



# 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  
Chief of Staff

June 16, 2008

## New York State Tumor Marker Proficiency Test 5/2008 Evaluation <sup>1</sup>

Dear Laboratory Director,

Attached is a summary and evaluation of the New York State Proficiency Test from May 6, 2008 for Tumor Markers AFP, CA125, CA15-3, CA27.29 and CA19-9, CEA, PSA, free PSA and complexed PSA.

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).

**Important Reminder:** 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.

---

<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.

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 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.

#### 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. For this PT, the initial solution was made with high levels of the various analytes and then four other solutions were prepared from this one using four serial two-fold dilutions. Thus, your results for each analyte should increase linearly from TM181 to TM185 if plotted as two-fold increasing concentrations (see figures 10-18). 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  $\pm$  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 AFP, 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 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  $\pm$  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.

For grading purposes, all results for **AFP, CA125, CA15-3, CA19-9, CA27.29, CEA, PSA, 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.  $D/D_{max}$  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  $D_{max}$  being the maximal allowable deviation, i.e.  $3SD$ . Thus,  $D/D_{max}$  needs to be between  $-1$  and  $+1$  for a result to be scored. **Note: If your  $D/D_{max}$  is not within  $\pm 0.66$  (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  $D/D_{max}$  for each analyte to help you assess your results. If your **average  $D/D_{max}$  is greater than  $\pm 0.5$** , 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.

Results were reported by 120 labs using 11 methods to measure **CA125** (Fig.1). The results from eight of the eleven methods used to measure CA125 showed relatively consistent differences in measurements from a high of  $+25\%$  (TOSOH STA1A) to a low of  $-10\%$  (Roche Elecsys and E170) relative to the medians, and reasonable consistency across all five samples. The exceptions were Ortho Vitros Eci and Fujirebio/Centocor. The Ortho Vitros Eci results, especially, stood out because their measurements for TM182 and 183 were quite low compared to the medians ( $58\%$  and  $45\%$  lower, respectively) while those for the other three samples were closer to the median. The reason for this unusual pattern of results by the Ortho Vitros Eci is not known at this time. Finally, although limited data from only one or two labs was available, the Fujirebio/Centocor method showed a trend toward substantially lower measurements. Overall, the average bias part of the figure shows two main groups of assays whose results differed by approximately  $20\%$ , and one high and two low “outliers”. Thus, consistent with previous results, there are substantial differences in how CA125 is measured. However, with the exception of the unusual results from the Ortho Clinical Vitros Eci method, figure 10 shows that measurements of CA125 by all the other methods appear to be relatively linear across the five samples prepared by serial dilutions.

The MUC1 breast cancer antigen was measured by 104 labs, almost evenly split between those that used one of seven different **CA15-3** assays (Fig. 2), and those that used one of four different **CA27.29** assays (Fig. 3). Results for CA15-3 from the Beckman Unicel and Access (BCU & BCX/BC1), Siemens/Bayer ADVIA Centaur (COB/BA1), Ortho Vitros Eci (JJC/JJ1) and Roche Elecsys and E170 (BME & BMR/BM1) methods were similar and within  $11\%$  of the medians, whereas those from the Abbott AxSYM, Imx & Architect (ABB, ABM & ABH/AB1) methods were consistently higher than the medians on average by about  $12-32\%$ . In contrast, measurements by the Siemens/DPC Immulite 1000/2000 and 2500 (DPB, DPD & DPF/DP5) methods stood out with measured concentrations for CA15-3 more than twice as high as the medians. Consequently, the Siemens/DPC results 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 these Siemens/DPC methods was also reported in the last two CAP surveys (TM-B 2007 & TM-A 2008), although the discrepancies were not quite as great for those surveys. For CA27.29, the results (Fig. 3) from all four methods were similar and were on average within  $14\%$  of the medians. However, the figure also shows that the differences between the TOSOH methods (TOM/TO1 & TO2) and the Siemens/Bayer methods (COB & COS/BA1) were concentration dependent. At the lowest concentration in TM181, the values obtained with the TOSOH methods were twice those from the Siemens/Bayer methods. The difference gradually decreased as the concentration increased. Overall, the CA27.29 results were about  $18\%$  higher than the corresponding CA15-3 results. Furthermore, figures 11 and 12 show that measurements of both CA15-3 and CA27.29 by the different methods

appear to be relatively linear across the five diluted samples, although there appears to be clear nonlinear upwards trend with increasing concentration of CA15-3 with the Siemens/DPC methods.

Results for **CA19-9** (Fig. 4) were reported from only 57 labs. Just over half of these labs used Siemens/Bayer ADVIA Centaur (COB/BA1), 12 labs (21%) used Beckman Unicel or Access (BCU or BCX/BC1), and 7 labs (12%) used Roche Elecsys or E170 (BME & BMR/BM1). All the other methods were used by 3 or fewer labs. The results from all of the methods except for the Siemens/Bayer ADVIA Centaur (COB/BA1) were similar and within 6% of the medians. Measurements by the Siemens/Bayer ADVIA Centaur method stood out with results that ranged from 35-83% higher than the medians. 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/Bayer ADVIA-Centaur method were included in the calculation of the all kit medians. Thus, there seems to be a clear difference between measurements from the high Siemens/Bayer ADVIA-Centaur compared to those made by the other five methods. Nevertheless, measurements of CA19-9 by all the different methods (figure 13) appear to be relatively linear across the five samples, albeit with different slopes.

