APLAC

Asia Pacific Laboratory Accreditation Cooperation

Final Report

APLAC T046

FOOD MICROBIOLOGICAL PROFICIENCY TESTING PROGRAM

December 2008

ACKNOWLEDGEMENTS

CNAS wishes to acknowledge gratefully the financial support provided by the APLAC secretariat and express our sincere thanks to the APLAC and EA members for their participation and cooperation. We would also like to appreciate Mr. Cao Zhijun, the PT manager, for coordinating this program and acting as technical adviser.

Also thanks to Dr. Lu Xingan and Dr. Liu Shuyan of Liaoning Entry-Exit Inspection & Quarantine Bureau of P. R. China for collaborating the supply, packaging, pre-testing of the samples, the data analysis, preparing the final report.

Contents

1. Foreword

This report summarizes the results of APLAC Food Microbiological Proficiency Testing Program T046 for testing laboratories organized under the auspices of the Asia Pacific Laboratory Accreditation Cooperation (APLAC). China National Accreditation Service for Conformity Assessment (CNAS) coordinated the program. Liaoning Entry-Exit Inspection and Quarantine Bureau of P. R. China collaborated the program. The program covered the microbiology analysis of four samples —Aerobic plate count, *Listeria monocytogenes*, & *Vibrio parahaemolyticus.*

This program applies only to the use of inter-laboratory comparisons of the purpose of proficiency testing. The aim of the program is to determine the performance of individual laboratories for specific tests. It does not include determining the effectiveness and precision of test methods.

2. Features of the Program

a) A total of 31 laboratories from 9 economies participated in the program. APLAC, EA (European Cooperation for Accreditation) members and some other laboratory accreditation bodies were invited to participate. Following are participation details of the participants representing economies, accreditation bodies and the number of laboratories:

Economy	**Accreditation Body**	**No. of Labs**
Australia	National Association of testing Authorities (NATA)	3
Canada	Canadian Association for Laboratory Accreditation (CALA)	4
Canada	Standards Council of Canada (SCC)	4
Hong Kong, China	Hong Kong Accreditation Service	2
India	National Accreditation Board for Testing and Calibration Laboratories (NABL)	4
Malaysia	Standards Malaysia (SM)	4
New Zealand	International Accreditation New Zealand (IANZ)	4
P. R. China	China National Accreditation Service for Conformity Assessment (CNAS)	4
Singapore	Singapore Accreditation Council (SAC)	2

b) Each of the participants was randomly allocated a unique code (between 1 and 31). 5 laboratories gave up the chance, and 2 laboratories didn't return any results. All reference to the participants in this report is via the code they were allocated, to assure confidential treatment of results.

c) Four food samples containing freeze-dried bacteria and cryoprotectants in small glass vials were distributed to each participant. They were labeled Sample PT A, PT B, PT C & PT D respectively. Participants were requested to determine:

PT A & B are tested for Aerobic plate count.

PT C is for determination of the presence or absence of *Listeria monocytogenes*;

PT D is for determination of the presence or absence of *Vibrio parahaemolyticus*;

For each test item, a single result was reported by participants—along with the test methods used. These are presented in Appendix A, together with calculated z-scores, summary statistics and graphical presentations of the data. (Appendix A)

d) Prior to distribution, the samples were tested for homogeneity and stability. Based on the results of the preliminary testing, it was concluded that the samples within each group were sufficiently homogeneous, and were stable for a limited period at ambient temperature without detriment to the microbiological content. (See Appendix B).

e) Laboratories were requested to perform the tests according to "Instructions to Participants" provided, and to record the results on the accompanying Results Sheet, both of which were distributed to participants with their samples. (Appendix C)

3. Format of Appendices

Appendix A

a) A table of results (*Aerobic plate count*) along with robust

statistical parameters, and rough estimations of uncertainty of the assigned values (consensus values).

b) Robust z-score bar chart for the test of Aerobic plate count.

c) A table of the test of *Listeria monocytogenes* (Sample PT C) results and method used along with their corresponding assigned values (positive or negative).

d) A table of the test of *Vibrio parahaemolyticus* results (Sample PT D) and methods used along with their corresponding assigned values (positive or negative).

Appendix B contains details of the samples used in the program--including sample source, preparation, homogeneity testing and stability testing.

Appendix C contains copies of Instructions to Participants, Receipt Form, Results sheet, and Testing Record Form.

Appendix D contains details of statistical procedures, calculations and formulae

4. Description of Samples

The bacteria strains having been used in the PT samples are below.

Sample	Assign value	Strain	Strain number
PT A		*Bacillus cereus* *Serratia marcescens* *Enterobacter aerogenes*	CMCC(B)63305 CMCC(B)41002 CMCC(B)45103
PT B		*Bacillus cereus* *Serratia marcescens* *Citrobacter freundii* *Enterobacter aerogenes*	CMCC(B)63305 CMCC(B)41002 CMCC(B)48016 CMCC(B)45103
PT C	Positive	*Listeria monocytogenes* *Listeria innocua* *Listeria ivanovii* *Citrobacter freundii* *Klebsiella pneumoniae* *Proteus mirabilis* *Enterobacter aerogenes* *Bacillus cereus*	CMCC(B)54003 ATCC 33090 ATCC 19119 CMCC(B)48016 CMCC(B)46114 CMCC(B)49003 CMCC(B)45103 CMCC(B)63305

Sample	Assign value	Strain	Strain number
PT C	Negative	*Listeria innocua* *Listeria ivanovii* *Citrobacter freundii* *Klebsiella pneumoniae* *Proteus mirabilis* *Enterobacter aerogenes* *Bacillus cereus*	ATCC 33090 ATCC 19119 CMCC(B)48016 CMCC(B)46114 CMCC(B)49003 CMCC(B)45103 CMCC(B)63305
PT D	Positive	*Vibrio parahaemolyticus* *Vibrio alginolyticus* *Vibrio metschnikovii*	CMCC(B)86-87 AS1.1833 CMCC(B)91-9
PT D	Negative	*Vibrio vulnificus* *Vibrio mimicus*	CMCC(B)92-3 CMCC(B)91-8

Four food samples consist of freeze-dried bacteria and cryoprotectants in small glass vials. Freeze-dried cryoprotectants used: Trehalose, Proline, Glutamine sodium and Sucrose.

For PT C and PT D, the inoculum level of the target bacteria is about 40～100cfu/check sample in the presence of high levels of competitive flora (10^4 cfu or so).

5. Statistical Design of the Program

- ***Aerobic plate count:***

Each participant was provided with two different but similar level samples PT A and PT B (split-level design -- slightly different between the pair of samples) Though the medians were the same, there were difference in the averages (Means). To statistically assess the participants' results, we used robust statistical procedures to generate the z-scores and summary statistics for each sample result—number of results, median, normalized interquartile range. We also calculated the between-laboratories z-score and within-laboratories z-scores for the pairs of results, based on the standardized sum and difference of the paired results of PT A & PT B. The table below shows the summary statistics and the parameters (The data has been transformed its $\log_{10}$ cfu/mL for statistical calculations).

Table 1 summaries the statistics calculated. Table 2 details the parameters used

to calculate the z-scores.

Table 1 Statistical Summary of Aerobic plate count in PT A & PT B

Sample	Number of results	Median (In $\log_{10}$ form)	Median (cfu/ml)	NIQR	Robust CV
PT A	**25**	**4.176**	**15000**	**0.175**	**4.2%**
PT B	**25**	**4.176**	**15000**	**0.180**	**4.3%**

Table 2 Parameters Used to Calculate between labs and within lab z-scores

Sample	Number of Results	Normalized Sum		Normalized difference	
		Median	NIQR	Median	NIQR
PT A & PT B	25	5.885	0.261	0.013	0.045

- ***Listeria monocytogenes:***

The sample PT C is tested for *Listeria monocytogenes*, the evaluation of the results is based on comparison to the corresponding known values. Samples with and without *Listeria monocytogenes* were randomly assigned to participants.

- ***Vibrio parahaemolyticus:***

The sample PT D is used for determining the presence or absence of *Vibrio parahaemolyticus* we deal with the results compared with the assigned values (known values). Samples with (positive) and without (negative) *Vibrio parahaemolyticus* were randomly assigned to participants.

6. Extreme Results

❶ Aerobic plate count in PT A & PT B

In order to achieve the program's aim of assessing laboratories' performance, a robust statistical method, which uses z-scores to assess participant's performance, has been utilized. The z-score is a measure of how far the result is from the consensus value—a normalized value, which gives a "score" to each result relative to the other results in the group.

Therefore a z-score close to zero means that the result agrees well with those from other laboratories. An outlier will be any result(s) which has an absolute z-score value greater than or equal to three. (i.e. z-score$\leqslant$-3 or z-score$\geqslant$3). Four kinds of z-scores—robust z-score of PT A, robust z-score of PT B, between laboratories z-score and within laboratories z-score are calculated and shown in

Appendix A. For further details on the calculation and interpretation of the robust z-score and the relative statistics, please see Appendix E.

