PT Frequency Subcommittee of the PT Expert Committee
Final Report
August 3, 2009

This report is respectfully presented on behalf of the PT Frequency Subcommittee, in response to their charge from the PT Expert Committee in April, 2008. It represents the collective efforts of everyone on the subcommittee, through the conference calls, investigative studies, and drafting and reviewing reports. It also reports the consensus recommendations approved by all members of the subcommittee.

Contents of the Final Report: **Page**

This report references documents available on the NELAC website (New Jersey Study) and an electronic file, referenced below.

http://www.nelac-institute.org/cms/posts/1213773688.php#pab1_6

PT Frequency Subcommittee of the PT Expert Committee
Final Report – August 3, 2009

I. Background

The Frequency Subcommittee Charter was established by the TNI Proficiency Testing (PT) Expert Committee in April 2008. The purpose was to gather objective information to be used by the TNI PT Expert Committee to assist their efforts to determine if the frequency requirements for PT in the TNI standard should be changed as was proposed by some TNI members during the draft standard development comment period.

The specific charge to the subcommittee was "…to gather and analyze information on the issue of the frequency of proficiency testing", and to include:

- existing information in the published literature,
- economic assessments on the impact of a change in frequency, and
- opinions of stakeholder interests.

The subcommittee's report is to present "it's findings/recommendations" to the Expert Committee. This report is to be submitted prior to the August 11 meeting of the Expert Committee.

The Subcommittee is comprised of eight members that represent accredited laboratories, PT providers, and accrediting bodies. Dan Tholen (A2LA) is the Subcommittee Chair and Jeff Lowry (ERA) is the Vice-Chair. Members: Judy Morgan, Gary Dechant, Chuck Wibby, Michella Karapondo, Reza Karimi, Rae Anne Haynes. Associate membership to the Subcommittee is open to any TNI member, by request. Associate members: Kirstin McCracken, Jim Pletl, Agnes vanLangenhoven.

The Subcommittee's activities in the last year include 20 teleconferences to manage progress on several fronts, as listed below. Approved minutes from all teleconferences are posted on the TNI website.

- comparison study of PT performance between laboratories that analyze 2 PT samples per year vs. laboratories that analyze 1 PT sample per year in New Jersey's designated programs;
- related studies of data from Maine and Wisconsin (both were unsuccessful due to unavailability of sufficient data);
- a survey of overall PT performance in NELAC and non-NELAC labs in different States, for each matrix;
- a survey of State AB's on key questions related to PT;
- survey of economic aspects of PT for NELAP accredited labs;
- discussion paper on factors to consider in assessing the economic impact of a change in required frequency;
- review of scientific literature and current international practices.

The following sections provide a brief overview of the material that was reviewed by the subcommittee.

A general statement on applicability of this review:
The committee agreed that the type of information that we could provide represents a very limited part of the question about optimal PT frequency or even the more limited question of whether twice a year is better than once. Questions of cost-effectiveness are important, as are opinions from stakeholders regarding credibility. Deeper investigation and review leads to the problems with any "one size fits all" PT policy, due to unavoidable differences between labs regarding analytes and ranges tested, accuracy required for customer use, the frequency of testing, past PT performance, and other QC practices

undertaken by the laboratory. Deeper issues also relate to the information value of PT as currently practiced compared to opportunities to disseminate the information in every study – if PT were used as a QC tool by the lab and as an educational vehicle by PT providers, the labs that perform acceptably in study after study would also obtain value from every exercise.

II. Summary of PT Study Data from New Jersey State PT Program

The analyses have not changed from what was presented in August, 2008. The analyses results and the raw data have been posted on the TNI website for review by any interested party. There was one additional analysis of the possible impact of "duplicate" results in the New Jersey data. These are results that are legitimate PT results for compliance with requirements, but they are not obtained independently from other results from that lab. The incidence of duplicate results was different in different classes of analytes and in the two groups of laboratories in this study. However the analysis showed that there was no change in the conclusions regarding different performance in the groups of laboratories that take PT once per year and labs that take PT twice a year. The subcommittee decided to retain the original analysis, since the PT results are all valid results.