Results for **CEA** (Fig. 5) were reported by 178 labs using 13 different methods. Not unexpectedly, the greatest variability in CEA results was observed for the lowest sample, TM181. Apart from this variability, seven methods, including the Abbott Architect or AxSYM and Imx (ABH or ABB & ABM/AB1), Siemens/DPC Immulite 1000 or 2000 or 2500 (DPB or DPD or DPF/DP5), and Beckman Unicel or Access (BCU or BCX/BC1), gave measurements that were similar and on average within 10% of the medians, while the remaining six methods showed substantial positive or negative biases. Among these six methods, the highest measurements were from the TOSOH A1A and ST A1A (TOM/T01 & T02) ranging from 18 to 39% higher than the medians, followed by the Siemens/Bayer ADVIA-Centaur from 9% to 17% higher, while, on the other side, Roche E170 and Elecsys (BMR & BME/BM1) measurements were low ranging from 9-25% lower. Finally, measurements from the Ortho Vitros Eci (JJC/JJ1) were inconsistent, with a high for TM181 of +57% and a low of -42% and -55% for TM182 and 183, respectively. The reason for this unusual pattern of results for CEA from the Vitros Eci is not apparent at this time, but this pattern is similar to what was observed for the CA125 results noted above. 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. However, the graph in figure 14 shows that measurements of CEA by the different methods appear to be relatively linear with concentration. The exception is the Ortho Vitros Eci method, which exhibited the same deviation from linearity as shown for CA125 (figure 10).

As in past events, we used international reference standards to determine target values for AFP, PSA and free PSA. However, for reasons that are unknown at this time and, perhaps, in part due to the low levels of AFP in TM181 and 182, there were substantial differences between the measurements and calculated target values for AFP between the three methods used. Thus, target values were assigned using traceable International Standards for **PSA, free PSA and % free PSA** only. Although results for AFP, PSA, free 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 the PSA, free PSA and % free PSA analytes is provided in the discussion below, as well as in the summary tables and graphs.

Absolute target values for 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; and, PSA (90:10), 96/670, 1 µg per vial. Each vial was resuspended as recommended by NIBSC, followed by serial dilution to obtain six different concentrations. Each dilution was measured in

duplicate on a Beckman Access and a Roche Elecsys 2010 instrument and the measurements repeated later, if possible, with a different reagent lot, and in collaboration with Siemens/Bayer Diagnostics, on an ADVIA Centaur. 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 TM181-185 samples that had been measured in the same run as the standards. Thus, two sets of target values were obtained from the Beckman Access and Roche Elecsys 2010 instruments for total and free PSA and one set of target values was obtained for 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 103 labs using 10 different methods. 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/all kit median is given for each sample to compare measurement biases between the different methods. Most results were within 15% of the all kit medians except for the Siemens/Bayer ADVIA-Centaur (COB/BA1) and the Ortho Clinical Vitros Eci (JJC/JJ1), which gave results higher than the medians by 105% and 63% on average, respectively. It should be noted that the relative AFP measurements by the last two methods are higher for these PT samples compared to those from other recent PTs, but the reasons for this are unknown at this time. With the exception of the last two methods, the overall results suggest that most methods for AFP are well standardized, and all methods displayed good linearity across the entire concentration range (figure 15).

Results were reported by 268 labs using 13 different methods to measure total **PSA** (Fig. 7). The samples were prepared as mixtures of 12% free and 88% ACT-complexed PSA (the % free PSA is indicated 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 this PT was 11%. As for the last PT, the results showed a more or less continuous gradient from +25% bias to essentially 0% bias, without the clear grouping into “high” and “low” methods seen in some of the previous events. Furthermore, figure 16 shows that measurements of PSA by the different methods appear to be linear across the five serially diluted samples.

Seventy-one labs measured **free PSA** (Fig. 8) with slightly less than half (44%) of the results reported with the Beckman Hybritech Access or Unicl (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  $\pm$ 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. Overall, the Beckman Unicl and Access (BCU and BCX/BC1) and Siemens/Dade Behring Dimension (DUD/DA1) results were consistently higher than the target values, on average 54% and 26%, respectively, whereas most of the results from the other methods were within 11% of the targets. Nevertheless, as figure 17 shows, the measurements of free PSA by the different methods appear to be relatively linear across the five diluted samples.

Figure 18 shows the % free in relation to the total PSA concentration in each sample. As can be seen, each method combination trends towards lower % free PSA values as the total PSA concentration increases from 1.8 to 28.7 ng/ml. This suggests that none of the method combinations are perfectly equimolar. However, the effect is generally relatively small and unlikely of clinical significance.

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 19-27% higher than the targets. In contrast, both Siemens/DPC and Abbott measured free PSA right on target, but the above target values for total PSA resulted in the % free PSA to be lower than the target by 5-18%. The other two method combinations, Roche Elecsys or E170 (BME or BMR/BM1) and Siemens/Dade Behring Dimension (DUD/DA1), gave ratios that on average were essentially almost identical to the targets, indicating that their respective measurements relative to the targets for both total and free PSA were comparable.

**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 11 labs measured **complexed PSA**, and all of these used the Siemens/Bayer ADVIA Centaur or ACS-180 method. Furthermore, the mean % complexed PSA calculated from these values of 85.1% compared well with the mean of 12.6% free PSA for the five 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 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.

If you have any questions or wish to discuss some of the issues alluded to 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 schedule for the last 2008 Tumor Marker Proficiency Test mail-out is:

**Mail-out date:**

September 9, 2008

**Due date:**

September 24, 2008

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)