Table 5 details the outliers and the laboratories' code numbers.

Table 5 Outliers on Aerobic plate count

z-score	**Outlier**
Between-laboratories	T046-003, T046-006
Within-laboratories	T046-004, T046-007
PT A	T046-003, T046-006, T046-007
PT B	T046-003, T046-006

Of the 24 participating laboratories, 4 (16.7%) laboratories reported one or more unsatisfactory results.

❷ *Listeria monocytogenes* in PT C

Comparing the laboratories' results with the assigned values, we use false positive or false negative to assess laboratories' testing performance. A total of 18 laboratories participated in this test, 4 (22%) laboratories reported false results, 2 laboratories' results were false positive, 2 were false negative.

Table 4 False results of *Listeria monocytogenes* in PT C

False Results (*L.monocytogenes*)	False Positive Results	False Negative Results
Laboratory's code	T046-025, T046-28	T046-024, T046-026

❸ *Vibrio parahaemolyticus* in PT D

We compared the laboratories' results with the assigned values to reach conclusions: if the results did not agree with the assigned value, we used false positive or false negative to assess laboratories' testing performance. A total of 14 laboratories participated in this test, 5 (35.7%) laboratories reported false results, 1 laboratories' result was false positive, 4 were false negative.

Table 3 False results of *Vibrio parahaemolyticus* in PT A

False Results (*Vibrio parahaemolyticus*)	False Positive Results	False Negative Results
Laboratory's code	T046-009	T046-014, T046-025, T046-028, T046-030

7. Technical Comments

● ***Aerobic plate count*** **in PT A & PT B**

Testing technical aspect

The predominant target bacteria, *Bacillus cereus*, which is chosen in PT A and PT B, is fit for proficiency testing because of the following two reasons: (a) it survives easily at adverse environmental conditions; (b) it is well known that there is no significant difference under the incubation temperature 22～37℃ on different non-selective media during total Aerobic plate count. However, this strain grows quickly and can easily form spreading colonies. At the beginning of the testing, the participants should do an overall estimate of the *Aerobic plate count* and have some techniques to prevent unsatisfactory results from happening, such as blank control, actions to prevent spreading of colonies, daily observation of the plates etc.

Overall performance (24 participating laboratories) here was very good with a total of 4 (16.7%) laboratories having reported one or more outliers. The count plates with spreading colonies might be the main reason.

● ***Listeria monocytogenes*** **in PT C**

A total of 18 laboratories participated in this test, 4 (22%) laboratories reported the false results, 2 laboratories' results were false positive, 2 were false negative.

False positive:

Some participants might lack the necessary experience to identify *Listeria spp.* Following are some ways to differentiate a few confused strains.

- *L.innocua* can be differentiated from *L.monocytogenes* by its lack of hemolysis on the blood agar plates and negative reaction in the CAMP test.
- *L.monocytogenes*, *L.ivanovii* produce hemolysis in horse or sheep blood agar and are also positive in the CAMP test. *L.monocytogenes* cannot utilize

xylose, but is rhamnose-positive. *L.ivanovii* shows enhanced hemolysis in the area of the *R.equi* streak.

- The serological testing may be useful for identifying *Listeria spp*.

False negative:

In samples containing mixed populations of *Listeria spp.*, *L.innocua* can easily overgrow *L.monocytogenes* as enrichment progresses and confound isolation from Fraser Broth. If the suspect Blood agar colony or colonies have been generated from follow-up of *Listeria* selective media plate, there is a high probability of confirmation compared to follow-up of *Listeria spp.* from extended enrichment in Fraser Broth. For the above reason, it is advised to extend the pre-enrichment period of the Fraser broth if a suspect colony was found on selective /blood agar upon primary isolation. Besides, this target *L.monocytogenes* strain has less haemolytic activity than normal and produce a small narrow zone of β-hemolysis.

- ***Vibrio parahaemolyticus* in PT D**

Except for *Vibrio parahaemolyticus*, *Vibrio alginolyticus* and *Vibrio metschnikovii* with sucrolytic activities (yellow colonies appeared on the TCBS agar plates) were added to the positive fish powder samples; But in the negative samples, *Vibrio vulnificus* and *Vibrio mimicus were added to them as the background bacteria*. The two strains can not hydrolyze sucrose which are similar with *Vibrio parahaemolyticus.* Here the factors influenced the results were analyzed.

- Addition of the others *Vibrio spp.*. In a way, it increased the difficulties of isolated and identified of *Vibrio parahaemolyticus.*
- The culture media and reagents. The culture media from different factories have different isolating effect and there are obviously dissimilarity of the different batches of the cultrue media from the same factory. For example, the growth of *Vibro spp*. were various on the TCBS agar plates of the different factories.
- Detection apparatus. Due to the special biochemical charateres, some

apparatuses can not identify well and truly *Vibro spp.*,but the test results of some paticipants depended mainly on such machines.

- ❖ The validation of the technology guidline of reagents and culture media. Some paticipants depended mainly on the experience but not validated the capability of the used culture media and chemical reagents by the standard strains, especially, to some key reagents and culture media, such as TCBS agar plates, reagents of Indol test and amidinohydrolase and so on.
- ❖ Detection methods. The biochemical reactions of the different standard detection methods of *Vibrio parahaemolyticus* disagree.

Meanwhile we think that the animal experiment, O/129 test and serological test would be helpful to identification of the *Vibrio parahaemolyticus*; Also some accessorial methods might redound to determine *Vibrio parahaemolyticus*, such as ATB、VITEK、BIOLOG、API、PCR、Real-time PCR、SENSITITRE、Chomagar、HS and so on.

8. Definition of Statistical Terms

Robust statistics Statistics which minimize the effect of extreme results, for example the median is robust, as opposed to the mean which is easily influenced by extreme values

Median The middle value or center of a data set calculated form the ordered values.

Normalized IQR The interquartile range (IQR) is the difference between the upper (Q_3) and lower (Q_1) quartiles, which are the values above and below which a quarter of the values lie. It is a measure of the spread of results calculated by multiplying the IQR by a factor (0.7413), which relates it back to the “normal distribution”.

Robust CV The robust coefficient of variation (CV) is the normalized IQR divided by the median and expressed as a percentage. This is

comparable to the "classical" CV, which is the standard deviation divided by the mean.

Z-score A normalized value which assigns a "score" to the result(s), relative to the other numbers in the group—e.g. (result-median) / normalized IQR

Extreme result A result which is significantly different from the assigned value—may be an outlier, gross error, mis-calculation, etc.

Outlier A result which has been identified as significantly different by statistical techniques, e.g. absolute ***z-score*** greater than or equal to three for testing programs, or a false positive/false negative result.

False positive erroneously reporting the presence of a parameter (e.g. analyze, organism) which is absent from the sample.

False negative failing to report the presence of a parameter (e.g. analyze, organism) which is present in the sample.

Assigned value value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose

Consensus value an assigned value obtained from the results submitted by participants (e.g. for most testing programs the median is used as the assigned value)

Normalized Sum the sum of the pair of results divided by $\sqrt{2}$.

Normalized Difference the difference between the pair of results divided by $\sqrt{2}$.

Between-labs z-score this is based on the standardized sum of the pair of results reported by the laboratory. The z-score is calculated as: = (Normalized Sum - Median) /Normalized IQR

Within-lab z-score This is based on the standardized difference of the

pair of results reported by the laboratory. The z-score is calculated as: = (Normalized Difference - Median) /Normalized IQR

Please refer to Appendix D for further details of statistical procedures, calculations and formulate.

APPENDIX A

Summary of Results

A.1 *Aerobic plate count* in PT A & PT B

Each of these tables contains the results returned by each laboratory, including the code number and the method used, four kinds of z-score—robust z-score of PT A, robust z-score of PT B, between laboratories z-score and within laboratories z-score for pair of results are calculated. Please see Appendix E for details on how these z-scores are calculated and interpretation of the robust z-score and the relative statistics.

