



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

June 15, 2007

New York State Tumor Marker Proficiency Test 5/2007 Evaluation ¹

Dear Laboratory Director,

Attached is a summary and evaluation of the New York State Proficiency Testing mail-out from May 8, 2007 for Tumor Markers AFP, CA125, CA15-3, CA27.29 and CA19-9, CEA, PSA, free PSA and complexed PSA.

Most labs submitted PT results electronically, thank you very much, but a small number of those did not give their instrument/reagent codes correctly. Please carefully check the instrument/reagent code listings in the pull-down menu on the electronic form to ensure that you have selected the correct method codes as this could impact your PT evaluation.

To those that have not yet done so, we remind you that electronic submission of results is now mandatory for PT submissions. The electronic proficiency testing reporting system (EPTRS) is a free utility on the Department's Health Provider Network (HPN). The HPN is a secure website and requires all users to obtain an HPN ID in order to access the HPN and EPTRS application. If your laboratory does not already have an HPN account, you should start the process by contacting the Help Desk at (866) 325-7743, or by email at ecrs@health.state.ny.us. Questions regarding the 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.

We would like to comment on some difficulties that were encountered with electronic submission of the PT results. A few changes in the electronic report form caused a little confusion, particularly for labs submitting this way for the first time. The absence of data in required fields was a common problem and this prevented movement from the event menu page to the result pages. In particular, the newly added PSA2 line in the event menu created some confusion. 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 on the PSA2 line. Some labs entered the same PSA results twice, possibly because they did not realize they had to select "test not offered" for PSA2 and thus were not able to move off the event page. Other required fields that caused problems were those for the range of total PSA for measuring free PSA and calculating the free/total PSA ratio. Text, such as "all levels", "NA" (not applicable) or "see comments" could be entered here. If the test was

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.

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.

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 and upper limits of detection. Another common issue occurred with the use of more complex age-specific reference ranges or other population demographic parameters for test interpretations instead of simply flagging results as normal or abnormal in a general quantitative manner. This was an issue particularly for labs that use probability tables for risk of cancer for interpreting PSA and free/total PSA ratio results. In addition, the free/total PSA ratio was also separated into low, intermediate and high risk or "test not reported" on the electronic form which was different from the normal/abnormal selection on the paper form. We asked the labs that contacted us to indicate in the comments section how they would handle interpretation in this case. In the past, labs frequently provided an additional page with probability or risk tables. Although the comment section will serve to provide some of this information at the present time, we will work to improve the electronic PT form to address these issues and facilitate collection of the required information in a more satisfactory way.

Lastly, a few labs encountered problems with using the drop-down menus. The first selection at the top of the menu is the blank and a few labs had trouble finding it to use when appropriate. Some labs selected qualifiers apparently in error and should have used the "blank" selection, and others selected the incorrect reagent code, for example, a reagent mis-matched with the incorrect instrumentation. Please check these entries carefully in the future to make sure your selection is correct. Please be aware that in the next event, these fields will be pre-populated based on what you entered this time.

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.

We would like to remind you that **all information** requested on the result sheet **must be provided**. **Please make sure that the codes for instruments and reagents are present and correct, including those that you select electronically, and that your results are written on the correct line and in the correct column.** 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 a number of labs that have failed to provide instrument and reagent codes. *This omission may result in an automatic failure.* **No changes can be made for incorrect or missing information once the submission deadline has past.** This also applies to the CQ code which must be entered next to your signature and which is found in the upper right corner of your Certificate of Qualification. We need this code in order to properly record and track your results.

For your information, we also included a tabular summary of all the results with high/low cut-off values (mean +/- 3 SD) 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.

