Evaluation of Interlaboratory Comparisons on Quality Testing Towards Pesticide Formulation of beta-Cyfluthrin, Chlorpyrifos, and Profenofos Active Ingredients

Ayutia Ciptaningtyas Putri[1], Bayu Refindra Fitriadi[1]*

[1]Center for Seeding and Protection of Plantation Plants (BBPPTP) of Surabaya
Jl. Raya Mojoagung No. 52, Jombang 61482, Indonesia

Abstract—The Surabaya Center for Seed and Plantation Protection (BBPPTP Surabaya) carried out interlaboratory comparison research to test the quality of pesticide formulations. This study aimed to determine the ability of each participating laboratory to analyze the concentration of the active ingredients' beta-cyfluthrin, chlorpyrifos, and profenofos in test samples expressed in z-score values. A total of 15 laboratories participated in this program. The quality test method of the pesticide formulations for the beta-cyfluthrin, chlorpyrifos, and profenofos active ingredients referred to the Association of Official Agricultural Chemists (AOAC) method, which was developed and validated. A homogeneity test was carried out before the test samples were distributed. The data were evaluated using a robust z-score statistical calculation algorithm, a method under the International Organization for Standardization (ISO) 13528:2016. The assigned values used to calculate the z-score were obtained by statistical processing of participants' test results. The results of the stability test calculation showed that the distributed sample data was statistically stable. Three laboratory participants were in the questionable category, and one laboratory participant was in the outlier category.

Keywords—Interlaboratory comparison; Pesticide formulation; beta-Cyfluthrin; Chlorpyrifos; Profenofos

1. INTRODUCTION

One of the technical requirements that must be fulfilled in applying SNI ISO/IEC 17025:2017 is the quality assurance of the test results. TesSts by a testing laboratory are part of a significant decision-making process; hence, a mechanism is required to assist the validity of the data issued by the laboratory concerned. This can be achieved by implementing interlaboratory comparisons [1].

Interlaboratory comparison is research to evaluate the performance of testing laboratories against predetermined criteria according to their competence. The general objectives of interlaboratory comparison include evaluating laboratory performance in testing and monitoring laboratory performance, identifying problems in the laboratory, determining the effectiveness and comparability of test and measurement methods, increasing customer confidence in laboratories, identifying differences between laboratories, educating participated laboratories based on the results of comparisons, validating the uncertainty claims, evaluating the performance characteristics of a method, and determining the reference material values [2].

Analysis of the quality of pesticide formulations is a mandatory requirement for pesticide companies to register these pesticides at the Ministry of Agriculture before distribution. In addition, the analysis of the quality of pesticide formulations is also an essential test because the large number of pesticides circulating in the market requires quality controls on the active ingredient content of the pesticide formulation. Khattab et al. [3] have found that pesticides are not registered, and several other pesticides contain active ingredients that do not match the claims on the label. In addition, pesticides that do not contain active ingredients or can be called counterfeit pesticides are also found. Based on data from the Directorate General of Agricultural Infrastructure and Facilities on www.pestisida.id, the number of pesticide formulations registered and permitted by the Minister of Agriculture in 2022 is 1,890 trademarks.

Quality testing laboratories for pesticide

formulations spread throughout Indonesia require comparison facilities to guarantee the quality of test results. However, until 2022, of the 28 laboratories conducting proficiency testing (PUP) that have been accredited by the National Accreditation Committee (KAN), no laboratories have carried out proficiency tests on the quality of pesticide formulations. This also occurs internationally, where most PUP institutions conduct proficiency tests on pesticide residues. It is recorded that only the National Accreditation Board for Testing and Calibration Laboratories (NABL) from India routinely conducts quality proficiency tests for pesticide formulation [4].

The Surabaya Centre for Seed and Plantation Plant Protection (BBPPTP Surabaya) is one of the testing laboratories appointed by the Minister of Agriculture as one of the institutions for testing the quality of pesticide formulation based on the Decree of the Minister of Agriculture Number 653/Kpts/SR.330/M/11/2021 attempting to organize comparison of quality testing of pesticide formulations.