The conclusions from the study, as described in the document posted on the TNI website are as follows:

1. The group of laboratories that participates in PT two or more times each year has consistently lower rates of unacceptable results on PT samples than does the group of laboratories that participates one time each year.
2. The average recoveries in the two groups are similar, although the variation of recovery is lower in the group that does PT twice per year, causing consistently lower average z scores in this group.
3. The groups of laboratories in this study differ in ways other than frequency of PT (for example, applicable standards, quality management, audit frequency), and these differences could contribute to the observed difference in performance. These other factors include an on-site audit for conformance with all NELAC requirements, and the size of the laboratory.

The subcommittee report and data tables are available on the TNI website
http://www.nelac-institute.org/cms/posts/1213773688.php#pab1_6

III. Studies of Performance in different states:

A. Laboratories accredited in Maine and Wisconsin

The subcommittee started investigating the effects of factors other than frequency of PT by looking at performance by laboratories in Maine and Wisconsin, because of unique opportunities to isolate the effects of frequency and regulatory standards. Both studies failed to produce results because the States did not have the information processing ability to summarize results and report in a way that we could analyze. The Maine study required the cooperation of all PT providers, in order to obtain meaningful numbers of results. This was not obtained.

B. Studies of Performance in different states: laboratories from all States

See spreadsheets in attached file: StatebyState.xls. The data file includes the following spreadsheets:

1. All States. The total number of graded results and the percentage of unacceptable results for all laboratories in every State, separated by NELAP accredited or not, and reported separately for WS, WP, and Soil PT, in 2007 and 2008.

2. NELAP %. Data for NELAP States only, with charts of percentage unacceptable in each State. The title of each chart gives the overall US average unacceptable rate for all States.
3. NELAP N. Data for NELAP States, showing the number of graded results, on which the Percentage Unacceptable rates are calculated.
4. Large States %. Percentage unacceptable for the 17 largest states (by number of results in the dataset).
5. Large States N. Number of graded results, on which the percentages are based.
6. State OU. "Over Under" totals for overall percentage unacceptable by matrix and year, counts of numbers of times a State average was over the national average.

B.1 The PT Frequency Subcommittee reviewed data from 3 PT providers that were able to separate out results from NELAP-accredited laboratories, and summarize results from these labs and all other labs, from every State. This was intended to provide insight into effects other than frequency of PT, such as standards used for accreditation or the intensity of oversight of laboratories in the State. However with all laboratories combined , this study could not provide insight into mixed factors, such States where the non-NELAP labs could be certified to two different standards, or where some special groups are required to participate in PT more frequently.

B.2 It is acknowledged by the Subcommittee that there are three significant limitations in the data that must be considered in any interpretation:
a. the results are summarized for all analytes and all laboratories in the three general classes of PT;
b. these data are from just three of the seven providers that offered Chemistry PT in 2007 and 2008. (However, the three providers were all among the largest providers, by numbers of samples, and these data represents a large majority of PT in the Chemical areas.);
c. these data cover only two years, so it is difficult to identify trends.

B.3 The Subcommittee agreed to focus on data from States with large numbers of graded results in all classifications; trends or effects noted in these States could suggest confirmation from smaller States. The states were ranked on the total number of graded results in this database, and examined for sufficient numbers for analysis. The 17 states with the most results were selected, and hereafter called "largest States".

B.4 The graphs that accompany the report show the variability in unacceptable rates in different States and over three general classification of PT (WS, WP, Soil). Unacceptable rates are shown for laboratories accredited to the NELAC standard (by a NELAP State) and laboratories that are not accredited by a NELAP State. See for example the spreadsheet **Large States %** and scroll down to the chart with unacceptable rates for 2007 WS. Note that the NELAC laboratories (Red) have higher unacceptable rates in most states; for example in California (CA) non-NELAC labs had 1.64% unacceptable results and NELAC labs had 2.28% unacceptable. The title states that the national average was 2.20%, so the non-NELAC laboratories were under the national average, and the NELAC labs were over the average. Scroll down to WP 2007 or Soil 2007 and see that the non-NELAP laboratories tend to have higher unacceptable rates. The analysis did not clearly identify any trends in performance other than the evidence of lower rates of unacceptable results in non-NELAP laboratories for WS (Drinking Water) and lower rates of unacceptable results in NELAP laboratories in WP and Soil PT studies.

