

October 26, 2009

**New York State Tumor Marker Proficiency Test 9/2009 Evaluation <sup>1</sup>**

Dear Laboratory Director,

Attached is a summary and evaluation of the New York State Proficiency Test from September 15, 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.

---

<sup>1</sup> 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,  $D/D_{max}=(x-\text{mean})/3SD$ , 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 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 11 methods to measure **CA125** (Fig.1); however, four methods were used by no more than 4 labs each. The results from seven of the eleven methods used to measure CA125 showed relatively consistent differences from a high of +15% (Abbott AxSYM, ABB/AB1) to a low of -15% (Roche Elecsys/Cobas e411 & E170/e601) relative to the medians, and reasonable consistency across all five samples. The four exceptions were results by the Abbott Architect (ABH/AB1) and TOSOH ST-A1A (TOM/TO1) methods that were on average 35% and 42% higher, respectively, and, on the other side, by Siemens Immulite 2500 (DPF/DP5) and Fujirebio/Centocor (GAA/CE1) that were on average 20% and 40% lower than the medians. Please note that since the data for the Fujirebio/Centocor method were from only two labs they were excluded from the calculations of the all kit means and medians. In conclusion, with the exception of those four methods that were used by less than 15% of the labs, there was a reasonably good consensus in how CA125 was measured with no more than 25% difference between each method.

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. 2), and the remainder using one of two different **CA27.29** assays (Fig. 3). It should be noted that CA15-3 was intentionally left out of sample TM202 as part of the testing development process and thus all labs automatically received a passing credit for this sample. While several labs reported either "0" or less than their lower limit of detection for TM202, other labs reported a background level of CA15-3 or CA27.29 ranging from 0.5-2.2 U/ml for this sample. As Figure 2 shows, there were large differences in the way that the different methods measured CA15-3. The Abbott Architect (ABH/AB1), Siemens Immulite 2000/2500 (DPD & DPF/DP5) and Siemens ADVIA-Centaur (COB/BA1) methods all measured CA15-3 in these PT samples ranging from 20% to 40% higher, while, on the opposite side, the Roche Elecsys/Cobas e411 and E170/e601 (BME & BMR/BM1) and Beckman Unicel and Access (BCU & BCX/BC1) measured 14% to 28% lower than the medians. The remaining two methods, including Abbott AxSYM (ABB/AB1) and Ortho Clinical Diagnostics Vitros Eci/Q (JJC/JJ1), were closer and within 6% of the medians. Overall, the values are spread over a two-fold range. For CA27.29, results from only two methods were reported (Fig. 3). While the TOSOH method showed a slight positive bias for the low CA27.29 sample, TM205, and a slight negative bias for the high sample, TM201, the Siemens Centaur ADVIA showed the opposite effect. Overall, however, results from these two methods were relatively close.

Results for **CA19-9** (Fig. 4) were reported by only 62 labs. It should be noted that CA19-9 was intentionally left out of sample TM201 as part of the testing development process and thus all labs automatically received a passing credit for this sample. While a little less than 40% of the Beckman and Siemens ADVIA-Centaur groups reported either "0" or less than their lower limit of detection for CA19-9 in TM201, the remainder of the labs using those methods reported a background level averaging 1.2 and 2.8 U/ml of CA19-9, respectively.

All of the labs using either Roche or TOSOH methods reported average background levels of 2.5 or 5.6 U/ml, respectively. Half (31) of the labs used Siemens ADVIA-Centaur (COB/BA1), 13 labs (21%) used Beckman Unicel or Access (BCU & BCX/BC1, and 7 labs (11%) each used either the Roche Elecsys/Cobas e411 or E170/Cobas e601 (BME & BMR/BM1), or the TOSOH ST-A1A (TOM/TO1) methods. 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 overall by about 35%. In contrast, measurements of CA19-9 by Siemens ADVIA-Centaur were almost twice as high as 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. Thus, as Figure 4 shows, there seem to be clear differences between how different methods measured CA19-9, resulting in an almost three-fold difference from lowest to highest.