The scope of this comparison research is to test the levels of the pesticide formulations of beta-cyfluthrin ($C_{22}H_{18}Cl_2FNO_3$), chlorpyrifos ($C_9H_{11}Cl_3NO_3PS$), and profenofos ($C_{11}H_{15}BrClO_3PS$) active ingredients (**Fig. 1**). These three pesticides are included in the insecticides class. The preference of this scope is due to the class of insecticides is being categorized as a type of pesticide with a high quantity and use in Indonesia, and many pesticide laboratories in Indonesia can analyze these active ingredients.
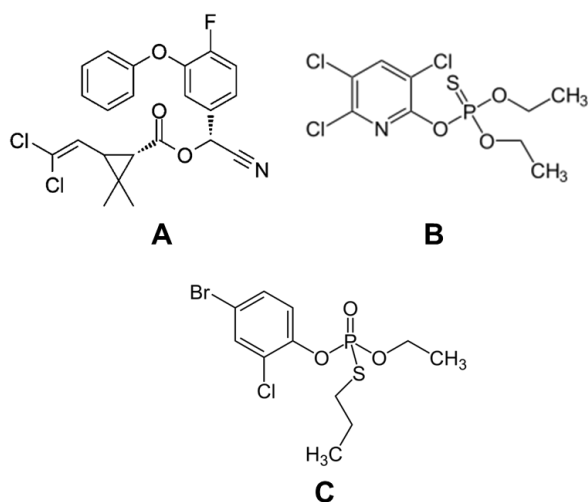
Fig. 1   Chemical structure of beta-cyfluthrin (A), chlorpyrifos (B) and profenofos (C)

The test results data from the laboratory participants are then processed and evaluated using robust statistical tests. The robust statistics are now the preferred approach for proficiency test (PT) providers, as suggested by ISO 13528. Robust statistics reduce the impact of outliers on the calculated statistical parameters, such as the mean and standard deviation [5].

To determine the performance of each participating laboratory, the z-score value needs to be calculated. Laboratory performance is good if the z-score value is between –2 to 2. There are various calculation methods for determining the z-score for each participant. Most commonly, the z-score is used to measure the PT scheme performance of participants on the basis of participant results, assigned value, and standard deviation of proficiency assessment [6]. Rosario et al. [7] compare four z-score calculation methods, concluding that algorithm A calculation method provides a high z-score value. Setiawati [8] also compares four z-score calculation methods, concluding that the median and scaled median absolute deviation (MADe) calculation methods provide the smallest coefficient of variation when compared with other methods. Hardjito [9] uses algorithm A method to calculate the z-score. SNI ISO 13528: 2015 has established several z-score calculation methods that can be used, including the sample mean and standard deviation method, the median and MADe method, the median and normalized interquartile range method (nIQR), algorithm A method, and Qn and Q/Hampel methods [10]. Several calculation methods can be employed to determine the z-score value of the comparison participants; the most unsatisfactory results (outliers) are then selected. This study aims to determine the performance of each participating laboratory based on the z-score value using the algorithm A calculation method.

2. EXPERIMENTAL SECTION

2.1. Materials

The materials used in this research were beta-cyfluthrin analytical grade standard (99.5% purity ChemService, USA), chlorpyrifos analytical grade standard (99.5% purity ChemService, USA), profenofos analytical grade standard (96.4% purity ChemService, USA), acetone analytical grade (JT Baker, USA), beta-cyfluthrin pesticide formulation (A), chlorpyrifos pesticide formulation (B) and profenofos pesticide formulation (C).

2.2. Instrumentations

The instruments used in this research included 10 mL volumetric flask (pyrex), pycnometer (pyrex), 20-100 μL micropipette (Eppendorf, Germany), analytical balance (Kern, Germany), Chromatography Gas (GC) Shimadzu type 2010 plus, rtx-1 column with column dimensions of 4.6 x 150 mm and particle size of 3.5 μm, and Flame Ionization Detector (FID) detector.

2.3. Standard Preparation

Each standard was weighed ±10 mg and put it in a 10 mL volumetric flask, then acetone was added. The standard solution was diluted to a concentration of ±100 μg/mL

2.4. Preparation of Samples

The test sample was weighed ±100 mg, put in a 10 mL volumetric flask, and then acetone was added. The solution was diluted to a concentration of ±100 µg/mL. In addition, the density of the test sample was measured using a 10 mL pycnometer.

2.5. Sample Analysis

Testing on the GC-FID instrument used a modified method from AOAC with conditions of beta-cyfluthrin injector temperature 300 °C, chlorpyrifos 300 °C, and profenofos 300 °C, column temperature beta-cyfluthrin 270 °C, chlorpyrifos 210 °C and profenofos 210 °C, detector temperature beta-cyfluthrin 300 °C, chlorpyrifos 250 °C and profenofos 300 °C.

