



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

October 15, 2008

New York State Tumor Marker Proficiency Test 9/2008 Evaluation ¹

Dear Laboratory Director,

Attached is a summary and evaluation of the New York State Proficiency Test from September 9, 2008 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 \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 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 \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.

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

For grading purposes, all results for **AFP, CA125, CA15-3, CA19-9, CA27.29, CEA, PSA, PSA2, free PSA, % free PSA and complexed PSA** were evaluated based on their respective peer group mean. In order for you to more easily compare your results to those of your peer group, we calculated a D/Dmax value and displayed it directly under your individual results. D/Dmax is a measure of how much your result (x) deviates from your peer group mean, $D/D_{max}=(x-\text{mean})/3SD$, 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 122 labs using 11 methods to measure **CA125** (Fig.1). The results from six of the eleven methods used to measure CA125 showed relatively consistent differences from a high of +10% (Siemens Immulite 1000) to a low of -14% (Siemens Immulite 2500) relative to the medians, and reasonable consistency across all five samples. The exceptions were the Abbott Axsym, IMx and Architect (ABB, ABM & ABH/AB1) and the two TOSOH A1A and ST A1A (TOM/TO1 & TO2) methods that were 26%, 42% and 49% higher, respectively, and, on the other side, the Ortho Clinical Vitros Eci (JJC/JJ1) that was 24% lower. Although limited data from only two labs were available for the Fujirebio/Centocor (GAA/CE1) method, this method's results were on average 40% lower than the medians. Thus, consistent with previous results, there were substantial differences in how CA125 was measured.

The MUC1 breast cancer antigen was measured by 109 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). Results for CA15-3 from three methods, Abbott Axsym, Imx & Architect (ABB, ABM & ABH/AB1), Beckman Unicel and Access (BCU & BCX/BC1), and Siemens ADVIA Centaur (COB/BA1) were almost identical, whereas those from the Ortho Vitros Eci (JJC/JJ1) and Roche Elecsys/Cobas e411 and E170/Cobas e601 (BME & BMR/BM1) methods were on average consistently lower than the medians by 12% and 20%, respectively. In contrast, measurements by the Siemens Immulite 1000/2000 and 2500 (DPB, DPD & DPF/DP5) methods stood out with concentrations for CA15-3 more than twice as high as the medians. Consequently, the Siemens Immulite 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 Immulite methods was also reported in the last two CAP surveys (TM-A 2008 & TM-B 2008), although the discrepancies were not quite as great for those surveys. For CA27.29, results with only two methods were reported (Fig. 3). Those obtained with TOSOH ST A1A (TOM/TO2) were between 12% and 79% (average 31%) higher than those from Siemens ADVIA-Centaur and ACS-180 (COB & COS/BA1). Furthermore, this difference was more pronounced at lower levels of CA27.29. In addition, excluding the Siemens Immulite methods, the CA27.29 results were about 20% higher than the corresponding CA15-3 results.

Results for **CA19-9** (Fig. 4) were reported by only 64 labs. Just under half (31) of these labs used Siemens ADVIA Centaur (COB/BA1), 14 labs (22%) used Beckman Unicel or Access (BCU or BCX/BC1), 8 labs (12.5%) used Roche Elecsys/Cobas e411 or E170/Cobas e601 (BME & BMR/BM1), and 7 labs (11%) used one of the TOSOH A1A or ST A1A (TOM/TO1 or TO2) methods. The results from the two Beckman and Roche methods were essentially identical. In contrast, measurements of CA19-9 by Siemens ADVIA Centaur were 57-85% higher. 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, there seems to be a clear difference between measurements from the high Siemens ADVIA-Centaur compared to those made by the other three methods. A rather unexpected result was obtained with the two TOSOH methods. Whereas for all the other methods, CA19-9 levels measured in TM190 were approximately two-fold those in TM186, and with TM188, 189 somewhere in between (TM187 is excluded from this analysis), the values measured by TOSOH for these samples were nearly identical across the four samples (Fig. 5). In other words, these methods do not seem to discriminate between high and low levels of CA19-9. At present, we have no explanation for this unexpected observation.

Results for **CEA** (Fig. 6) were reported by 181 labs using 16 different methods. Changes to the assay configuration and/or reagents for both the Beckman Access and Unicel, and the Siemens Centaur and ACS-180 instruments were introduced by those companies recently and required the participants to carefully choose their reagent codes, as both the old and new reagents were still in use. A few labs did not indicate which reagent they used, which made evaluation of their results difficult. Overall, there was poor concordance between the different methods, with the results spanning an approximately two-fold range. The highest measurements were from the TOSOH A1A and ST A1A (TOM/T01 & TO2) methods ranging from 35 to 43% higher than the medians, whereas, on the other side, the two Roche methods, Elecsys/Cobas e411 and E170/Cobas e601 (BME & BMR/BM1), and the Ortho Clinical Vitros Eci (JJC/JJ1) gave measurements that were between 9% and 54% lower than the medians. Overall, there was an almost two-fold difference from lowest to highest. Interestingly, the results by the Ortho Vitros Eci method (8 labs) separated into two groups. The first group's measurements were higher and generally closer to the all lab means, while measurements by the second group were substantially lower. Additional information provided by the labs using this method indicated that two different older (lower) lot numbers were associated with the higher measurements, while two different newer (higher) lot numbers were associated with the lower measurements. Although the method group was small, the differences between the results from the two sub-groups were statistically significant. In addition, this negative measurement bias by the newer lots was greatest for the two lowest CEA samples, TM187 and 188. As a consequence, the mean measurement of TM187 by the newer lots was only about half as much as that made by the older lots (1.6 vs 2.9 ng/ml), and for TM188 was about 25% lower (6.6 vs 8.8 ng/ml). These differences might possibly have clinical significance, since a common reference range cut-off for CEA is 3.0 ng/ml.