A.1.1 Table of Results & z-score of Aerobic plate count

Lab Code	PT A (cfu/mL)	PT B (cfu/mL)	Method used	Transformed Results (log_{10} cfu/mL)		z-score			
				PT A	PT B	PT A	PT B	Between-labs	Within-lab
T046-001	11000	9800	MFHPB-18	4.04	3.99	-0.77	-1.03	-0.78	-1.09
T046-002	15000	14000	NZTMZ Method 43.1	4.18	4.15	0.00	-0.17	0.00	-0.77
T046-003	75000	60000	ISO 4833:2003	4.88	4.78	3.99	3.34	3.61	-1.84
T046-004	32000	21000	APHA(2001)6.331	4.51	4.32	1.88	0.81	1.37	-3.20
T046-006	130000	140000	FDA/BAM Chapter 3	5.11	5.15	5.35	5.38	5.25	0.22
T046-007	3000	12000	AOAC 990.12	3.48	4.08	-3.99	-0.54	-2.08	9.27
T046-008	5300	7000	Australian standard 1766.2.1	3.72	3.85	-2.58	-1.84	-2.04	1.62
T046-009	13350	11250	AOAC 15th	4.13	4.05	-0.29	-0.69	-0.39	-1.48
T046-011	18000	19000	AS 5013.1-2004	4.26	4.28	0.45	0.57	0.57	0.08
T046-013	6400	4900	MFHPB-33	3.81	3.69	-2.11	-2.70	-2.24	-2.14
T046-014	11400	11900	MFHPB-18	4.06	4.08	-0.68	-0.56	-0.51	0.00
T046-015	14000	13000	MFHPB-18/HPB	4.15	4.11	-0.17	-0.34	-0.17	-0.81
T046-019	32000	34000	MFHPB-33	4.51	4.53	1.88	1.97	1.94	0.12
T046-020	12000	13000	FDA/BAM Chapter 3	4.08	4.11	-0.55	-0.34	-0.35	0.26
T046-021	15000	17000	FDA/BAM Chapter 3	4.18	4.23	0.00	0.30	0.23	0.57
T046-021	39000	36000	AOAC 986.33	4.59	4.56	2.37	2.11	2.24	-0.85
T046-022	16000	16000	FDA-BAM Chapter 3	4.20	4.20	0.16	0.16	0.23	-0.30
T046-024	10000	15000	USFDA-BAM 8TH EDN 2001	4.00	4.18	-1.00	0.00	-0.40	2.50
T046-025	42500	30500	AOAC 986.33	4.63	4.48	2.58	1.71	2.14	-2.58
T046-026	11600	12600	IS:5402-2002	4.06	4.10	-0.64	-0.42	-0.43	0.27
T046-027	13409	14182	IS:5402-2002/ISO-4833-1991	4.13	4.15	-0.28	-0.14	-0.12	0.09
T046-028	12000	9800	GB/T4789.2-2003	4.08	3.99	-0.55	-1.03	-0.68	-1.69
T046-029	20000	23000	GB/T4789.2-2003	4.30	4.36	0.71	1.03	0.92	0.67
T046-030	16000	16000	GB/T4789.2-2003	4.20	4.20	0.16	0.16	0.23	-0.30
T046-031	15000	16000	GB/T4789.2-2003	4.18	4.20	0.00	0.16	0.16	0.15

			Normalized Sum	Normalized Difference
No of results	25	25	25	
Median	4.176	4.176	5.885	0.013
NIQR	0.175	0.180	0.261	0.045
Median (cfu/mL)	15000	15000		
Robust CV	4.2%	4.3%		

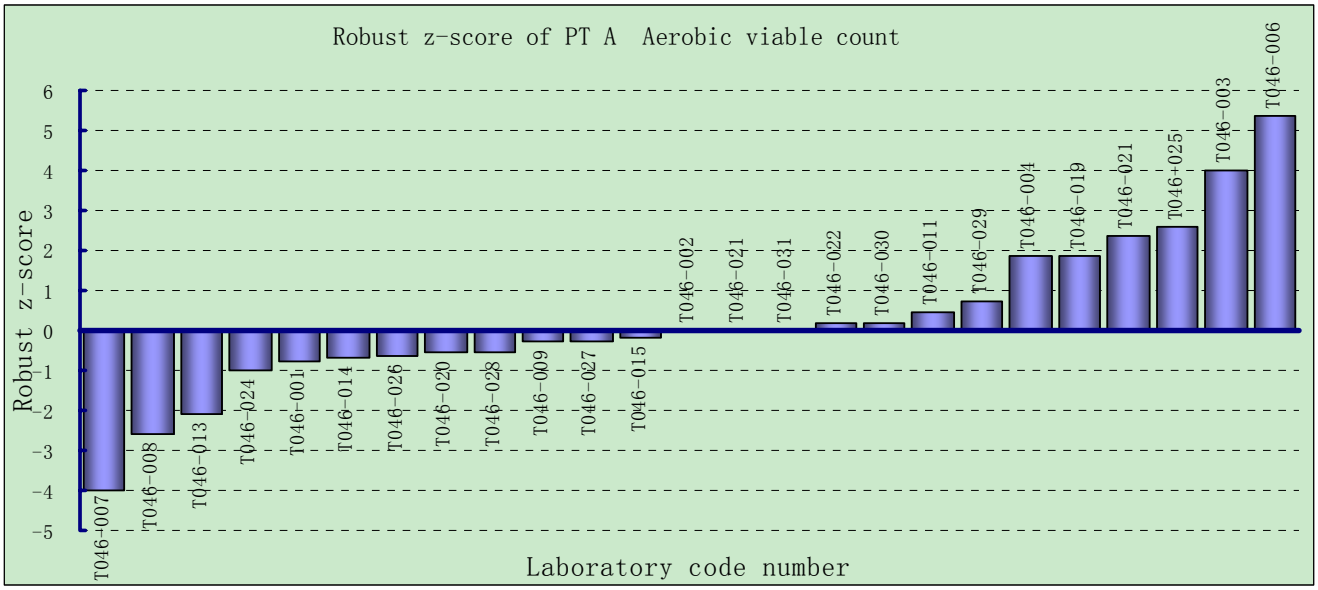
Note: § RED denotes an outlier (i.e. |z| ≥3), YELLOW denotes questionable result.

A.2 Ordered z-score bar-charts

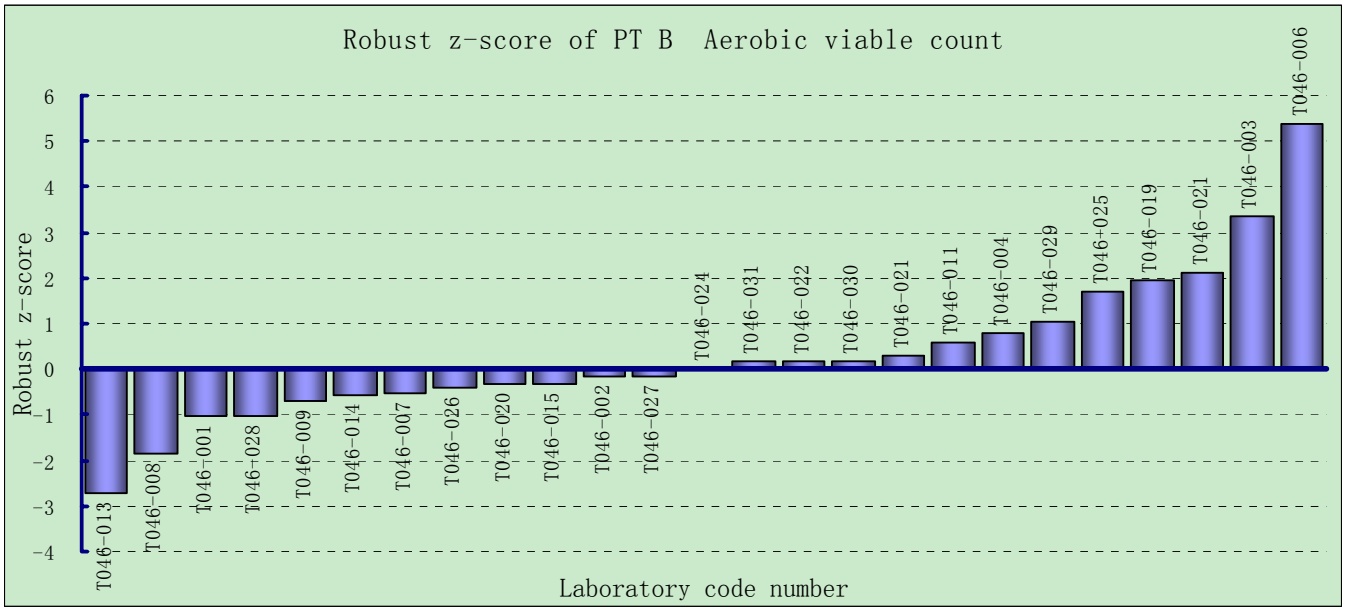
On these charts each laboratory's z-score is shown, in order of magnitude, and is marked with its code number. From these each laboratory can readily compare its performance relative to the other laboratories.

These charts contain solid lines at 3 and –3, so the outliers are clearly identifiable as the laboratories whose bar extends beyond those "cut-off" lines. The y-axis of these charts has been limited to range from –5 to 5, so in some cases very large or small (negative) z-scores appear as extending beyond the limit of the chart.

