

NATIONAL REFERENCE LABORATORY
Department of Health, East Avenue Medical Center
Tel. No./Fax No.: 435-71-36; E-mail: nrlcamcdoh@yahoo.com

PHILIPPINE INSTITUTE OF PURE AND APPLIED CHEMISTRY
Ateneo de Manila University Campus, Loyola Heights, Quezon City
Tel. 426-6072 Fax. 426-6073 E-mail: info@pipac.com.ph

APPENDIX I

PROFICIENCY TESTING PROGRAM FOR DRUG SCREENING LABORATORY

FINAL REPORT: PT CYCLE 2004-01 Pilot Screening Test for Met and THC-COOH in urine

1. **The participating laboratories** in the pilot run of the Proficiency Testing program were randomly selected from the list of all screening drug testing laboratories (SDTL) within Metro Manila and all government laboratories nationwide. To keep confidentiality of all results, all laboratories were each given a unique Laboratory Code which was used throughout the PT cycle up to the statistical scoring of the results.
2. **The test samples.** Each participating drug testing laboratory was given a set of ten (10) test samples. The samples were synthetic urine formulations with the compositions shown in Table 1 below. Every vial in the set is randomly selected and coded from each formulation using a randomization protocol.
 - a. *Concentration.* The formulations were spiked with standard reference grade MET or THC-COOH with a certified purity of at least 99 %. The prepared concentrations were confirmed by solid phase extraction/GC-MS analysis.
 - b. *Homogeneity.* Ten random samples from each formulation were tested for homogeneity using a randomly selected screening test kit. Four tests per vial were done. Thus, a total of $10 \times 10 \times 40 = 400$ tests were done for the homogeneity testing of the formulations that were used in the PT cycle. The test results were analyzed using a statistical test, which confirmed sufficient homogeneity of the prepared formulations suitable for proficiency testing.
 - c. *Stability.* Random samples per formulation were also set aside for stability tests by screening and SPE/GC-MS analyses. The results show no significant change in concentration levels within the two-week period of the PT cycle.

Table 1. Test sample formulations (the samples distributed to the laboratories are randomly coded and do not follow the order given below).

Formulation Code	As-prepared Concentration of analyte (ng/mL) ^a	
	MET (presence)	THC-COOH (presence)
A	1,500 (+)	0 (-)
B	0 (-)	88 (+)
C	1,500 (+)	88 (+)
D	500 (-) ^b	0 (-)
E	0 (-)	25 (-) ^b
F	500 (-) ^b	25 (-) ^b
G	0 (-)	0 (-)
H	1,500 (+)	0 (-)
I	0 (-)	88 (+)
J	0 (-)	0 (-)

^aThe as-prepared concentration values are traceable to the standard reference used and for spiked samples the concentrations were subsequently confirmed by SPE/GC-MS analyses. The qualitative test result is identified as presence (+) or absence (-) of the drug based on the cutoff level.

^bThis concentration level is below the screening cutoff of 1000 ng MET/mL for the test kits used by all laboratories in this PT cycle, and below the cutoff of 50 ng THC-COOH/mL.

3. **Scoring of the results of participating laboratories.** In this pilot run, the results submitted on-line by majority of the laboratories as well as data from hardcopy submissions of some of the laboratories were included in the statistical. However, in future PT cycles, only on-line data will be accepted for statistical evaluation, and thus, those that do not meet the on-line submission requirement will be considered as non-submission of results.

The performance of the laboratory is scored using an acceptability criterion for a number of incorrect responses for the set of 10 tests for a particular drug. The acceptability criterion was set by a scoring system using the

probability for a correct result for a given test sample formulation. The overall performance was taken as the combined score for the two drugs tested in this PT cycle. This scoring system is explained below.

The scoring system. Each test result from the laboratory is assigned a quality score q using the formula:

$$q = I \times P \times 3 \quad (1)$$

where I is the indicator value of 0 or 1 for correct and incorrect result, respectively; P is a probability factor, and 3 is a scaling factor. The nature of these factors is explained below.

Figure 1 shows the plot of the average Proportion of Correct results (in percent) versus the formulation concentration relative to the cutoff level, C/C_{cutoff} . The ideal result is 100 %. Noteworthy observations are: (a) the laboratories had no problem detecting a true blank as evidenced by the almost 100 % correct results for such a test sample, and (b) the proportion of correct results is less than 100 % for spiked samples, which is lower when the concentration level is closer to the cutoff (compare, for example, the result for positive samples of THC-COOH and MET). The variation of the performance of the laboratories across different formulations is a manifestation of the inherent concentration-dependence of the screening test method near the cutoff levels. There is higher success rate for the blank sample compared with the spiked samples.

