.
Executive Deputy Commissioner

March 4, 2010

New York State Tumor Marker Proficiency Test 1/2010 Evaluation [1]

Dear Laboratory Director,

Attached is a summary and evaluation of the New York State Proficiency Test from January 26, 2010 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, together with the grades from the previous two PT events and your performance status. Starting with the present evaluation, only the laboratory's individual result and score report will be mailed, whereas the overall evaluation with the summary tables and graphs will be sent electronically and will also be posted on our website at http://www.wadsworth.org/labcert/clep/PT/oncology/index.htm. **Please review and sign your score report and keep it in your files.** You will need it for your next laboratory survey to demonstrate successful participation in the NYS PT program.

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. **Note: If your D/Dmax is not 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

[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.

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.

For your information, we 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 generally 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. Note: the numbers in parentheses behind each legend represent the number of labs that used that particular method.

Discussion:

Results were reported by 117 labs using 10 methods to measure **CA125** (Fig.1); however, only four methods were used by 10 or more labs each. Two methods including TOSOH ST-A1A (TOM/TO1) and Abbott Architect (ABH/AB1 used by only two labs) gave results that were on average 24% and 26% higher than the medians, respectively. Two other methods, including Siemens Immulite 2500 (DPF/DP5 used by four labs) and Roche Elecsys/Cobas e411 and E170/e601 (BME & BMR/BM1) gave results that were on average 21% and 23% lower than the medians, respectively. In contrast, the results from the other six of the ten methods used to measure CA125 all were within +/-10% of the medians, and were reasonably consistent across all five samples. In total, 76% of the labs agreed reasonably well on how CA125 was measured in these samples with no more than 20% difference between each method, whereas higher results were obtained from two methods used by 7% of labs and lower results from two methods used by 17% of the labs.

Results for **CA19-9** (Fig. 2) were reported by only 57 labs. Over half (30) of the labs used Siemens ADVIA-Centaur (COB/BA1), 11 labs (19%) used Beckman Unicel or Access/2 (BCU & BCX/BC1), 8 labs (14%) used Roche Elecsys/Cobas e411 or E170/Cobas e601 (BME & BMR/BM1), and 6 labs (11%) used the TOSOH ST-A1A (TOM/TO1) method. Only two of the methods, Beckman and Roche, gave CA19-9 results that were close to each other and represent the medians, whereas those from TOSOH were lower than the medians by about 33%. In contrast, measurements of CA19-9 by Siemens ADVIA-Centaur were more than twice as high as those from the other methods. Consequently, the results from this method were not included in the calculations for the all lab and kit means as specified in the summary tables 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. In addition, the graph shows that the new Abbott Architect (ABH/AB1 used by only 1 lab) method gave measurements for CA19-9 that were very high, ranging from 4 to almost 8 times higher than the all kit medians and, thus, they were excluded from the all lab and kit mean and median calculations. These high measurements are consistent with previous CA19-9 results by this method that were also at least four-fold higher than the medians. Thus, as Figure 2 shows, there seem to be clear differences between the various methods used to measure CA19-9, resulting in at least three-fold differences from lowest to highest for the majority of methods. Additionally, the extremely high results by the Abbott Architect method stood out further from all the others.

The MUC1 breast cancer antigen was measured by 96 labs, with slightly more than half (55%) using one of seven different **CA15-3** assays (Fig. 3), and the remainder using one of two different **CA27.29** assays (Fig. 4).

As Figure 3 shows, the results from the different methods used to measure CA15-3 spanned a two-fold range. The highest three methods, Siemens Immulite 2000/2500 (DPD & DPF/DP5), Abbott Architect (ABH/AB1), and Siemens ADVIA-Centaur (COB/BA1), all measured CA15-3 in these PT samples from 20% to 30% higher than the medians, while the lowest, Beckman Unicel and Access/2 (BCU & BCX/BC1) measured on average 35% lower than the medians. The remaining three methods, which include Abbott Axsym (ABB/AB1), Ortho Clinical Diagnostics Vitros ECi/Q (JJC/JJ1), and the Roche Elecsys/Cobas e411 and E170/e601 (BME & BMR/BM1) were in between those extremes and within 14% of the medians on average. For CA27.29, results from only two methods were reported (Fig. 4) and they were within 10% of each other. However, the values were 18% lower on average than those from the CA15-3 assays.

Results for **CEA** (Fig. 5) were reported by 172 labs using 11 different methods of which four were used by less than ten labs each. Seven methods gave results for CEA that were relatively consistent and on average were within +/-10% of the medians. In contrast, the Ortho Clinical Diagnostics Vitros ECi/Q & 5600 (JJC & JJF/JJ1) and the TOSOH ST-A1A (TOM/TO1) methods gave results that were on average 20-39% higher, while the Roche Elecsys/ Cobas e411 and E170/Cobas e601 (BME & BMR/BM1) results were on average 20% lower than the medians. These methods, used by 25% of the labs, appear to exhibit clear differences in how they measure CEA.