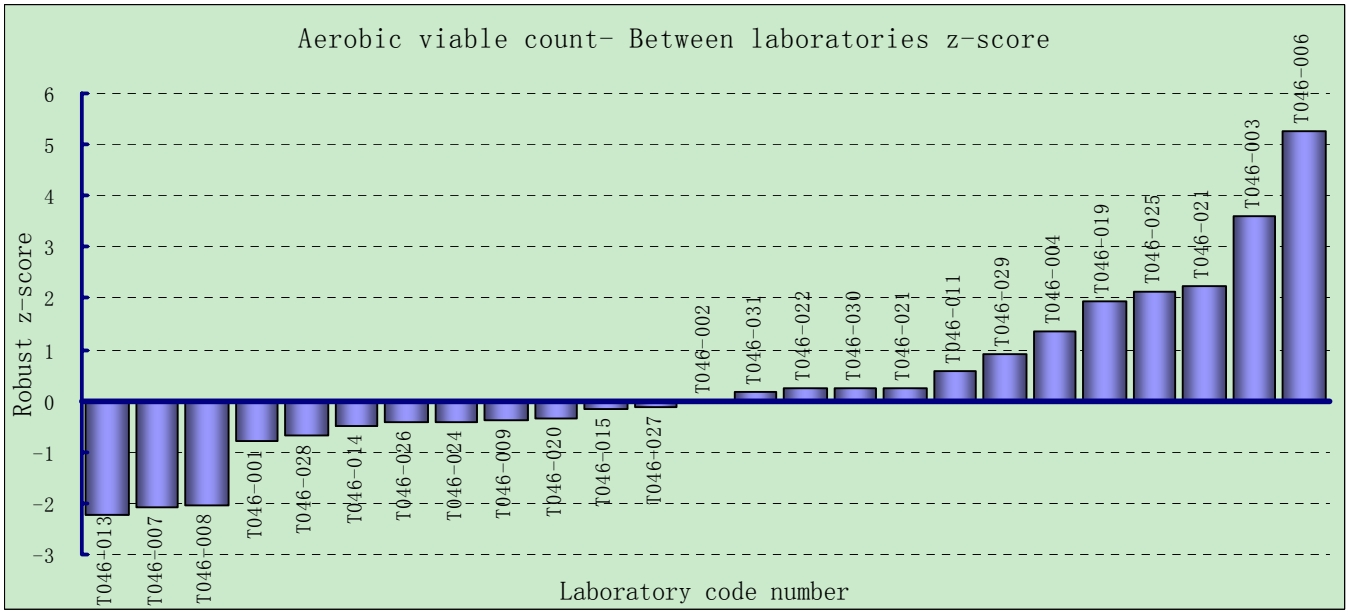
A.2.1 Ordered robust z-score of PT A bar-chart



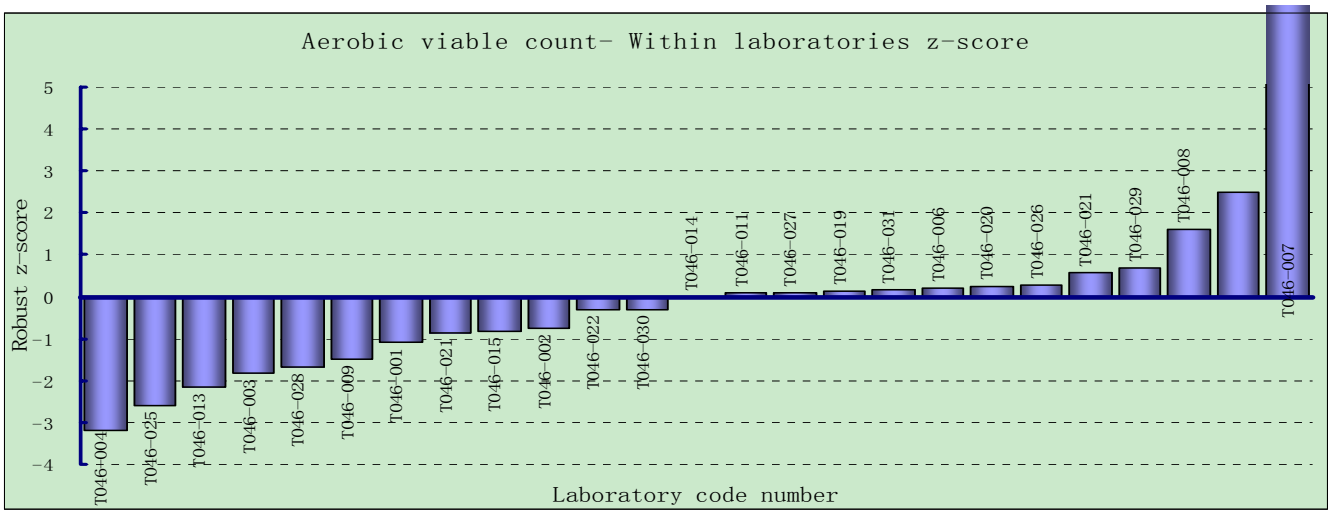
A.2.2 Ordered robust z-score of PT B bar-chart



A.2.3 Ordered between labs z-score bar-chart



A.2.4 Ordered within lab z-score bar-chart



A.3 *Listeria monocytogenes* in PT C

A.3.1 Table of *Listeria monocytogenes* results and the evaluation

Lab code	Assigned value	Lab result	Method used	Evaluation
T046-001	Positive	Positive	MFHPB-30	satisfactory
T046-002	Positive	Positive	FDA BAM 8th	satisfactory
T046-003	Positive	Positive	ISO:11290-1;1996	satisfactory
T046-004	Positive	Positive	FDA 8th Chapter 10	satisfactory
T046-006	Positive	Positive	FDA 8th Chapter 10	satisfactory
T046-011	Positive	Positive	AS/NZS 1766.2.15-1998	satisfactory
T046-013	Negative	Negative	MFHPB-30	satisfactory
T046-014	Negative	Negative	MFHPB-30	satisfactory
T046-015	Positive	Positive	MFHPB-30/HPB	satisfactory
T046-019	Negative	Negative	MFHPB-30	satisfactory
T046-022	Positive	Positive	FDA BAM	satisfactory
T046-024	Positive	Negative	USFDA-BAM 8TH EDN 2001	unsatisfactory
T046-025	Negative	Positive	BAM CHAPTER 10	unsatisfactory
T046-026	Positive	Negative	BAM CHAPTER 10	unsatisfactory
T046-027	Negative	Negative	IS:14988,p-1-2001/ISO:11290-1;1996	satisfactory
T046-028	Negative	Positive	GB?T 4789.30-2003	unsatisfactory
T046-029	Negative	Negative	GB/T 4789.30-2003	satisfactory
T046-030	Negative	Negative	GB/T 4789.30-2003	satisfactory
T046-031	Positive	Positive	GB/T 4789.30-2003	satisfactory

A.4 *Vibrio parahaemolyticus* in PT D

A.4.1 Table of *Vibrio parahaemolyticus* results and the evaluation

Lab code	Assigned value	Lab result	Method used	Evaluation
T046-002	Positive	Positive	FDA BAM 8th	satisfactory
T046-006	Positive	Positive	FDA BAM 8 Chapter-9	satisfactory
T046-008	Negative	Negative	Australian standard 1766.2.9	satisfactory
T046-009	Negative	Positive	FDA BAM 8 Chapter-9	unsatisfactory
T046-011	Positive	Positive	AS 1766.2.9-1997	satisfactory
T046-014	Positive	Negative	US FDA BAM-9	unsatisfactory
T046-024	Positive	Positive	USFDA-BAM 8TH EDN 2001	satisfactory
T046-025	Positive	Negative	BAM 8 Chapter-9	unsatisfactory
T046-026	Negative	Negative	BAM Chapter-9	satisfactory
T046-027	Negative	Negative	IS:5887-1976/ISO:6579-1993	satisfactory
T046-028	Positive	Negative	GB/T 4789.7-2003	unsatisfactory
T046-029	Positive	Positive	GB/T 4789.7-2003	satisfactory
T046-030	Positive	Negative	GB/T 4789.7-2003	unsatisfactory
T046-031	Positive	Positive	GB/T 4789.7-2003	satisfactory

APPENDIX B

Samples Preparation

And

Homogeneity & Stability Testing

B.1 Preparation of Sample

Test samples were prepared by Food Testing Laboratory of Liaoning Entry-Exit Inspection and Quarantine Bureau which was the contracted collaborator of CNAS. The laboratory, which has been accredited according to ISO/IEC 17025:2005, met with the Management System Requirements (including quality management system, document control, internal audit, corrective action) and the Technical Requirements (including management, staffing and training, choice of method or procedure, data analysis and interpretation of scheme results, confidentiality etc). It also has been assessed and approved by CNAS to comply with the relevant requirement of ILAC-G13 Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes, applicable to the characterization, homogeneity and stability testing of proficiency testing test materials.