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/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. 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 123 labs using 11 methods to measure **CA125** (Fig.1). In contrast to previous events, we did not see a clear separation into different subgroups. Nevertheless, there were substantial differences between methods, particularly for the highest TOSOH ST-A1A (TOM/TO2) at +26% and lowest Fujirebio/Centocor RIA (GAA/CE1) at -20%. This corresponds to absolute values for TOSOH that ranged from 41-80% higher than those for Centocor. In addition, results from TOSOH A1A (TOM/TO1), Beckman Unicel and Access (BCU and BCX/BC1), and Abbott AxSYM (ABB/AB1), were consistently higher than the medians on average by 9-14%, while, on the other side, those from Roche Elecsys and E170 (BME & BMR/BM1) and Johnson & Johnson Vitros Eci (JJC/JJ1) were consistently lower by 14-16%. Finally, four methods, including Bayer/Siemens ADVIA Centaur (COB/BA1) and DPC Immulite 1000, 2000 and 2500 (DPB, DPD, & DPF/DP5) were essentially identical and within 2% of the medians. Thus, consistent with previous results, there are large differences in how CA125 is measured.

The MUC1 breast cancer antigen was measured by 105 labs, 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 Abbott AxSYM (ABB/AB1), Beckman Unicel and Access (BCU & BCX/BC1) and Bayer/Siemens ADVIA Centaur (COB/BA1) methods were similar and within 8% of the medians, whereas those from the Roche Elecsys and E170 (BME & BMR/BM1) and Johnson & Johnson Vitros (JJC/JJ1) methods were consistently lower than the medians on average by about 25%. In contrast, measurements by the 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 DPC results were not included in the calculations for the all lab mean or for the all kit means and medians

because of the impact that these large differences would have had on these values. Interestingly, a similar result was also reported in the latest CAP survey (TM-A 2007). For CA27.29, results (Fig. 3) from the two TOSOH methods, A1A (TOM/TO1) and ST-A1A (TOM/TO2), were 7% and 15% higher than the medians, whereas those from the two Bayer/Siemens, ADVIA Centaur (COB/BA1) and ACS-180 (COS/BA1), methods were 7% and 17% lower than the medians. Overall, the CA27.29 results were about 20% higher than the CA15-3 results.

Results for **CA19-9** (Fig. 4) were reported from only 56 labs. More than half of these labs (29 or 52%) used Bayer/Siemens ADVIA Centaur (COB/BA1), 11 labs (20%) used Beckman Unicel (BCU/BC1) or Access (BCX/BC1), and 6 labs (11%) used Roche Elecsys or E170 (BME & BMR/BM1). All the other methods were used by 3 or fewer labs. The results from the Roche Elecsys and E170 (BME & BMR/BM1), Beckman Access (BCX/BC1) and Unicel (BCU/BC1), DPC Immulite 2000 (DPD/DP5) and TOSOH A1A (TOM/TO1) methods were similar and within 10% of the medians. In contrast, those from Bayer/Siemens ADVIA Centaur (COB/BA1) and Fujirebio/Centacor (GAA/CE1) were on average 50-60% higher, while, on the other side, those from TOSOH ST-A1A (TOM/TO2) were 22% lower than the medians. Thus, there seems to be a clear difference between measurements from the high Bayer and Fujirebio and the low TOSOH ST-A1A methods compared to those made by the other five methods.

Results for **CEA** (Fig. 5) were reported by 182 labs using 15 different methods. Whereas the TOSOH ST-A1A (TOM/TO2) method measured CEA substantially higher than any of the other methods, ranging from 28-39% higher than the medians, the majority (12/15) of the methods gave measurements that were on average within 15% of the medians. In contrast, the two remaining methods, Roche E170 and Elecsys (BMR & BME/BM1), gave consistently lower measurements ranging from 20-26% lower than the medians. Figure 5 also shows that the positive bias seen with some of the methods is greatest at the highest level of CEA (TM166). The CEA results from this PT, as well as those from previous PTs, suggest that 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, with a different reagent lot, and in collaboration with Bayer/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 TM166-170 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 AFP, total and free PSA and one set of target values was obtained for AFP and total PSA from the Bayer/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 103 labs using 11 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 8% of the target, with the exception of the Johnson & Johnson Vitros Eci (JJC/JJ1), which gave results on average 15% lower than the targets. These results suggest that overall the methods for AFP are well standardized. Indeed, had we evaluated labs by target \pm 30%, none of the labs would have failed.