The values of the probability factor P are taken to be equal to the proportion of correct answers. Thus, P represents the probability that a single test trial will yield a correct outcome. The values of P are not identical for all test trials because of the inherent uncertainty of the screening test method at analyte concentrations near the cutoff level. This implies that for samples whose concentrations are way below (e.g., a blank) or way above the cutoff (the “reliable” region), the absence or presence of the analyte is more clearly defined compared with those which are near the cutoff level (the uncertain or “unreliable” region). Therefore, the factor P compensates for this effect so that the “penalty” for an incorrect result depends on the “certainty” of the test sample formulation. A lower P is used for an incorrect answer for a concentration within the unreliable region of the test method, and a full $P (= 1)$ is given to those that are in the reliable region. Here, the values of P for the test samples are consensus values because they are taken from the average performance of the laboratories in this PT cycle (Table 2).

The factor 3 is a scaling factor that converts the score to a *pseudo-z-score*. The z -score is the common method of scoring quantitative results and correspond to the acceptability criteria shown in Table 3.

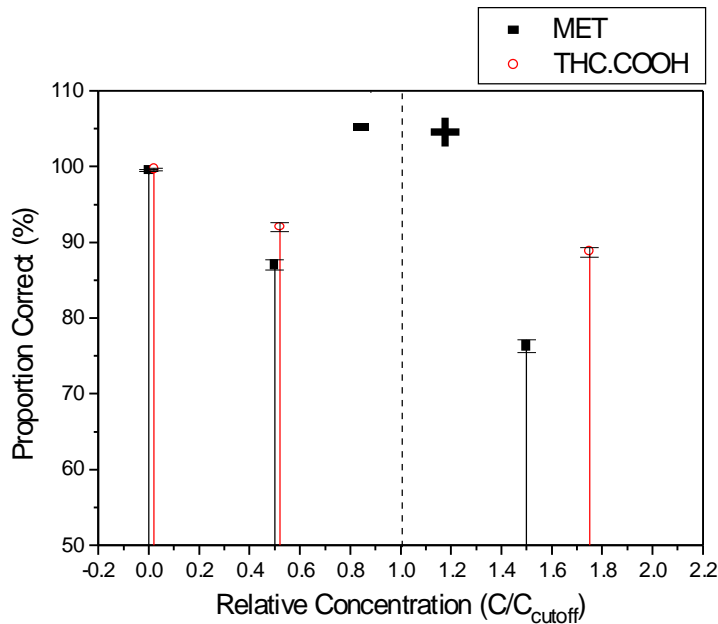


Figure 1. Plot of the average performance of the laboratories (97) in this PT cycle versus concentration level of analyte in the test sample. The error bars are the expanded uncertainty of the averages at 95 % level of confidence.

Table 2. *P* and *q*-scores derived from this PT cycle. U_q is the expanded uncertainty for the *q* value at 95 % level of confidence.

Relative Concentration	MET			THC-COOH		
	<i>P</i>	<i>q</i>	U_q	<i>P</i>	<i>q</i>	U_q
0.0	0.996	2.99	0.004	0.996	2.99	0.004
0.50	0.870	2.61	0.020	0.920	2.76	0.020
1.5	0.763	2.29	0.025	na	na	na
1.75	na	na	na	0.887	2.66	0.019

na = not applicable, not tested in this cycle

Table 3. Classification of score

Score	Classification	Performance Evaluation
0	Excellent	Pass
$0 < score \leq 2$	Acceptable	Pass
$2 < score < 3$	Questionable	Conditional Pass*
$score \geq 3$	Fail	Fail

*A laboratory that obtains Conditional Pass performances in two consecutive PT cycles will be given a Performance Evaluation of a Fail in the second instance, and will be subject to the same regulatory requirements as a laboratory that obtains a Fail in a PT cycle in the first instance.

The same criteria were used for the *q*-score. The assessment of the performance of the laboratory in testing for a particular drug analyte uses all ten results per analyte by combining the *q* scores to get the re-scaled sum of *q* scores (RSQ):

$$RSQ = \frac{\sum_{i=1}^n q_i}{\sqrt{n}} = \frac{q_1 + q_2 + \dots + q_{10}}{\sqrt{10}} \quad (2)$$

This method of combining scores is again akin to the method of combining *z*-scores used in quantitative PT results. The acceptability criteria for RSQ will be the same as described in Table 3.

The overall performance of the laboratory in the PT was assessed by combining the scores for MET and THC-COOH as follows:

$$RSQ_{combined} = \frac{RSQ_{met} + RSQ_{thc}}{\sqrt{2}} \quad (3)$$

The combined overall score will also follow the acceptability criteria described in Table 3. In this PT cycle, the values of the factors and the corresponding *q*-scores are summarized in Table 2.