The collaborator refined the food microbiology proficiency testing process by using lyophilized microorganisms derived from traceable reference cultures (CMCC Chinese Medical Cultures Center and ATCC). Among lots of strains, the strains with high survival rate and relative long storage period after being freeze-dried were chosen for this program. According to the different characteristics of reference cultures, the collaborator used different cryoprotectants (e.g. trehalose, glutamine and sucrose) and froze at different rates.

The samples consisted of freeze-dried bacteria and cryoprotectants in small glass vials.

B.2 Homogeneity Testing

The test materials were packaged in their final forms and samples were selected randomly for homogeneity testing and stability testing.

***Aerobic plate count* in PT A & PT B**

For each gross sample, 16 samples were randomly selected and tested. Each sample was divided into two test portions, all test portions were tested in random

order on the repeatability condition, that is to say, the testing was carried out during a short period by one person using the same testing method and apparatus in one laboratory. The results on each sample are listed in Table B.1~B.2.

The between-samples standard deviations (s_s) were calculated using the following formulae, where g is the number of samples used:

The sample average is $\bar{x}_t = (x_{t1} + x_{t2})/2$

The between-test-portion range is $w_t = | x_{t1} - x_{t2} |$

General average is $\bar{\bar{x}} = \sum \bar{x}_t / g$

The standard deviation of general average is $s_x = \sqrt{\sum (\bar{x}_t - \bar{\bar{x}})^2 / (g-1)}$

The within-sample standard deviation is $s_w = \sqrt{\sum w_t^2 / 2g}$

The between-samples standard deviation is $s_s = \sqrt{s_x^2 - s_w^2 / 2}$

Table B.1 Homogeneity Testing Results of Sample PTA

Sample	x_{t1}	x_{t2}	$\log(x_{t1})$	$\log(x_{t2})$	$\bar{x}_t$	w_t	w_t^2
1	11000	12000	4.04	4.08	4.060	0.040	0.001600
2	12000	13000	4.08	4.11	4.095	0.030	0.000900
3	10000	10000	4.00	4.00	4.000	0.000	0.000000
4	12000	11000	4.08	4.04	4.060	0.040	0.001600
5	12000	11000	4.08	4.04	4.060	0.040	0.001600
6	11000	10000	4.04	4.00	4.020	0.040	0.001600
7	12000	12000	4.08	4.08	4.080	0.000	0.000000
8	14000	14000	4.15	4.15	4.150	0.000	0.000000
9	15000	15000	4.18	4.18	4.180	0.000	0.000000
10	11000	14000	4.04	4.15	4.095	0.110	0.012100
11	13000	12000	4.11	4.08	4.095	0.030	0.000900
12	11000	15000	4.04	4.18	4.110	0.140	0.019600
13	11000	12000	4.04	4.08	4.060	0.040	0.001600
14	15000	15000	4.18	4.18	4.180	0.000	0.000000
15	13000	11000	4.11	4.04	4.075	0.070	0.004900
16	14000	13000	4.15	4.11	4.130	0.040	0.001600
$\bar{\bar{x}}$ =4.09cfu/mL, Median=4.176cfu/mL							
s_x =0.051cfu/mL, s_w =0.039cfu/mL, s_s=0.043cfu/mL, NIQR=0.175cfu/mL,							
s_s/NIQR=0.24<0.3							

Table B.2 Homogeneity Testing Results of Sample PTB

Sample	x_{t1}	x_{t2}	$\log(x_{t1})$	$\log(x_{t2})$	$\bar{x}_t$	w_t	w_t^2
1	14000	14000	4.15	4.15	4.148	0.004	0.000015
2	14000	15000	4.15	4.18	4.163	0.026	0.000681
3	16000	16000	4.20	4.20	4.202	0.004	0.000017
4	16000	15000	4.20	4.18	4.188	0.024	0.000572
5	18000	17000	4.26	4.23	4.245	0.030	0.000873
6	19000	18000	4.28	4.26	4.268	0.025	0.000611
7	18000	16000	4.26	4.20	4.232	0.056	0.003123
8	14000	14000	4.15	4.15	4.148	0.004	0.000015
9	19000	15000	4.28	4.18	4.228	0.104	0.010797
10	17000	14000	4.23	4.15	4.188	0.084	0.007035
11	15000	18000	4.18	4.26	4.218	0.075	0.005666
12	21000	15000	4.32	4.18	4.248	0.144	0.020710
13	17000	14000	4.23	4.15	4.188	0.084	0.007035
14	16000	15000	4.20	4.18	4.188	0.024	0.000572
15	13000	13000	4.11	4.11	4.112	0.004	0.000016
16	18000	16000	4.26	4.20	4.232	0.056	0.003123
$\bar{\bar{x}}$=4.20cfu/mL, Median=4.176cfu/mL							
s_x=0.043cfu/mL, s_w=0.044cfu/mL, s_s=0.029cfu/mL, NIQR=0.180cfu/mL,							
s_s/NIQR=0.16<0.3							

Comparing the between-samples standard deviation (s_s) with the corresponding standard deviation for proficiency assessment (NIQR in this program), all s_s are less than 0.3 NIQR, it indicates that the variation between samples could be neglected in subsequently statistical analysis and proficiency assessments, in another word, the samples are homogenous enough for the PT purposes.

Listeria monocytogenes in PT C

One set of samples was positive, and the other was negative. 12 samples were randomly selected for homogeneity testing, the result showed all positive and negative samples conformed to their assigned values.

Vibrio parahaemolyticus in PT D

One set of samples was positive, and the other was negative. 12 samples were randomly selected for homogeneity testing, the results showed all positive and negative samples conformed to their assigned values.

B.3 Stability Test

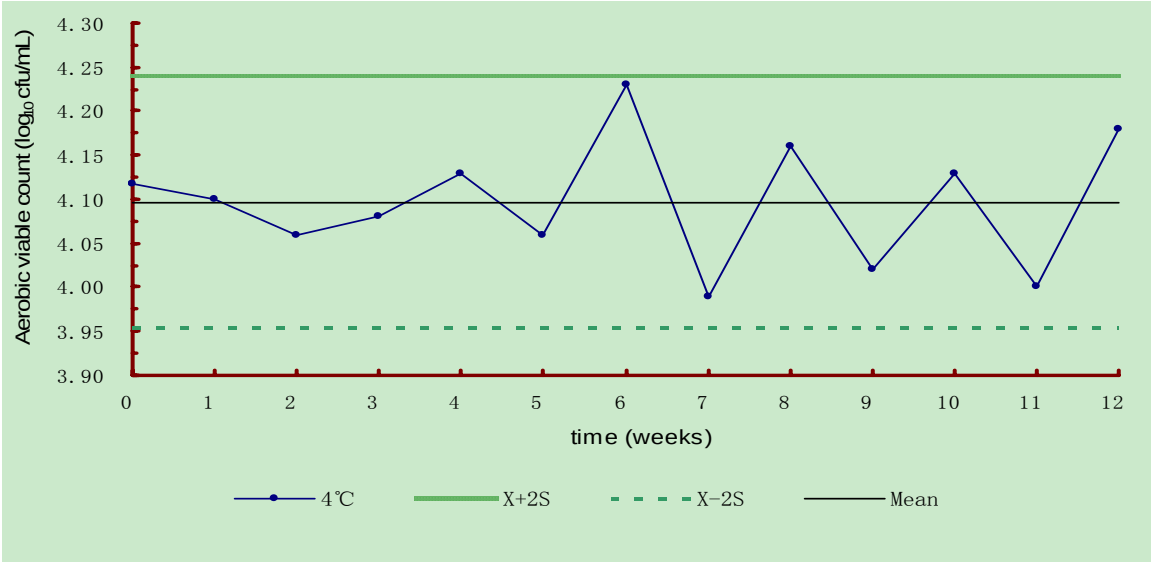
We use TNT Express to delivery the samples, normally taking about 7 days to arrive at participant laboratories. The samples undergo three phases, first in the origin country (our country) for waiting delivery, the samples were persevered at 4℃; the second is in the course of delivery aircraft, the inner of aircraft is about 24℃; the third phase is at the location country, it depends the local temperature. Base on the transport condition, we designed the stability testing as described below.