The new enhanced CEA2 assay for the Beckman instruments did not show a significant difference when results from the Beckman Unicel using the enhanced CEA2 (BCU/BC5) assay were compared to the Unicel using the original assay (BCU/BC4). However, when the enhanced CEA2 assay (BCX/BC5) was compared to the original assay on the Access (BCX/BC4), the results were higher for all five samples by about 11% (see Figure 6), p value <0.002. Similarly, the CEA results from the Siemens ADVIA-Centaur and ACS-180 assays using the original antibody pool (COB & COS/BA2) were on average about 7% higher for all five samples compared to the new pool (COB/BA3), p value <0.001. While we can't assess whether these differences between the old and new reagents from those manufacturers are also seen with actual patient samples, we would like to remind labs that it is their responsibility to validate any new set of reagents. Overall, the CEA results from this PT, shown in Figure 6, as well as those from previous PTs, suggest that large differences exist among the methods used to measure CEA.

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

Absolute target values for AFP, PSA and free PSA were established based on the following International Standard preparations that were obtained from NIBSC (National Institute for Biological Standards and Control, A WHO International Laboratory for Biological Standards, Blanche Lane, South Mimms, Poters Bar, Hertfordshire EN6 3QG, UK, <http://www.nibsc.ac.uk>): PSA (free), 96/668, 1 µg per vial; PSA (90:10), 96/670, 1 µg per vial; and AFP, 72/225, 100,000 IU per vial with a conversion factor of 1.21 ng/IU. Each vial

was resuspended as recommended by NIBSC, followed by serial dilution to obtain six different concentrations. Each dilution was measured in duplicate on a Beckman Access and a Roche Elecsys 2010 instrument and the measurements repeated later, if possible, and in collaboration with Siemens Diagnostics, on an ADVIA Centaur (only AFP and total PSA). 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 TM186-190 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, total PSA and free PSA and from the Roche Elecsys 2010 instrument for AFP and free PSA with only one set for total PSA on the Elecsys. Additionally, one set of target values was obtained for AFP and total PSA from the Siemens/Bayer ADVIA Centaur. These were then averaged to obtain the target values for each sample and analyte. The respective target values with their standard deviations can be found in the summary tables.

Results for **AFP** (Fig. 7) were reported from 107 labs using 9 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/target value is given for each sample to compare measurement biases between the different methods. Most results were within 10% of the target values except for the Ortho Clinical Vitros Eci (JJC/JJ1), which gave results lower than the medians by 18% on average. With the exception of this method, the overall results suggest that most methods for AFP are well standardized.

Results were reported by 268 labs using 13 different methods to measure total **PSA** (Fig. 8). The samples were prepared as mixtures of free and ACT-complexed PSA mixed in different proportions as indicated by the % free PSA in Figure 8. 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 11.5%. As for the last PT, the results showed a more or less continuous gradient from +23% bias to essentially 0% bias, without the clear grouping into “high” and “low” methods seen in some of the previous events. Sample pairs TM186, 187 and 188, 189 were targeted for identical total PSA but different % PSA values. All methods measured the total PSA levels in each sample pair fairly equally, indicating that all methods are essentially equimolar.

Eighty-two labs measured **free PSA** (Fig. 9) with the majority (39%) of the results reported with the Beckman Hybritech Access or Unicel (BCX or BCU/BC1) methods. All results were evaluated according to traditional peer group statistics and received a passing score if they fell within the mean $\pm 3SD$. 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 Unicel and Access (BCU and BCX/BC1) and Siemens Dimension (DUD/DA1) results were consistently higher than the target values, on average by 49% and 20%, respectively, whereas most of the results from the other methods were within 11% of the targets.

As in prior surveys, the figure for **% free PSA** (Fig. 10) 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 21-30% higher than the targets. In contrast, both Siemens Immulite and Abbott measured free PSA right on target, but the above target values for total PSA from these methods resulted in the % free PSA to be lower than the target by 2-16%. The other two method combinations, Roche Elecsys/Cobas e411 or E170/Cobas e601 (BME or BMR/BM1) and Siemens 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 ADVIA Centaur method, with good agreement between the labs. Furthermore, the mean % complexed PSA calculated from these values of 88.5% for TM186, 188 and 190 compared well with the mean of 10.1% free PSA for these three samples and, that of 62.2% for TM187 and 189 compared well with 35.9% for these two samples.

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

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

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

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

Note: Please be aware that in each subsequent event, fields will be pre-populated based on what you entered this time or a previous time. **Therefore, make sure that the selected instruments and reagents are correct, whether this is pre-populated from the last event or newly entered information.** This is

important and in your interest since we need this information to properly evaluate your results and compare them to those of your peers. **You are at risk** of receiving a technical failure for results evaluated outside of the correct peer group or an administrative failure for incorrect methodology. **No changes can be made for incorrect or missing information once the submission deadline has passed.**

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

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

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

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

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

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

Mail-out date:

January 27, 2009
May 5, 2009
September 15, 2009

Due date:

February 11, 2009
May 20, 2009
September 30, 2009

Erasmus Schneider, Ph.D.
Director, Oncology Section
Clinical Laboratory Evaluation Program
Wadsworth Center
Empire State Plaza
Albany, NY 12201-0509

Ph: (518) 474-2088
FAX: (518) 474-1850
email: schneid@wadsworth.org