Results for **CEA** (Fig. 5) were reported by 174 labs using 12 different methods. The method groups in the summary table reflects the fact that all Beckman Unicel and Access/2 labs now use the enhanced CEA2 assay and all Siemens ADVIA-Centaur labs use the new CEA antibody. Nine methods gave results that were relatively consistent and on average were within +12 to -12% of the medians. In contrast, the three remaining methods including, the Ortho Clinical Diagnostics Vitros Eci/Eci Q (JJC/JJ1) and the new Vitros 5600 (which did not have a separate method code listing for this PT event (JJC/ZZZ) and the TOSOH ST-A1A (TOM/TO1) gave results that were on average 20-81% higher than the medians. Although there were limited data from only two labs, the Vitros 5600 gave results that were substantially lower than those from the Vitros Eci/Q.

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, 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, twice 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 TM201-205 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 4 sets for total and free PSA (2 sets for each different set of calibration standards), and two sets 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 106 labs using 8 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, 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  $\pm$ 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. On average, with the exception of the Vitros Eci/Q (JJC/JJ1) method, results varied from +11% by Siemens Dimension VISTA (DUV/DA2) to -6% by Siemens Immulite 1000/2000/2500 (DPB,DPD &

DPF/DP5). Not unexpectedly, the greatest variability was observed for the lowest AFP sample, TM201, especially for the Siemens ADVIA-Centaur which showed an above average positive bias for this low level of AFP. Overall, however, the AFP results suggest that most methods are reasonably well standardized.

Results were reported by 263 labs using 13 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  $\pm$ 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 PSA measurements in this PT was 19.2%, which is higher than in some recent PTs. As observed for the last several PTs, the results showed a more or less continuous gradient from +37% to +8% bias on average, without the clear grouping into “high” and “low” methods that was observed in some of the earlier events. Expectedly, the greatest variability was observed for the lowest PSA sample, TM203, which also had the highest percent free PSA. For this PT, two labs reported measurements by Beckman Unicel and Access/2 (BCU & BCX/BC3) calibrated with the WHO standards and these were much closer to the target values than those obtained by these same methods calibrated using the original Hybritech standards. 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.

Eighty-four 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 (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  $\pm$ 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. As seen in Figure 8, there was a clear distinction in results obtained with the two Beckman instruments and, to a lesser extent, with the Siemens Dimension, and the rest of the methods at the very low concentration in TM201 and TM203. Overall, the Beckman Unicel and Access and Siemens Dimension results were consistently higher than the target values, on average by 68%, for both Beckman instruments and 35% for the Siemens Dimension, respectively, whereas the results from the other methods ranged from 11% to 26% higher than the targets. 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 were comparable to the results from the other methods. Again, the 25% lower results are consistent with the information in the Beckman Technical bulletin referenced in the PSA discussion above. Across all methods, the results from the two low free PSA samples with identical concentrations of free but different % free PSA, TM203 and TM 201 at 0.2 ng/ml free PSA target for both with 22.7% and 9.8% ratio of free/total, respectively, were identical suggesting that the amount of total PSA in these samples did not substantially affect the measurements of free PSA.

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 22-43% 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 lower free PSA values coupled with the higher PSA values compared to the target values resulted in % free values that on average were lower than target by 15%.

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

89.9% for TM 201, 202 and 204 combined, 71.8% for TM203, and 85.0% for TM205, compared well with the mean of 10.9%, 25.1%, and 15.6% free PSA, respectively, for these 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 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](mailto:clepeptrs@health.state.ny.us), or directly to Kathi Wagner at (518) 402-4266 or by e-mail at [klw05@health.state.ny.us](mailto: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/lep/PT/oncology/index.htm>.

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

**Mail-out date:**

January 26, 2010  
May 11, 2010  
September 14, 2010

**Due date:**

February 10, 2010  
May 26, 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](mailto:schneid@wadsworth.org)