Aerobic plate count in PT A & PT B

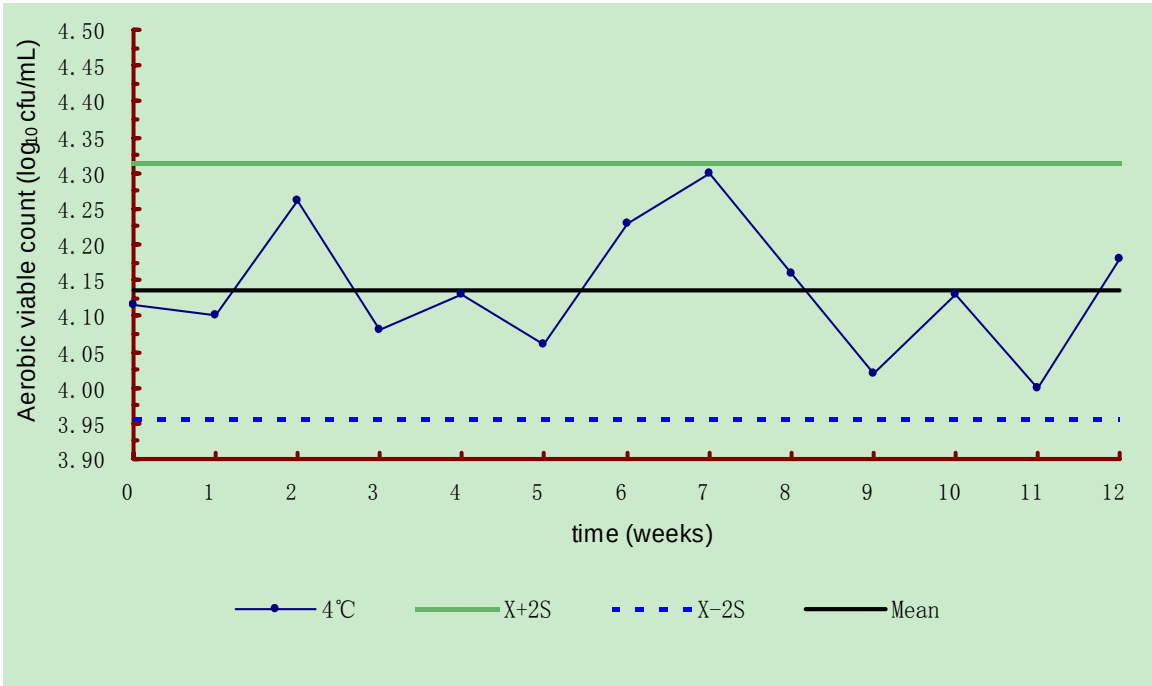
Bacillus cereus was chosen in PT A and PT B as it easily survives adverse environmental conditions, fit for proficiency test.

Two types of stability tests were done, a stability test at a storage temperature (4℃) and a stability test at higher temperatures simulating transport conditions. For testing the stability of the sample stored at 4℃, 12 ampoules were examined at regular time interval, total 12 weeks. The stability of the samples at higher temperatures was determined at three different temperatures. The temperatures tested were 20℃, 37℃and 45℃. Once three days the samples from each storage temperature were examined respectively; To 20℃ over a period of 42 days, 14 samples were examined; To 37℃ over a period of 30 days, 10 samples were examined; To 45℃ over a period of 21 days, 7 samples were examined. The counts obtained for each storage temperature were logarithm transformed. The following figures show that there is not significant different at various higher temperature up to 37℃ for 30 days (or up to 45℃ for 21 days) to simulate transport conditions.

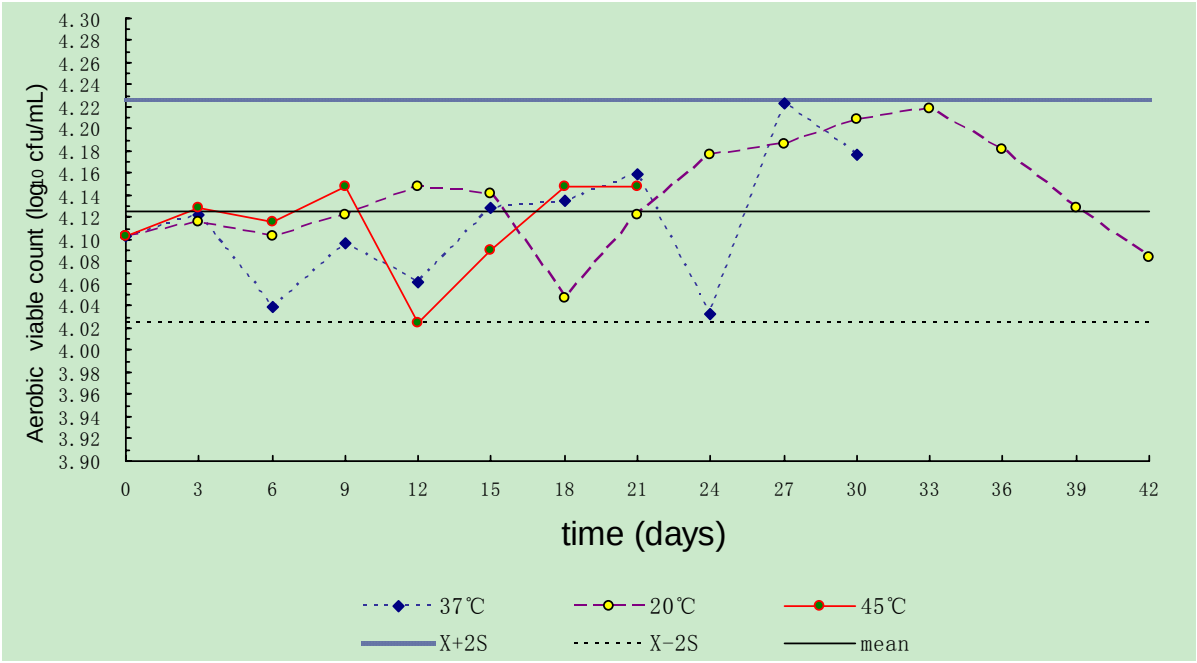
B.3.1 Figure of Results of stability testing of PT A Aerobic plate count at 4℃



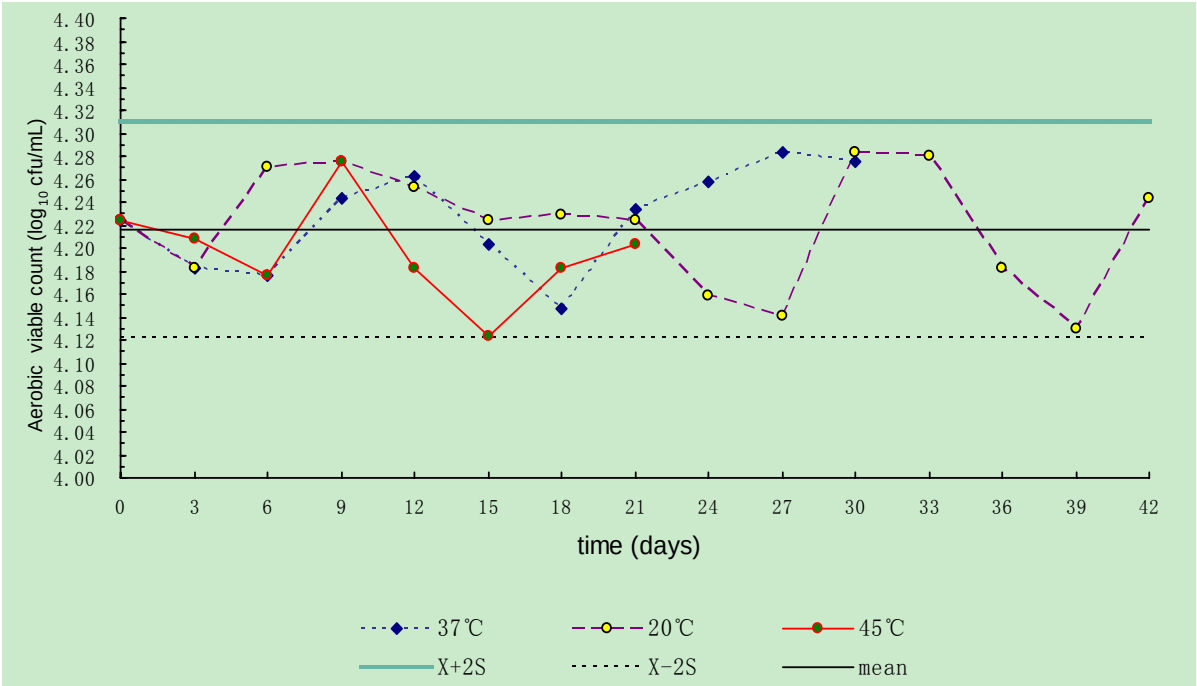
B.3.2 Figure of Results of stability testing of PT B Aerobic plate count at 4℃



B.3.3 Figure of Results of stability testing of PT A Aerobic plate count at various high temperatures over a period days



B.3.4 Figure of Results of stability testing of PT B Aerobic plate count at various high temperatures over a period days



- ***Listeria monocytogenes* in PT C & *Vibrio parahaemolyticus* in PT D**

Two types of stability tests were done, a stability test at a storage temperature (4℃) and a stability test at higher temperatures simulating transport conditions. For testing the stability of the sample stored at 4℃, 10 samples were examined at regular time interval, total 60 days. The stability of the samples at higher temperatures was determined at three different temperatures. The temperatures tested were 20℃, 37℃ and 45℃. To 20℃, once three days over a period of 30 days, ten samples were examined. To 37℃ and 45℃, once a day over a period of 12 days, twelve samples from each storage temperature were examined. The results are below.