B.5 The number of graded results in the NELAP and non-NELAP groups is quite variable, between states and within states, and varies by states. See for example the spreadsheet **Large States N** and scroll down to 2007 WS. Note that several states (CA, NC, MI) have large numbers of results from non-NELAP labs, and other states (e.g., FL, TX) have many more NELAP results. Scroll to Soil results and see that there are more results from NELAP

laboratories, with some states having relatively high numbers of non-NELAP results (e.g., CA, NJ) and some having very few (e.g., SC, OR).

B.6 Of the largest States, the frequency of PT is 2 for NELAP laboratories, but is variable for other laboratories. For example, only Florida (FL) and Illinois (IL) require 2 PT for all laboratories. There is no apparent trend to lower rates for non-NELAP laboratories in these two States. This is also not apparent in results from Maine and New York, where 2 PT per year are required for all laboratories.

B.7 The data for laboratories in New Jersey showed similar effects seen in more limited data in the initial study. However this effect is not replicated in other NELAP States. The subcommittee checked the hypothesis that States with reputations for rigorous monitoring of PT, such as New Jersey, would show similar performance. Subcommittee members had different opinions of relative rigor of monitoring of PT in different States, but there was better agreement that California, Oregon and Georgia were especially attentive to PT results. However laboratories in these States do not have consistent patterns of lower rates of unacceptable results.

B.8 The difference between NELAP and non-NELAP laboratories for WS Drinking Water was discussed by the Committee. There were several suggestions for reasons, including the types of laboratories that do this testing exclusively, and the nature of technology used only for drinking water. None of the reasons were thought to be related to the frequency of PT, and so are not immediate concern to this Subcommittee.

B.9 The data allowed assessment of relative rates in different States over all 3 areas of PT, in each of the two years. A simple statistic of "Over Under", with counts of times where the State average was over the national average. Spreadsheet "**State OU**" presents this analysis, including some observations on different patterns. For example, see that in Florida (FL, row 6) for 2007 non-NELAP, laboratories had "0 0 1" indicating that the non-NELAP unacceptable rate was under the national average for WS and WP, but not for Soil. The NELAP labs ("0 1 1") were under the national average for WS, but over the average for WP and Soil. In 2008 the FL labs had patterns of "1 0 1" for non-NELAP and "0 1 1" for NELAP. This showed that FL was over the national average 7 times, and under 5 times, so the State had rates about average in both sets of laboratories. Notice that MI, TX, and WA had similar patterns indicating average relative performance. This showed some States with consistently lower rates (CA, MA), and some where the performance of NELAP and non-NELAP laboratories were significantly different (NJ, OH, PA, TN). Shadings show these groupings. Other shadings show the statistical significance of these patterns, testing various hypotheses. Other shadings on the states show the opinion questions on preferred frequency and (*) whether two PT are required for all laboratories. There was no apparent relationship between the State OU count and current practices or opinions regarding frequency of PT.

B.10 The differences in performance could not be attributed to any one factor, including the following:
1. PT frequency
2. Standards used for non-NELAP laboratories
3. NELAP States vs others
4. Reputation for rigorous follow-up of unacceptable PT results.
5. Expressed preference for frequency

B.11 The subcommittee agrees that the State data are interesting, and worthy of further investigation. However the analysis did not identify any useful conclusions regarding the frequency of PT.

IV. Surveys of Opinions and Costs

A. Summary of a Survey of State Accrediting Authorities

A survey of States was conducted by Judy Morgan (Environmental Science Corp.). It was conducted in two stages that included an e-mail survey and follow-up telephone calls as needed for clarification. Judy accumulated complete survey results from 46 States, including all of the NELAP states.