4. Overall PT Results. A summary of the performance of the laboratories is plotted in chart form shown in Figure A.

In summary, 3 laboratories had an overall combined score greater than 3 and therefore were considered to Fail the PT. These laboratories performed poorly in the test for MET and THC-COOH, either failing or getting questionable performances for the two analytes. Overall, there were 12 laboratories with questionable performance. The rest were either excellent (46) or acceptable (36).

For the analysis of THC-COOH only, one laboratory had a Fail and 3 had questionable performances; these labs failed to detect positive samples of THC-COOH. For the analysis of MET, one lab also failed but 17 had questionable ratings; again, these are the labs that failed to detect a positive sample of MET.

Based on a chi-square statistical test, there was no relationship between the kit brand used and the results of this Proficiency Test Cycle.

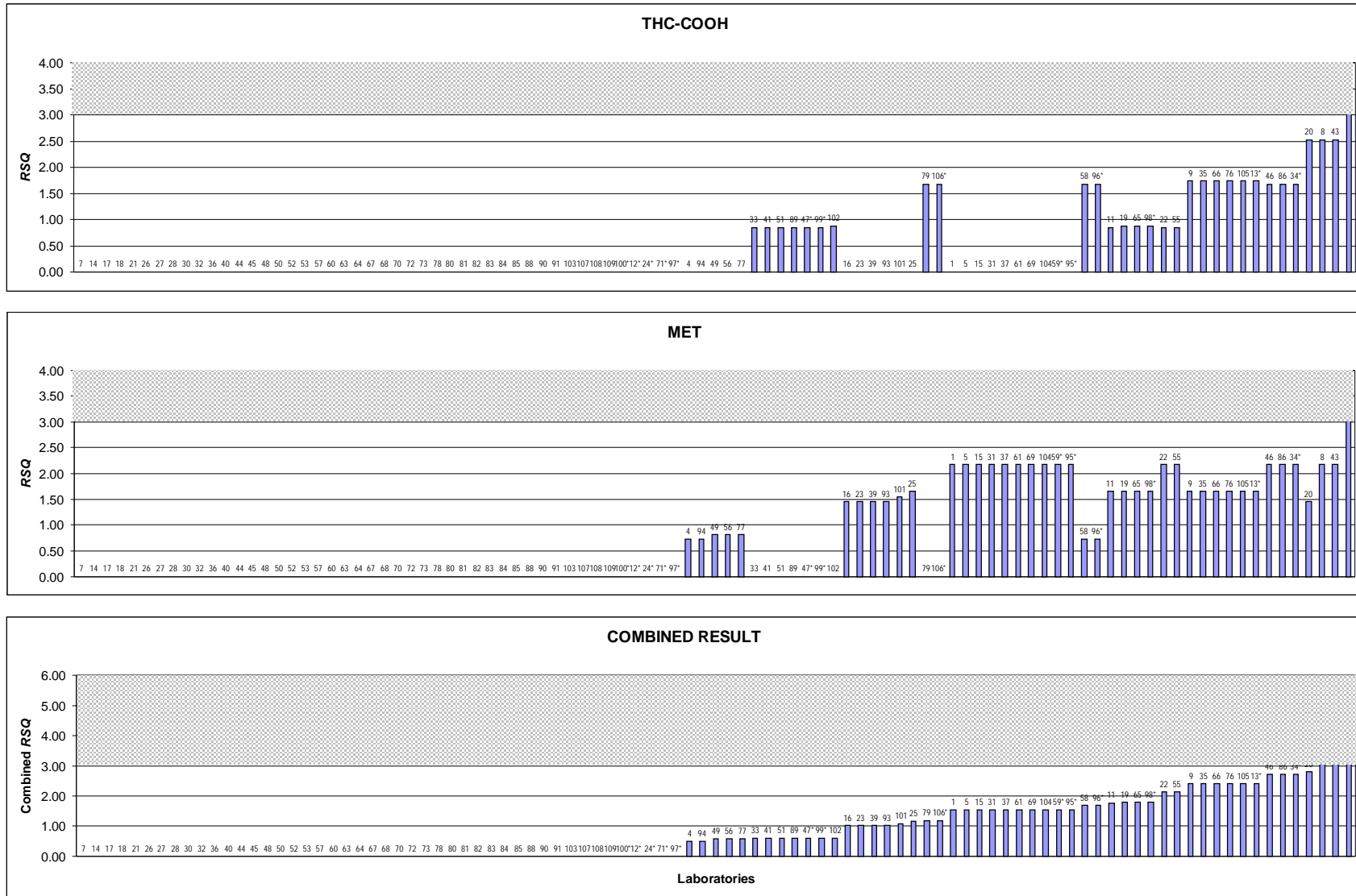


Figure A. Histogram of q -scores for THC-COOH, MET, and combined results.