Please note that the method groups in the summary table reflect the fact that Beckman Unicel and Access/2 CEA tests are now all converted to the enhanced CEA2 test and that all Siemens ADVIA-Centaur CEA tests use the new CEA antibody. The two special Beckman CEA reagent listings that were in place for the original CEA and the enhanced CEA2 tests were removed from the reagent list because of this. Similarly, there is only one antibody available for the Siemens ADVIA Centaur CEA test and there are no longer two separate reagent listings to indicate the original antibody and the new CEA antibody. Thus, the general Beckman Unicel Access/2 reagent and the general Siemens ADVIA Centaur reagent should be selected for either the Beckman or the Siemens ADVIA Centaur CEA test, respectively.

For **AFP, free PSA and PSA,** target values were assigned using traceable International Standards. Although results for AFP, PSA and free PSA were evaluated based on their respective peer group means for grading purposes, the performance of the individual methods was compared to the target values established through the use of the international standards as described below.

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 once in duplicate plus once in singlet on a Beckman Access, once in duplicate on the Roche Cobas e411 instrument, and, in collaboration with Siemens Diagnostics, on an ADVIA-Centaur (measurement for AFP and total PSA only). Additionally, each dilution of total PSA and free 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 TM206-210 samples that had been measured in the same runs as the standards. Thus, two sets of target values were obtained from the Beckman Access for AFP and 4 sets for total and free PSA (2 sets for each different set of calibration standards), and one set from the Roche Cobas e411 instrument for AFP, total PSA and free PSA. Additionally, one set of target values was obtained for AFP and total PSA from the Siemens 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 102 labs using 8 different methods. Three of these methods were used by less than 10 labs each. 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, and the Beckman Unicel and

Access, 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 and/or calibration biases between the different methods. With the exception of the Vitros ECi/Q (JJC/JJ1) method that gave results that on average were 21% lower than the targets, results from the remaining methods varied from +6% by Siemens Dimension VISTA (DUV/DA2) to -7% by Siemens Immulite 1000/2000/2500 (DPB, DPD & DPF/DP5). Overall, however, the AFP results were quite consistent across all five samples and suggest that most methods are well standardized. Thus, as the summary table shows, the AFP all kit means and medians are very close to the target values for these samples.

Results were reported by 259 labs using 13 different methods to measure total **PSA** of which all but two were used by at least ten labs (Fig. 7). The samples were prepared as mixtures of 10% free and 90% ACT-complexed PSA 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 and/or calibration biases between the different methods. The average bias for all methods in this PT was 18.7% which is similar to that observed in the previous PT. However, this time the methods seemed to fall into either a high or a low group, which had an average bias of +27.8 +/-6.6% and +11.0 +/- 1.7%, respectively, $p<0.002$ for the difference. Presumably, the methods in the low group are all calibrated against the international WHO standard, whereas those in the high group presumably are calibrated against the older Hybritech standard. For the Beckman Unicel or Access/2 assays, which are available with either standard, the difference between the results based on the two calibration standards was about 20%, in accordance with the technical bulletin from Beckman on the PSA assay ("Access Hybritech PSA Hybritech and WHO Calibration Information #A59476A, 2008) that reports a difference of 22% between results generated from the original Hybritech versus the WHO calibration. Clearly, it is important that you and your physician clients are aware of this difference and what calibration standard is used.

Eighty-two labs measured **free PSA** (Fig. 8) with the most (38%) results reported with the Beckman Hybritech Access/2 or Unicel (BCX or BCU/BC1) methods (2 of those were calibrated with the WHO standards). 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 and/or calibration biases between the different methods. As seen in Figure 8, there was a clear distinction in results obtained with the two Beckman instruments calibrated with Hybritech calibrators and, to a lesser extent, with the Siemens Dimension, and the rest of the methods. However, as seen in figure 8, almost all methods showed a more or less positive bias compared to the target value, which ranged from a maximum of +88% down to -5%. Also, as the graph clearly shows, the WHO calibrated Beckman methods gave results that were substantially lower than those from the original Hybritech calibrated Beckman methods and that were comparable to the results from the other methods.

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. 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 22-47% higher than the targets. In contrast, measurements of the free and total PSA by Abbott Axsym and Architect and Roche Elecsys/Cobas e411 and E170/Cobas e601 were relatively close to the targets, and consequently the % free PSA calculated for these three methods was also close to the targets. For Siemens Immulite, the comparatively lower free PSA values combined with the comparatively higher PSA values resulted in % free values that on average were lower than the target by 20-24%.

Only 11 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 90.0% for TM 206-210 compared well with the mean of 10.3% free PSA for these samples.

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. There are still too many instances where individuals have either inadvertently selected a qualifier (< or >) or an incorrect instrument/reagent pair when scrolling through the instrument or reagent lists and this has resulted in failed samples or tests. **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. A quantitative range such as “4-10” 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 in 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.

For your information, the schedule for the remaining 2010 Tumor Marker Proficiency Test mail-outs follows:

Mail-out date:	**Due date:**
May 11, 2010	May 26, 2010
September 14, 2010	September 29, 2010

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