Survey questions included:

1. How many PT's are required annually by your State for the DW, WW, RCRA, UST or Air program?
2. Do you have other programs (not mentioned above) that require PT's?
3. Does your State support or implement the current NELAP PT policy which requires 2 PTs per year per FoPT?
4. If your State does not currently support or participate in the NELAP program, is the PT frequency requirement the reason?
5. Do you believe that your current PT program structure is effective?
6. If you could recommend changes to the current PT program, what do you think would strengthen it?

Current requirements (46 States):

Field of testing	**Number of States requiring PT**
Drinking water	46
Waste water	29
Solid matrices	19
Underground Storage Tank	17
Air	3
Asbestos	Also mentioned
Lead	Also mentioned

PT Frequency: Of the 46 States that responded, 24 required that their labs participate in PT once per year and 15 states require PT two times per year. Other States said they allowed either 1 or 2, depending on the laboratories' accreditations. Of the 24 States requiring PT once per year, 4 States administer both the State and NELAP requirements.

A total of 46 States responded to the question "Do you believe that 1 PT is sufficient for a PT program or do you believe 2 are more effective?" The response distribution is as follows:

- 19 believe 1 is sufficient (Two NELAP Accreditation Bodies support reducing frequency to one.)
- 15 believe that 2 samples are more effective
- 16 stated that 1 or 2 could be sufficient, depending on performance history (9 of these respondents are NELAP Accreditation Bodies)
- 2 non-NELAP Accreditation Bodies would prefer a greater frequency

Other opinions: Another question received less clear response: "If your State does not currently support or participate in the NELAP program, is the PT requirement the reason?" Several responses included the observation that the cost of PT is only one factor; other factors include the cost to revise systems and to implement all of the other NELAC requirements, including managing receipt and follow-up on more PT results.

Other comments included the need to tighten the acceptance limits for many analytes. There were several suggestions to require 2 successful PTs initially, then one PT/year as long as performance is acceptable. Many States asserted that PT is only one part of an overall competence assessment.

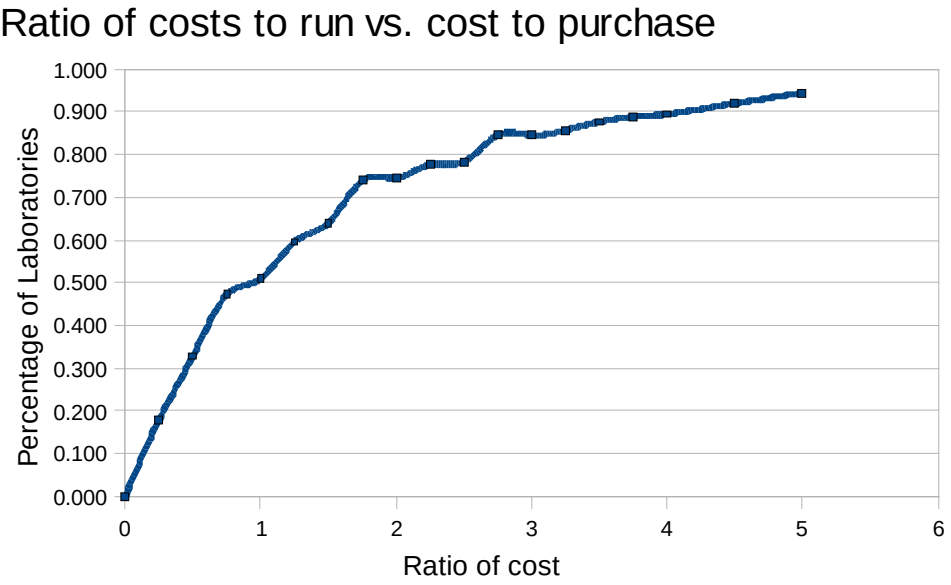
IV. B. Survey of NELAP accredited laboratories regarding Economic Factors of PT

Judy Morgan distributed a web-based survey to all NELAP accredited laboratories, sponsored by ACIL. One part of the survey concerned PT and asked the following (paraphrased from questionnaire):

- Total Direct Cost for purchased PT studies
- Total Direct Testing Cost for processing PT samples
- Six opinion questions regarding value of PT and preference for one or two PT per year.

