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

June 14, 2010

New York State Tumor Marker Proficiency Test 5/2010 Evaluation [1]

Dear Laboratory Director,

Attached is a summary and evaluation of the New York State Proficiency Test from May 11, 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. As for the previous 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 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

[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.

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.

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 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 ten methods to measure **CA125** (Fig.1); however, only half of the methods were used by ten or more labs each. Sixty-four percent of the labs used five of the methods, and these five gave results that were similar and within +/- 6% of the medians. Another three methods were on average 15% higher (Beckman Unicel and Access) or lower (Roche Elecsys, Cobas and E170, and Siemens Immulite 2500), respectively. Thus, in total, 93% of the labs using eight of the ten methods agreed reasonably well on how CA125 was measured in these samples with no more than +/-15% deviation from the median. In contrast, two methods, Abbott Architect (two labs) and TOSOH ST-A1A (five labs), which together were used by only 7% of the labs, gave results that were on average 33% and 39% higher than the medians, respectively.

Results for **CA19-9** (Fig. 2) were reported by only 59 labs. Over half of the labs (32 or 54%) used Siemens ADVIA-Centaur, 12 labs (20%) used Beckman Unicel or Access/2, 7 labs (12%) used Roche Elecsys/Cobas e411 or E170/Cobas e601, and 6 labs (10%) used the TOSOH ST-A1A method. Only two of the methods, Beckman and Roche, gave CA19-9 results that were close to each other and represent the medians. In contrast, measurements of CA19-9 by TOSOH ST-A1A were lower than the medians by about 31%, and, on the opposite side, those by Siemens ADVIA-Centaur were more than twice as high as those from the other methods. As a consequence, the all lab means (excluding Abbott Architect) for the CA19-9 measurements are substantially higher than the medians, reflecting the higher measurements from the large ADVIA Centaur group. Finally, as the graph shows, the Abbott Architect method (used by only 1 lab) gave measurements for CA19-9 that were between 6 and 7 times higher than the all kit medians, and more than ten times higher than the lowest results obtained with the TOSOH ST-A1A. These high measurements by the Abbott Architect are consistent with previous CA19-9 results by this method, which were also at least four-fold higher than the medians. Thus, as Figure 2 shows, there seems to be little agreement between the various methods used to measure CA19-9.

The MUC1 breast cancer antigen was measured by 100 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). Four of the methods for CA15-3 were used by less than ten labs each, and these together

accounted for 9% of the labs. Five of the methods used to measure CA15-3, accounting for 56% of the results, gave measurements that were within +/-15% of the medians. While there was only a small difference (<10%) between the two methods used for CA27.29, the differences between the methods used for CA15-3 were substantial. In particular, the Siemens ADVIA-Centaur method used by 34% of the labs gave results 36% higher than the medians, whereas the Beckman Access and Unicel methods used by 9% of the labs gave results that were 36% lower than the medians. Consequently, as Figure 3 shows, the results from the different methods used to measure CA15-3 spanned a two-fold range. Finally, the median values from the CA27.29 measurements were on average 16% lower than those from the CA15-3 assays.

Results for **CEA** (Fig. 5) were reported by 169 labs using eleven different methods, although four of these were used by less than ten labs each and together accounted for 11% of the total labs. The seven methods used by the majority (75%) of labs gave results that were relatively consistent and on average were within +/-6% of the medians. In contrast, the Ortho Clinical Diagnostics Vitros ECi/Q & 5600 and the TOSOH ST-A1A methods gave results that were on average 29% and 39% higher, respectively, while, on the opposite side, the Roche E170/Cobas e601 and Elecsys/Cobas e411 results were on average 17% and 19% lower, respectively, than the medians. These four methods were used by 25% of the labs. For reasons that are unknown, there was a substantially higher variability among the CEA measurements from the Ortho Clinical Vitros Eci/Q & 5600 method used by 11 labs, which resulted in high %CVs ranging from 14.3% to 20%. This is in contrast to the relatively good intra-method correlation for the other ten methods with %CVs ranging from 0.8% to 8.8%.

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 twice in duplicate for AFP and free PSA and once in duplicate plus once in singlet for PSA on a Beckman Access, 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 TM211-215 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). 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 eight different methods. Half of these methods were used by less than ten labs each, together accounting for 17% of the total labs. 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. Overall, the AFP results were quite consistent across all five samples, suggesting that most methods are well standardized. The exception was

the Ortho Clinical Vitros ECi/Q method, used by only 3 labs (3% of all), whose results on average were 24% lower than the targets. 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 256 labs using 17 different methods to measure total **PSA,** although seven of these were used by less than ten labs, together accounting for 9% of the total results (Fig. 7). The samples were prepared as mixtures of 10% free and 90% ACT-complexed PSA. 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. Expectedly, measurements for the lowest PSA sample, TM211, showed the greatest variability among methods. The average bias for all methods in this PT was 18.2% which is similar to that observed in the previous PT. However, in contrast to previous PT events, there didn't seem to be the clear separation of results into a high and low group. Rather, there is a gradient from the highest method (Siemens Dimension) at +40% to the lowest method (Beckman Unicel/Access with WHO calibration) at exactly the target value. For the Beckman Unicel or Access/2 assays, which are available with either calibration, the difference between the results based on the two calibration standards was about 27%. At this time, it is unclear why the previously observed separation into the high and low group did not occur.

Eighty-two labs measured **free PSA** (Fig. 8) with most results (40%) reported with the Beckman Hybritech Access/2 or Unicel 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 the Siemens Dimension, and the rest of the methods. Additionally, measurements by the Siemen's Dimension for the lowest level of free PSA in TM 211 showed an especially large positive bias (more than twice the target value for this sample). In contrast, as seen in figure 8, with one exception, all the other methods only showed a relatively small positive bias compared to the target value, from +4 to +14%. The exception is Siemens Immulite 1000/2000, whose results were approximately 10% lower than the target. As expected, the results from the WHO-calibrated Beckman methods were substantially lower than those from the original Hybritech-calibrated Beckman methods and these were comparable to the results from the other methods.

We would like to make some comments about the changes made for this PT event, wherein labs were required to measure free PSA for all samples if they test for free PSA, but they were not required to calculate % free PSA, whereas, previously, labs were asked to measure and calculate free PSA/%free PSA in samples according to their lab policy. We understand that this may in some cases be a deviation from a lab's policy in dealing with free PSA and mean that PT samples are not exactly treated like patient samples. However, a key reason behind these changes is that a lab's ability to accurately measure free PSA is an essential process for a testing laboratory, while calculating % free PSA is a secondary operation usually done by a computer. In addition, some labs do not calculate % free PSA at all, but only report free and total PSA values and leave the calculation to the physician. Consequently, the variability in the labs' policies coupled with the variable levels of PSA and free PSA in the PT samples has often led to difficulty in analyzing results and evaluating an individual lab's testing performance for free PSA. Thus, the survey changes were made to improve our ability to evaluate a lab's performance for measuring free PSA as an essential part of laboratory testing. The question regarding whether a range is used in normal laboratory procedures to decide whether free PSA is measured and/or calculated was included in this PT event for information purposes only.

Finally, only 11 labs measured **complexed PSA**, and all of these used the Siemens ADVIA-Centaur method, with good agreement between the labs.

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**.

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 to be used in conjunction with free PSA measurements. 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 use in conjunction with their free PSA results 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 final 2010 Tumor Marker Proficiency Test mail-out follows:

Mail-out date:	**Due date:**
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