Results were reported by 276 labs using 17 different methods to measure total **PSA** (Fig. 7). The samples were prepared as a mixture of 10% free and 90% ACT-complexed PSA for TM166 and 167 and as a mixture of 30% free and 70% ACT-complexed PSA for TM168, 169 and 170. 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 1.09. The majority (twelve) of the methods, including Johnson & Johnson Vitros (JJC/JJ1), Dade Behring Dimension (DUD/DA1), Abbott AxSYM, Imx and Architect (ABB, ABM & ABH/AB1), Bayer/Siemens ADVIA-Centaur (COB/BA2), DPC Immulite new generation PSA (DPB, DPD or DPF/DP6), DPC Immulite 1000, 2000, and 2500 (DPB, DPD & DPF/DP5), Roche E170 (BMR/BM1) and Elecsys (BME/BM1) gave results for PSA with an average positive bias of 5.8 \pm 4.0%. However, five methods, including TOSOH A1A and ST-A1A (TOM/TO1 & TO2), Bayer/Siemens ACS-180 (COS/BA1) and Beckman Access and Unicel (BCX & BCU/BC1), consistently gave higher PSA results with an average positive bias of 17.5 \pm 3.2%. This difference is statistically significant with $p < 0.0001$. There are currently two different calibration methods for the various PSA assays as described in an article by Julie McDowell in the June 9, 2005, online newsletter of the American Association of Clinical Chemistry (AACC): [Clinical Laboratory Strategies](#). One of these is the original Hybritech Tandem-R assay (or “traditional” method), and the second is the WHO standard based on the First International Reference Standard for PSA available from NIBSC (PSA 90:10, 96/670). The latter is the same standard that was used to determine the target values for our PT samples. Since the two calibration standards can result in differences in the PSA measurements obtained from different assays, it is likely that the five methods with substantially higher PSA results than those observed from the majority of the other methods were calibrated against the original Hybritech standard, while the other methods used the WHO calibration standard. As noted previously, the standard used for calibration can have clinical implications when the result measured is close to a decision point, such as the 4 ng/ml cut-off, and thus it may be important for the physician to know what calibrator you used.

Sixty-seven labs measured **free PSA** (Fig. 8) with more than half (52%) 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. As observed in previous PTs, Beckman Unicel and Access (BCU and BCX/BC1) results were consistently higher (on average 40% for this PT) than the target values for these samples, whereas most of the results from the other methods were within 10% of the targets.

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 Access assay resulted in %free PSA values that were between 17-21% higher than the targets. In contrast, the

overall lower free PSA values measured with the DPC Immulite assay resulted in %free PSA values that were between 8-15% lower than the targets. The remaining three method combinations, including Roche Elecsys and E170 (BME & BMR/BM1), Abbott AxSYM (ABB/AB1) and Dade Behring Dimension (DUD/DA1) gave ratios that on average were within 5% of the target. It should be noted that the %free PSA that is calculated from measurements of complexed PSA does not always give results that are comparable to those calculated directly from measurements of free PSA in the same samples, especially if non-homogeneous methods are used. In the latter case, the results are usually substantially different from those obtained either directly from free PSA measurements or from a homogeneous combination of assays for complexed and total PSA.

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' policy, 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, results must be given as percent free PSA, not as a fraction.

Only 12 labs measured **complexed PSA**, and all of these used the Bayer Centaur or ACS-180 method. Furthermore, the mean % complexed PSA calculated from these values of 85.9% compared well with the mean of 10.6% free PSA for TM166 and TM167, and the mean % complexed PSA of 72.2.% compared well with the mean of 25.9% free PSA for samples TM167-170.

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 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 3b). 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 AFP, PSA, and free PSA clearly shows that there are a few methods that appear to be calibrated differently from the rest of the methods used to measure these analytes 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 be posted on our website at <http://www.wadsworth.org/labcert/lep/PT/oncology/index.html>.

For your information, the schedule for the remaining 2007 Tumor Marker Proficiency Test mail-out is:

Mail-out date:

September 11, 2007

Due date:

September 26, 2007

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