J

Responses indicate very large variability in costs and in ratios of costs to purchase and cost to run. Ratios of cost to run PT vs. cost to purchase PT ranged from 0.03 to 46 (3% to 4600%). Laboratories could easily calculate the cost to purchase PT, but the costs of processing were not clearly defined, so laboratories had more discretion in estimating this number. The average ratio of cost to run vs. cost to purchase was 2.1 (that is, twice as expensive to run as to purchase), but the median was 0.89, indicating that for 50% of labs the cost to run PT was 90% of the cost to purchase. The middle 50% of 161 responses ranged from 0.33 to 1.82 - or 33% to 182%. The chart below shows the cumulative distribution of this ratio; that is, for any ratio (bottom axis) the percentage of responses with lower numbers is shown on the vertical axis.



The opinion questions are reported below. The tables show the percentages of respondents that agreed with the statement. Table A shows the overall agreement and agreement by laboratory size, as indicated by number of FTE employees. Table B shows the agreement rates by type of laboratory. The standard deviations on the right reflect the extent to which the different groups agree with each other. For example in Table A, the second question has very broad support, consistent across laboratory sizes; the third

question on process improvement due to PT shows substantial differences by size, with the group 51-100 employees much less in agreement.

The following general conclusions can be drawn from the opinion questions:
1. Most laboratories (60-90%) agree that PT provides value and improves processes (Questions 1-3). These percentages are fairly constant across lab size and type.
2. About two thirds of laboratories think that one PT per year is sufficient, and about one third think that two or more are appropriate. This is also generally true across sizes, but by Type, we see that State/Govt. labs and Research/Specialty labs are more likely to prefer 2 PT per year than are other types.
3. Few laboratories think that more PT improves marketability. The largest laboratories and Research and Specialty Laboratories are more likely to agree that more PT improves marketability than are other sizes or types.

Total # of Respondents	510	188	167	78	47	30	
Response Type/Question	All Resp.	5 emp or less	6 - 25 emp	26 - 50 emp	51 - 100 emp	101 - 250 emp	Std. Dev.
The cost of performance testing provides good value for my laboratory.	69.2	74	69.8	67.1	55.3	64.3	**6.4**
My laboratory's quality is enhanced by the use of performance testing.	84.2	84.1	83.6	84.2	87.2	82.8	**1.5**
My laboratory's processes are improved by the use of PT.	71.4	73.9	69.7	76.3	57.4	75.9	**7.0**
One PT per year is sufficient to meet the needs of our quality program.	65.7	70.9	62.3	64	57.4	71.4	**5.3**
Two or more PT samples per year are necessary to meet the needs of our quality program.	37.3	33.7	37.9	39.5	46.8	35.7	**4.5**
More frequent use of PT samples improves the marketability of our lab.	15.3	12.6	11.4	20.6	17.8	27.6	**6.0**

TABLE A and B: Percentage of 510 respondents that agree with statement by Size (A) and Type (B)

Total # of Respondents	236	74	72	55	44	29	
Response Type/Question	Comm.	Water Supply	Muni-cipal	Fed/St./ Local Gov't	Private Ind. Lab	Research & Specialty	Std. Dev.
The cost of performance testing provides good value for my laboratory.	59.3	77.1	77.6	75	83.7	77.8	**8.3**
My laboratory's quality is enhanced by the use of performance testing.	82.4	83.3	85.9	83	95.5	80.8	**5.3**
My laboratory's processes are improved by the use of PT.	70.8	66.2	71.4	64.2	86	81.5	**8.6**
One PT per year is sufficient to meet the needs of our quality program.	62.6	67.1	72.9	63.5	73.8	61.5	**5.3**
Two or more PT samples per year are necessary to meet the needs of our quality program.	36.8	34.3	33.3	45.1	33.3	51.9	**7.7**
More frequent use of PT samples improves the marketability of our lab.	15.4	15.6	15.8	6.9	16.7	20.8	**4.5**

V. Policies of US Federal agencies and opinions of representatives

The Subcommittee has started gathering information from other agencies on their policies for the frequency of PT. To date this includes the Department of Energy (DOE), the US Geologic Survey (USGS), and the Department of Defense. Similar data could be gathered for the Center for Medicare and Medicaid Studies (CMS) and Food and Drug Administration (FDA), but those agencies do not deal directly with environmental testing.