B.3.5 Stability testing of *Listeria monocytogenes* and *Vibrio parahaemolyticus,*

Testing Item	Sample	temperature	survival time limit	methods used
Listeria monocytogenes	PT C (Positive)	4℃	>60 days	AFNOR BIO 12/3-03/96 VIDAS *Listeria monocytogenes* kit used on the mini VIDAS immunoanalyzer USDA/FSIS Microbiology Laboratory Guidebook 3rd ed/1998
		20℃	>30 days	
		37℃	11 days	
		45℃	7 days	
Listeria spp.	PT C (Negative)	4℃	>60 days	CCFRA Microbiological Methods Manual Method 11.1:1995, 11.2:1995 & 11.2:1995 (VIDAS LIS)
		20℃	>30 days	
		37℃	11 days	
		45℃	7 days	
Vibrio parahaemolyticus	PT D (Positive)	4℃	>60 days	NMKL NO.22 FDA/BAM CHAPTER 9
		20℃	>30 days	
		37℃	7 days	
		45℃	3 days	
Vibrio parahaemolyticus	PT D (Negative)	4℃	>60 days	NMKL NO.22 FDA/BAM CHAPTER 9
		20℃	>30 days	
		37℃	7 days	
		45℃	3 days	

APPENDIX C

Instructions to Participants

and

Results Sheet

C.1

INSTRUCTIONS TO PARTICIPANTS

:

Your laboratory has been nominated by the laboratory accreditation organization in your economy to participate in the Asia Pacific Laboratory Accreditation Cooperation (APLAC) Food Microbiological Proficiency Testing Program T046. Your laboratory's code number is T046-xxx. All references to your laboratory in reports associated with this program will be with this code number.
I have sent you **four** samples in small glass vials labeled PT Sample A, B C and Sample D.
Enclosed with the samples are the following documents (total **7** pages):

- Instructions to Participants
- Receipt Form
- Results sheet
- Testing Record Form

To ensure that results from this program can be analyzed properly, please read these instruction carefully, both upon arrival of samples and before starting the analysis.

- Four simulated food samples in the form of freeze-dried bacteria in small glass vials are distributed to your laboratory.
- Samples were sent by TNT at ambient temperature and the laboratories are instructed to store them at -18℃ immediately upon arrival.
- Once you receive the samples, you will need to complete RECEIPT FORM **in print** and E-mail the completed form to the coordinator on **the same day** as you receive the samples.

- Testing requested:

Name of sample	Test item	Matrix	PT type
PT sample A	Aerobic Plate Count	Milk powder	Quantitative scheme
PT sample B	Aerobic Plate Count	Milk powder	Quantitative scheme
PT sample C	*Listeria monocytogenes*	Milk powder	Qualitative scheme
PT sample D	*Vibrio parahaemolyticus*	Fish powder	Qualitative scheme

- In the trials all samples should be examined during a specified two-weeks period. At the time of examination the freeze-dried micro-organisms in each glass vial were reconstituted in 26 mL of sterile dilution liquid. Be sure to mix the content in the glass vial thoroughly before taking out the analytical portion. This liquid was treated as the sample for testing. The laboratories were asked to examine the samples by their routine procedures or the methods from international or national standards and record the method on the results sheet.
- Please **E-mail** or **post** the completed results sheet (including the Testing Record Form) **NO LATER THAN May 8 2008 to**:

Dr. Liu Shuyan
Technology Center
Liaoning Entry-Exit Inspection & Quarantine Bureau of P.R.China
No.21 Hutan Beigou, Zhongshan District, Dalian, 116015, P.R. China
E-mail: liusy@lnciq.gov.cn; liushuyan05@163.com; luxa@online.ln.cn
Tel: +86-411-82867706 Fax: +86 411 82867690

Good luck with the work!
If you feel that a delay in a postal return means that the deadline may pass, please inform as soon as possible and state the reason.

C.2

RECEIPT FORM

We are pleased to announce that in accordance with the distribution schedule we have dispatched the samples on April 3, 2008 through TNT, along with Instruction, Receipt of the Samples, Result Sheet.

In order to monitor the samples status (Samples as distributed are stable for a limited period at ambient temperature), we kindly ask each participating laboratory on receipt of the samples to fill out this RECEIPT FORM and return copy of this data sheet by **email URGENTLY to:**

Attn: Dr. Liu Shuyan
Fax: +86 411 82867690
E-mail: liusy@lnciq.gov.cn; liushuyan05@163.com; luxa@online.ln.cn

Thank you in advance for your cooperation.

The APLAC T046 samples were received on:_____*(date)*

NAME OF LABORATORY:Cawtron InstituteCODE NO.T046-002

COUNTRY:_____ CONTACT PERSON: _____

LOCAL TEMPERATURE:_____~_____℉（_____~_____℃）

E-MAIL ADDRESS:_____

LIST ALL ITEMS RECEIVED:_____

_____

_____

Inspect the packaging. Was there any visible damage to the artifacts? Yes___No___

If yes, describe problem below, fax and e-mail the form to the Above Attention.

_____

_____

_____

C.3

Results Sheet (1)

Lab Code T046-xxx

DATE SAMPLE RECEIVED: _____

DATE SAMPLE PROCESSED: _____

REPORT DATE:_____

PT Sample A	PT Sample B
Aerobic plate count:_____	Aerobic plate count:_____
NOTE:	NOTE:

Method:

Signed: Date:

Results Sheet (2)

Lab Code T046-xxx

DATE SAMPLE RECEIVED: _____

DATE SAMPLE PROCESSED: _____

REPORT DATE:_____

PT Sample C

***Listeria monocytogenes*:**_____
(POSITIVE or NEGATIVE)

NOTE:

Method:

Results Sheet (3)

Lab Code T046-xxx

DATE SAMPLE RECEIVED: _____

DATE SAMPLE PROCESSED: _____

REPORT DATE:_____

PT Sample D

Vibrio parahaemolyticus:_____
(POSITIVE or NEGATIVE)

NOTE:

Method:

Please E-mail the completed results sheet (including the Testing Record Form) NO LATER THAN May 8 2008 to the coordinator:

Dr. Liu Shuyan
Fax: +86 411 82867690
E-mail: liusy@lnciq.gov.cn liushuyan05@163.com; luxa@online.ln.cn

C.4

Testing Record Form

(If necessary, another sheet may be added to this Results Form)

Laboratory I.D. #: _____Sample label:_____
Organism(s) Isolated: Genus _____
Species (optional) _____

Isolation Media:

Name_____ Manufacturer _____Catalog# _____
Name_____ Manufacturer _____Catalog# _____
Name_____ Manufacturer _____Catalog# _____
Did you add antibiotics?____Yes___No
If yes, please specify: _____

Identification Media:

Name_____ Manufacturer _____Catalog# _____
Name_____ Manufacturer _____Catalog# _____
Name_____ Manufacturer _____Catalog# _____
Did you add antibiotics?__Yes__ No
If yes, please specify: _____

Incubation Conditions:

___Temperature(s):_____
___Relative Humidity_____
Other: _____

For the following section, please check all procedures used:

____Macrosopic examination ____ Microsopic examination
____Colony morphology ____brightfield
___Pigmentation ____phase contrast
___Other (please specify) ____stains (please specify):
_____Biochemical Tests (please specify automated, non-automated and type)

_____
_____
_____
_____
_____

Bacterial references used for this isolate (texts, manuals):

_____
_____
_____
_____
_____

Authorizing Signature: Date:

Appendix D

Details of statistical procedures, calculation and formulae

The Statistical procedures

The procedures are based on robust statistics and use z-scores to assess participant's performance.

Z-scores are a normalized value which gives a "score" to each result, relative to the other numbers in the group – so a z-score value close to zero means that the result agrees well with those from the other laboratories.

Robust statistics are statistics which are not highly influenced by the presence of extreme results. In a mathematical environment, robustness is the "ability" of a statistical method to be unaffected by outliers. The "classical" (and most commonly used) measure of the center a dataset, the mean (average), is not robust. A robust alternative to the mean is the median (the middle value). The difference between these two is best illustrated by the following example.

