



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.
Acting Commissioner

March 6, 2007

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

Dear Laboratory Director,

Attached is a summary and evaluation of the New York State Proficiency Testing mail-out from January 27, 2007 for Tumor Markers AFP, CA125, CA15-3, CA27.29 and CA19-9, CEA, PSA, free PSA and complexed PSA.

Many 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 strongly encourage you to submit results electronically. Among the benefits of submitting electronically is that there will be no questions about timely filing, and it will save you the expense of mailing results by certified mail. 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 [eclrs@health.state.ny.us](mailto: eclrs@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](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).

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.**

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.

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 from 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, $D/Dmax = (x - \text{mean}) / 3SD$, 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), 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 122 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, from a high of +33% with TOSOH ST-A1A (TOM/TO2), to a low of -29% for the original Fujirebio/Centocor (GAA/CE1) RIA method, which amounts to an almost two-fold difference in absolute values. Results from four other methods, including TOSOH A1A (TOM/TO1), Beckman Unicel and Access (BCU and BCX/BC1), and Abbott AxSYM (ABB/AB1), were also consistently higher than the medians by 7-18%, 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 14-26%. Finally, four methods, including Bayer ADVIA Centaur (COB/BA1) and DPC Immulite 1000, 2000 and 2500 (DPB, DPD, &

DPF/DP5) were similar 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 103 labs, evenly split between those that used one of six different **CA15-3** assays (Fig. 2), and those that used one of four different **CA27.29** assays (Fig. 3). Results from the Bayer ADVIA Centaur (COB/BA1) and Abbott AxSYM (ABB/AB1) methods were nearly identical, but on average 1/3 higher than those from the Roche Elecsys and E170 (BME & BMR/BM1), Beckman Unichel and Access (BCU & BCX/BC1) and Johnson & Johnson Vitros Eci (JJC/JJ1) methods, which were similar to each other. In contrast, measurements by the DPC Immulite 1000/2000 (DPB/DPD/DP5) method stood out with measured concentrations for CA15-3 between 74-80% higher than the medians. Interestingly, a similar result was also reported in the latest CAP survey (TM-A 2007). The four methods used to measure CA27.29 (Fig. 3), including TOSOH A1A and ST-A1A (TOM/TO1 & TO2) and Bayer Centaur and ACS-180 (COB & COS/BA1), gave results that were generally similar and were on average within 8% of the medians. No significant difference between values obtained for CA15-3 and CA27.29 were observed.

Results for **CA19-9** (Fig. 4) were reported from only 54 labs. More than half of these labs (30 or 56%) used Bayer ADVIA Centaur (COB/BA1), 10 labs (19%) used Beckman Unichel (BCU/BC1) and Access (BCX/BC1), and 5 labs (9%) used Roche Elecsys and 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 Unichel (BCU/BC1), DPC Immulite 2000 (DPD/DP5) and TOSOH A1A (TOM/TO1) methods were similar and within 11% of the medians. In contrast, those from Bayer and Fujirebio/Centocor (GAA/CE1) were on average more than 50% higher, while, on the other side, those from the TOSOH ST-A1A (TOM/TO2) method were up to 45% 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 33-47% higher than the medians, the majority (10/15) of the methods gave measurements that were on average within 11% of the medians. However, the four remaining methods, Beckman Unichel and Access (BCU & BCX/BC1) and Roche E170 and Elecsys (BMR & BME/BM1), gave consistently lower measurements ranging from 11-26% lower than the medians. 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 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 TM161-165 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 total PSA from the 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 102 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 5% of the target, with the exception of the Bayer ACS 180 (COS/BA1) and Johnson & Johnson Vitros Eci (JJC/JJ1), which gave results on average 15 and 17% lower than the targets, and the Roche Elecsys (BME/BM1) and Beckman Access (BCX/BC1), with an average negative bias of 8% (range -5 to -10%) across all samples. These results suggest that overall the methods for AFP are well standardized.

Results were reported by 275 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 TM162, 164 and 165 and as a mixture of 25% free and 75% ACT-complexed PSA for TM161 and 163. 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 1.11 which is the same as the value reported by Rafferty et al. (Clin. Chem. 46; 1310-1317, 2000). The majority (ten) of the methods, including Bayer ADVIA-Centaur (COB/BA2) and ACS 180 (COS/BA1), Roche E170 (BMR/BM1), DPC Immulite1000, 2000 (DPB & DPD/DP5) and 1000 or 2000 new generation PSA (DPB or DPD/DP6), Dade Behring Dimension (DUD/DA1), and Abbott AxSYM, Imx and Architect (ABB, ABM & ABH/AB2) gave results for PSA with an average positive bias of 8.8 \pm 2.8%. However, five methods, including Johnson & Johnson Vitros Eci (JJC/JJ1), TOSOH A1A and ST-A1A (TOM/TO1 & TO2) and Beckman Access and Unicel (BCX & BCU/BC1), consistently gave higher PSA results with an average positive bias of 20.2 \pm 3.3%, while the last two methods, Roche Elecsys (BME/BM1) and DPC Immulite 2500 (DPF/DP5), consistently gave lower results with an average negative bias of -3.7% \pm 1.4%. These differences are statistically significant with $p < 0.001$ between each group. In addition, the Johnson & Johnson Vitros (JJC/JJ1) method exhibited an above average positive bias in the measurement of PSA at the lowest level of 2.1 ng/ml in TM162. 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.

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 35% for this PT) than the target values for these samples, whereas most of the results from the other methods were closer to the targets. Interestingly, an especially large positive bias was observed in the free PSA measurements by most methods for TM163 with a mean total PSA of 3.7 ng/ml and 25% free PSA. Figure 8 clearly shows that, although TM163 and 165 have essentially the same level of total PSA (mean 3.7 and 3.8 ng/ml, respectively), most

methods gave higher results for free PSA for TM163 with 25% free compared to those for TM165 with 10% free. This was particularly the case for the Beckman Access and Unicel (BCX and BCU/BC2) and Roche Elecsys (BME/BM1) with results for TM163 that were between 12-20% higher than their results for TM165 when compared to the targets. In contrast, for TM161 and 164, which also have similar levels of mean total PSA (9.2 and 9.6 ng/ml, respectively) but different proportions of free PSA, the relative performance of most methods was reversed, i.e., their results relative to the target was lower in TM161 with 25% free PSA than for TM164 with 10% free PSA. Again this effect was most pronounced for the Beckman methods, which gave free PSA results 11-22% lower in TM161 with 25% free PSA than they did in TM164 with 10% free PSA when compared to their respective targets. Most other methods showed only about 8-10% negative bias in measurements for TM161 compared to TM164. Thus, it appeared that there was a greater positive bias for most methods in samples with low levels of total PSA but high %free PSA, while the reverse was true at lower levels of %free PSA but high levels of total 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 Access assay resulted in %free PSA values that were between 4-24% higher than the targets. In contrast, all the other method combinations gave a ratio that on average was between 5-13% below 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 91.9% compared well with the mean of 11.3% free PSA for TM162, 164, 165, and the mean % complexed PSA of 80.0% compared well with the mean of 22.35% free PSA for samples 161 and 163.

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 write that information on the form and/or if possible attach a table or example result 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 asterisk and write a note on the back that there are two or more reference values and list the other values.

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>.

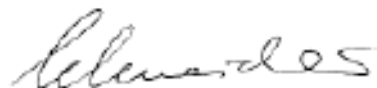
For your information, the tentative schedule for the 2007 Tumor Marker Proficiency Test mail-outs is:

Mail-out date:

**May 8, 2007
September 11, 2007**

Due date:

**May 23, 2007
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