The DOE Office of Health, Safety, and Security has a policy requiring two or more successful PT events annually for radionuclides, stable inorganics, and organics, in four separate matrices - soil, water, vegetation, and air filters. In an undated report titled "U.S. Department of Energy, Office of Health, Safety and Security Analytical Services Program Guidance for Proficiency Testing" that was distributed at the January, 2008 TNI meeting, Mr. George Detsis offered an explanation that "The typical analytical laboratory is too dynamic for just one PT session annually. Staff turnover, loss of key personnel, changes in procedures or instrumentation, etc., can directly impact performance."

The USGS does not have a policy for laboratories in use. Rather, they require site-specific testing on custom PT samples that demonstrate competence for the contaminants of interest for each site. Currently the Water Resources Discipline of USGS relies predominantly on PT samples for laboratory evaluation but the number is not prescribed. Some projects may require quarterly PTs, they could be at random times, or according to a variety of other plans.

The US Department of Defense Quality System Manual (revision 3) makes repeated reference to the NELAC requirements; it requires two successful PT events annually for every field of testing for which PT is available. This Manual is under eminent revision, so the requirements may change, although it is expected that the requirements will be consistent with current TNI standards.

The EPA Office of Water requires one successful PT each year, by method.

VI. Discussion on the economic impact of a change in PT frequency

Purpose: The subcommittee was not clear on what aspects on economic impact should be investigated. They have prepared the discussion below to help the PT Committee provide more detailed direction on the investigations that might be most useful.

There is some general agreement that cost should be broken down to two categories of direct costs and indirect costs. The direct and indirect costs are different for each of the three main stakeholder groups: Laboratories, PT providers, and data users (including customers, regulatory agencies, permit holders, accreditation bodies, etc.)

The review of costs and benefits of proficiency testing needs to consider costs and benefits of changes in policy, which means detailed assessment of the current situation. **Costs** include some that are quantifiable and some that are entirely or partially subjective. **Benefits** are compelling, but almost entirely speculative.

Costs: It is possible to quantify the laboratories' direct costs for the PT samples and the laboratory costs to process the samples. It is more difficult to quantify the laboratories cost for any undeserved evaluations of "unacceptable" on PT samples, which could have costs ranging from $0 to loss of large contracts or failure to win contracts. The cost of supplemental PT could be the only cost, or a small part of the cost of a false unacceptable evaluation. It is even more difficult to quantify the regulators' (or customers') costs of having a laboratory produce inaccurate results.

Similarly, there are substantial costs for accrediting bodies to monitor PT results from their accredited (or certified) laboratories. While it is possible to quantify these costs, it is not easy. Costs include staff time for designing requirements, checking for compliance, and following up on non-compliance. Costs also include building and maintaining the data processing infrastructure to collect, store, and access PT results.

Benefits: It is difficult to quantify any of the benefits of PT, other than the business opportunities that are made available as a result of NELAP or State accreditation, or more recently, access to DoD contracts that will require accreditation to 17025. For the broader population, of laboratories and customers, the benefits are best expressed as the merits that accrue from more accurate testing. Benefits also include detecting systematic errors that would not be found otherwise. Additional benefits accrue to all accredited laboratories and Accrediting Bodies if the credibility of the program is enhanced by more thorough demonstration of performance, or consistency with best international practices.

Impact of change: It was not clear how the PT providers would respond to a change in required frequency. It is difficult to estimate what the actual decrease would be, and therefore to estimate how prices or availability might change. The opinion questions suggest that not all laboratories would reduce their frequency of PT, if they were allowed to by their State.