Example : When twenty laboratories produce the test results beneath a mean of 12.52 and a standard deviation of 0.54 would be calculated. If just one laboratory made a writing error and reported 122 instead of 12.2 the mean would increase to 18.01 (+44%) and the standard deviation would increase to 24.48 (+400%). Apparently, both mean and standard deviation are extreme sensitive to deviating results. the median would increase to 12.55 (+1%) and the standard deviation would increase to 0.57 (+16%). The median and NIQR are in comparison to the mean and the standard deviation - much less sensitive to the presence of deviating results. Mean and standard deviation are (very) sensitive to deviating results; Median and NIQR are (very) insensitive to deviating results.

test results	12.20 12.5 12.3 12.2 11.9 11.6 11.4 12.4 12.6 13.2
	13.20 13.2 12.3 12.8 12.2 12.7 13.4 12.7 12.5 13.0
n	20
mean	12.52
st.dev	0.54
Median	12.50
NIQR	0.48

test results	122.0 12.5 12.3 12.2 11.9 11.6 11.4 12.4 12.6 13.2
	13.20 13.2 12.3 12.8 12.2 12.7 13.4 12.7 12.5 13.0
n	20
mean	18.01
st.dev.	24.48
Median	12.55
NIQR	0.57

The statistical procedures use z-scores depends on the statistical design of the program. In most cases programs are designed so that pairs of results are obtained, i.e. either two related samples are distributed or (less frequently and avoided if possible) two results on one sample are requested. Related samples can be identical (uniform pair) or similar (split pair). Occasionally it may be the case that a program can only be designed to have a single result on a single sample.

If paired samples are used they may be identical ("blind duplicates") or slightly different (i.e. the properties to be tested are at different levels). The pairs of results which are subsequently obtained fall into two categories: uniform pairs, where the results are expected to be the same (i.e. the samples are identical or the same sample has been tested twice); and split pairs, where the results should be slightly different.

Pairs of results are necessary to evaluate both sources of variation – the variability within z laboratory termed as "within-laboratory variation" and the variability between laboratories termed as "between-laboratories variation". The analysis (and interpretation) of results is the same, regardless of whether a uniform or split sample design has been used. For 'single result on a single sample's programs, the statistical analysis is even simpler than for the paired case. The summary statistics are calculated, as before.
A single result on a single sample is obtained. In this case the statistical analysis is simpler, but it is not possible to differentiate between the two types of variation. Only one set of z-scores is calculated based on the raw results (i.e. the z-score is the difference between the laboratory's result and the median, divided by the normalized IQR). Again, any result with a z-score greater than 3 or less than –3 would be classified as an outlier, and the sign of the z-score indicates whether the result is too high or low. As there is only a single result, a Youden diagram cannot be drawn, so just the z-scores bar chart would appear.
In addition to tables of the results and z-scores, and summary statistics, a number of graphical displays of the data are included in the report for a program. The two most commonly used graphs are the ordered z-score bar-chart and the Youden diagram - both of which are described in detail below.

These charts are to assist the coordinator and technical advisers with the interpretation of the results and are very useful to participants - especially those participants with outliers because they can see how their results differ from those submitted by other laboratories.

Ordered Z-score Chart

One of these charts is generated for each type of z-score calculated - so for our example data there are three of them. On these charts each laboratory's z-score is shown, in order of magnitude, and is marked with its code number. From this each laboratory can readily compare its performance relative to the other laboratories.

These charts contain solid lines at +3 and -3, so the outliers are clearly identifiable as the laboratories whose "bar" extends beyond these cut-off lines. The y-axis is usually limited to range from -5 to +5, so in some cases very large or small (negative) z-scores appear as extending beyond the limit of the chart.

Outlier criteria, Statistical calculations and formulae
An outlier is defined as any result/pair of results with an absolute z-score greater than three, i.e. $Z>3$ or $Z<-3$. This outlier criteria, $|Z|>3$, has a confidence level of about 99% (related to the

normal distribution) - i.e. there is a less than 1% chance that the result(s) is a true member of the population and it is far more likely that there is a problem with this result/pair of results. Similarly a z-score cut-off of two has a confidence level of approximately 95%. Laboratories which have a z-score in this region (i.e. $2<|Z|<3$) are encouraged to "take a close look at" their results.

$|z| \leqslant 2$ ------satisfactory

$2<|z|<3$ ------questionable

$|z| \geqslant 3$ ------unsatisfactory

The "classical" (and most commonly used) measure of the center of a dataset, the mean (average), is not robust.

A robust alternative to the mean is the median (the middle value).

IQR = Q3-Q1

Normalized IQR =IQR×0.7413

z-score = x-median/normalized IQR

The most important statistics used are the median and the normalized IQR - these are measures of the centre and spread of the data (respectively), similar to the mean and standard deviation. The median and normalized IQR are used because they are robust statistics, which means that they are not influenced by the presence of outliers in the data.

The no. of results is simply the total number of results received for a particular test/sample, and is denoted by N. Most of the other statistics are calculated from the sorted results, i.e. from lowest to highest, and in this appendix X[i] will be used to denote the ith sorted data value (e.g. X[1] is the lowest value and X[N] is the highest).

The median is the middle value of the group, i.e. half of the results are higher than it and half are lower. If N is an odd number the median is the single central value, i.e. X[(N+1)/2]. If N is even, the median is the average of the two central values, i.e. (X[N/2] + X[(N/2)+1])/2. For example if N is 9 the median is the 5th sorted value and if N is 10 the median is the average of the 5th and 6th values.

The normalized IQR is a measure of the variability of the results. It is equal to the interquartile range (IQR) multiplied by a factor† (0.7413), which makes it comparable to a standard deviation. The interquartile range is the difference between the lower and upper quartiles. The lower quartile (Q1) is the value below which, as near as possible, a quarter of the results lie. Similarly the upper quartile (Q3) is the value above which a quarter of the results lie. In most cases Q1 and Q3 are obtained by interpolating between the data values. The IQR = Q3–Q1 and the normalized IQR = IQR×0.7413.

The minimum is the lowest value (i.e. X[1]), the maximum is the highest value (X[N]) and the range is the difference between them (X[N]–X[1]).

Once the summary statistics have been calculated for each of the samples and tests in a program, the medians and normalized IQR are tabulated and sent to each laboratory which has returned results as "early information". Following the issue of this information no further changes or additions (e.g. of late results) to the data are permitted.

NOTE: The factor comes from the "standard" normal distribution, which has a mean of zero and a standard deviation (SD) equal to one. The interquartile range of such a distribution is [–0.6745, +0.6745] and this is narrower than the familiar ±1 SD interval. So, to convert an IQR into a ±1 SD range, it must be scaled up by the ratio of the interval widths, namely 2/1.3490. To then convert this ±1 SD range (whose width is 2 standard deviations) into an amount equivalent to 1 SD, this range is then halved. Hence the IQR is divided by 1.3490 (or equivalently multiplied by 0.7413) to convert it into an estimate of the standard deviation.

Suppose the pair of results are from two samples called A and B. The median and normalized IQR of all the sample A results are denoted by median (A) and normIQR (A), respectively. (Similarly for sample B.) A simple robust z-score (denoted by Z) for a laboratory's sample A result would then be:

$$Z = \frac{A - \text{median}(A)}{\text{normIQR}(A)}$$

The standardized sum (denoted by S) and standardized difference (D) for the pair of results are:

$$S = (A+B)/\sqrt{2} \quad \text{and} \quad D = \begin{cases} (B-A)/\sqrt{2} & \text{if } \text{median}(A) < \text{median}(B) \\ (A-B)/\sqrt{2} & \text{otherwise.} \end{cases}$$

Each laboratory's standardized sum and difference are calculated, followed by the median and normalized IQR of all the S's and all the D's - i.e. median(S), normIQR(D), etc. (these summary statistics are usually tabled in the report, to allow participants to calculate the z-scores themselves).

The between-laboratories z-score (denoted by ZB) is then calculated as the robust z-score for S and the within-laboratory z-score (ZW) is the robust z-score for D, i.e.

$$ZB = \frac{S - \text{median}(S)}{\text{normIQR}(S)} \qquad \text{and} \qquad ZW = \frac{D - \text{median}(D)}{\text{normIQR}(D)} \ .$$

The calculated z-scores are tabulated in the report for a program, alongside the corresponding results and the results are assessed based on their z-scores.

When interpreting results which have been identified as outliers, the sign of the z-score and

the design of the program must be considered. For both uniform and split pairs a positive between-laboratories outlier (i.e. ZB>3) indicates that both results for that pair are too high. Similarly a negative between-laboratories outlier (i.e. ZB<-3) indicates that the results are too low.

For uniform pairs, where the results are on identical samples, a within-laboratory outlier of either sign (i.e. |ZW|>3) indicates that the difference between the results is too large. For split pairs, where the analyte is at different levels in the two samples, a positive within-laboratory outlier (i.e. ZW>3) indicates that the difference between the two results is too large and a negative within-laboratory outlier (i.e. ZW<-3) indicates that the difference is too small or in the 'opposite direction' to the medians.

For situations where a program involves a single result on one sample (X) a simple robust z-score is calculated as Z = {X–median(X)}/normIQR(X) and outliers are classified as above - values of X for which |Z| > 3. When an outlier is identified the sign of the z-score indicates whether the result is too high (positive z-score) or too low (negative z-score), but whether this is due to between-laboratories or within-laboratory variation, or both, is unknown.