VII. Related activities in other areas of testing and literature

A. Introduction

There are several sources to look for guidance on the recommended frequency of proficiency testing. Recommendations could be specific for certain testing areas or geographic regions. Recommendations (or requirements) could be in the form of laws, regulations, or professional consensus on best practice.

In most areas of testing and calibration the main focus currently is on increasing the access to proficiency testing and creating standards for the development and operation of PT. The main problem is availability and affordability of PT, not frequency. In some developed economies (Europe, US, Japan, Australia, New Zealand) there is a growing demand for determining optimal PT practices, regarding having PT cover a laboratory's scope, the frequency of PT, and rules for determining acceptable performance.

B. International practices: ILAC, APLAC, EA, Other

Currently there is little international harmony in requirements for participation in PT. There is no consistency on which laboratories are to participate in PT, over what period they need to have PT for their entire scope, the frequency of participation, or rules for determining satisfactory performance. The only general international agreement is at the ILAC level, as expressed in ILAC P9. This requires a laboratory to have acceptable performance in at least one PT prior to gaining accreditation, and at least one PT event each year, with the entire scope covered in four years.

B.1 **ILAC.** There is general agreement in ILAC that the current P9 minimum is not acceptable – in many developed economies and established fields of testing, PT is much more available and common, and these practices need to be addressed. There needs to be allowance for geographic regions and fields of testing and calibration where PT is not available, but there needs also to be encouragement to follow best practices when PT is available. ILAC P9 is currently under revision; there is a rough consensus that the document will give guidance for determining the optimal PT plan for every laboratory individually.

In brief, many ILAC members, particularly those in the APLAC (Asia Pacific) and IAAC (North, Central, and South America) regions support having specific recommendations for PT participation. Many other members, particularly those in Europe, think that the decision on PT participation, including scope and frequency, should be determined by the laboratory and their accrediting body. **All groups agree that frequency should be based on risk (volume of testing, criticality of results), but balanced by previous performance and customers' needs.**

B.2 **APLAC**. The Asia Pacific Laboratory Accreditation Cooperation (APLAC) recently published PT006 to provide "Benchmarks" for frequency of PT and a recommendation for accreditation bodies' policies for participation in PT. This document provides a benchmark of <u>two testing events per year for environmental laboratories</u> (and, for example 2 per year for food testing and 10 per year for medical). PT006 also recommends a policy for accrediting bodies regarding policies for PT. This is reproduced below; the document is available on request.

Guidance for Accreditation Body PT Policy

The accreditation body should ensure [their accredited laboratories'] participation when PT is available, appropriate and at a frequency that reflects best practices and/or best local norms. When determining the appropriate frequency of PT, the accreditation body should take into account the test range, method capabilities and regulatory limits, where available.

Furthermore, accreditation bodies should have a policy to update their specific program requirements as PT schemes become available. This policy should promote PT participation requirements that are consistent with best practices for laboratories in that economy or, if

possible, with the APLAC benchmark frequencies in that field of testing or calibration, even if these exceed the minimum prescribed by ILAC.

B.3 **EA**. The European Cooperation on Accreditation (EA) and the European Standards Organisation (CEN) are developing a policy for PT. The current proposal has been in development for several years and represents a consensus of current EA opinion. The document presents guidance for determining a PT plan based on the laboratory's scope, their customers' needs, and the availability and applicability of PT. The proposal is for a laboratory to develop a plan and then to have approval by the accrediting body. There will of course be field-specific norms and specific legal requirements within some countries, especially for medical laboratories, drinking water laboratories, and environmental testing laboratories. The current draft of this document is available on request.

B.4 **IUPAC:** International Union of Pure and Applied Chemistry recently published a revision of their Harmonized Protocol for Proficiency Testing. This document represents the consensus of IUPAC's Analytical Chemistry Division (which includes chemical environmental testing). One relevant section from the protocol concerns the frequency of PT (Frequency of "Rounds"), reproduced below:

Section 3.10 Frequency of Rounds
The appropriate distribution frequency is a balance between a number of factors of which the most important are
• the difficulty of executing effective analytical QC;
• the laboratory throughput of test samples;
• the consistency of the results in the particular field of work covered by the scheme;
• the cost/benefit of the scheme;
• the availability of CRMs in the analytical sector; and
• the rate of change of analytical requirements, methodology, instrumentation, and staff in the sector of interest.

Objective evidence about the influence of round frequency on the efficacy of proficiency testing is very sparse. Only one reliable study on frequency has been reported, and that showed (in a particular scheme) that changing the round frequency from three to six per year had no significant effect (beneficial or otherwise) on the participants' performance.

In practice, the frequency will probably fall between once every two weeks and once every four months. A frequency greater than once every two weeks could lead to problems in the turn-around time of test samples and results. It might also encourage the belief that the proficiency testing scheme can be used as a substitute for IQC, an idea that is strongly to be discouraged. <u>If the period between distributions extends much beyond four months, there will be unacceptable delays in identifying and correcting analytical problems, and the impact of the scheme on the participants could be small. There is little practical value, in routine analytical work, in proficiency tests undertaken much less than twice a year.</u> (underline added)

B.5 **ILAC.** Within ILAC currently it is likely that P9 will present general guidance based on the European recommendation, balanced with availability and local convention within that field of testing. The document will then present different ways to meet the general recommendations, such as the APLAC benchmarks or related efforts, or more detailed guidance on self-determination.

VII. C Published / presented studies

There are several published studies demonstrating the benefits of experience with PT, as defined by the ability to perform well in PT. In general (as in other settings) if a lab takes more PT, they have better results on PT. This is generally true regarding number of years of experience with PT and the number of PT events per year.

VII. D Referenced documents:

DOE paper: U.S. Department of Energy, Office of Health, Safety and Security
Analytical Services Program Guidance for Proficiency Testing (available on request)
APLAC PT 006: http://www.aplac.org/documents/pt/aplac_pt_006_issue_1.pdf
EEE-PT (draft EA policy?) on PT plans (available on request)
ILAC P9 (draft revision) (available on request)
A2LA policy:
http://www.a2la.org/requirements/A2LA_General_Requirements_for_Proficiency_Testing.pdf
CAEAL Study (Middlebrook) http://www.cala.ca/perfacred-2004.pdf

VIII. Conclusions:

1. The data reviewed by the subcommittee showed that laboratories that take more PT can have generally better performance as measured by unacceptable rates and consistency of recovery (SD between labs). However, performance differences could be caused by a variety of factors other than the frequency of PT, including laboratories' practices, and in how these data were assembled for this analysis. The subcommittee's analyses did not confirm the cause of the performance differences.

2. The data reviewed by the subcommittee did not confirm any proposed theory for isolated causes of performance differences. The subcommittee agreed that the effects of differences between States' policies could be complicated, and they may not be apparent in overall summary data.

3. State AB's have various opinions, experiences, and recommendations regarding the frequency of PT. For most states this is a consideration in their decision whether to join NELAP or not, and whether to remain in NELAP or not. However frequency of PT was not the only issue of general concern, since the TNI process involves more than a difference in the frequency of PT. Some states appreciate the rigor of the TNI requirements, some may see it as a barrier.

4. An authoritative association of scientists (IUPAC) has recommended that PT should occur at least twice a year in all areas of analytical chemistry. An international cooperation of accreditation bodies (APLAC) has concurred, for the environmental area among others.

5. Other US Federal agencies (specifically, DOE and DOD) and stakeholders recommend PT of two or more times a year for many circumstances.

6. External publications consistently show that more experience with PT is related to better performance on PT studies.

7. The economic impact of a reduction is not clear with available information. Costs of regulatory oversight were not compiled. Benefits are more difficult to quantify – they include detection of unexpected error, assisting with internal quality improvement activities, and providing evidence of competence to customers and oversight bodies.

8. Efforts are underway internationally to provide guidelines for defining a PT plan where frequency is based on general principles and laboratories' individual competences, demand, and risk.

9. Based on the available information collected by the TNI PT Frequency Subcommittee, the recommendation to the TNI PT Expert Committee is that there is not compelling evidence to support changing the current requirement for frequency of PT in the TNI standard.