Standard solution 1 µL was injected into the GC equipment, then 1 µL of the test sample solution was injected. The active ingredient content in the test sample is calculated using the eq. (1).

$$\frac{(height/peak\ area) - intercept}{slope} \times \frac{volume}{weight} \times DF \times \rho \qquad (1)$$

Where DF is dilution factor, and ρ is density of the sample.

2.6. Homogeneity Test

The homogeneity test was carried out by testing ten samples randomly and calculating the Mean Square Between (MSB) (Eq. 2) and Mean Square Within (MSW) (Eq. 3) values.

$$MSB = \frac{\sum \left[(a_i + b_i) - \overline{X}_{(ai+bi)}\right]^2}{2(n-1)} \qquad (2)$$

$$MSW = \frac{\sum \left[(a_i + b_i) - \overline{X}_{(ai+bi)}\right]^2}{2n} \qquad (3)$$

The comparison ingredients will be homogeneous if the F_{count} = MSB/MSW is smaller than the F_{table}.

2.7. Analysis of Participants' Test Results Data

The assigned value was determined based on the consensus value of the comparison participant report. The duplo data analysis results sent by each laboratory were calculated statistically using the robust z-score statistical calculation algorithm A method, according to SNI ISO 13528: 2016. Algorithm A is a Z-score calculation method with strong statistical analysis, identification, and removal of outliers can be avoided. One of algorithm A's advantages is its high relative effectiveness (97%) in determining the population's mean and standard deviation, even with a moderate number of participants [11]. If the participants' test results are not homogeneous, they are selected first using the Dixon test. Furthermore, the selected data are processed using the robust Z-score algorithm A.

2.8. Stability Test

Data from the homogeneity test results were employed as the first data for stability tests. The second data was obtained by analyzing the participants who had committed a comparison by taking three test samples. An example is said to be stable if the first data and the second data do not indicate a significant difference and are determined by the eq. (4).

$$|\overline{X}_i - \overline{\overline{X_{HM}}}| < 0.3 \times n\ IQR \qquad (4)$$

where $\overline{X}_i$ is average sample results of the second test, $\overline{X_{HM}}$ is average homogeneity test results, 0.3 value is constants defined by APLAC, and nIQR is normalized difference between the 3rd and 1st quartiles.

3. RESULT AND DISCUSSION

The homogeneity of the test samples is a critical component in proficiency testing programs. Each participant must receive an identical sample. Thus, the homogenization of test samples in this research was carried out using pesticide formulations originating from the same producer and production code, as well as mixing and stirring. Each active ingredient was labeled and divided into 30 bottles of 30 mL sample size.

3.1. Homogeneity Test

A homogeneity test is required in comparison to reveal that a group of bottles of the test material prepared has fairly homogeneous properties; thus, if there is a difference in the test results of the participants with the reference value, it is not due to the inhomogeneous test material but due to the performance of the laboratory. Ten bottles were randomly selected and analyzed in duplicate in this homogeneity test using a predetermined method.

Series1 (blue dots) and series2 (red dots) show duplo repetition data for each bottle (**Fig. 2**). The results are analyzed using the F-test to evaluate whether the prepared test material is homogeneous. The calculation used the Mean Square Between (MSB) and Mean Square Within (MSW). An example of a comparison is considered homogeneous if the F_{count} = MSB/MSW is smaller than the F_{table}.

The homogeneity graphs above demonstrate the consistency of the test results between 10 test samples and between replicates. To determine the homogeneity level, statistical calculations were carried out.

Based on the statistical data (**Table 1**), the beta-cyfluthrin, chlorpyrifos, and profenofos active ingredients had a F_{count} value smaller than F_{table}; hence,

the test samples prepared for this research were homogeneous and suitable for distribution to participating laboratories.

Table 1. Results of homogeneity test statistics

Active Ingredients	MSB	MSW	F $_{count}$	F $_{table}$*
Beta-cyfluthrin	2.093	0.894	2.339	3.02
Chlorpyrifos	388.31	329.98	1.117	3.02
Profenofos	949.23	609.66	1.557	3.02

*P = 0.05; V1 = 9; V2 = 10
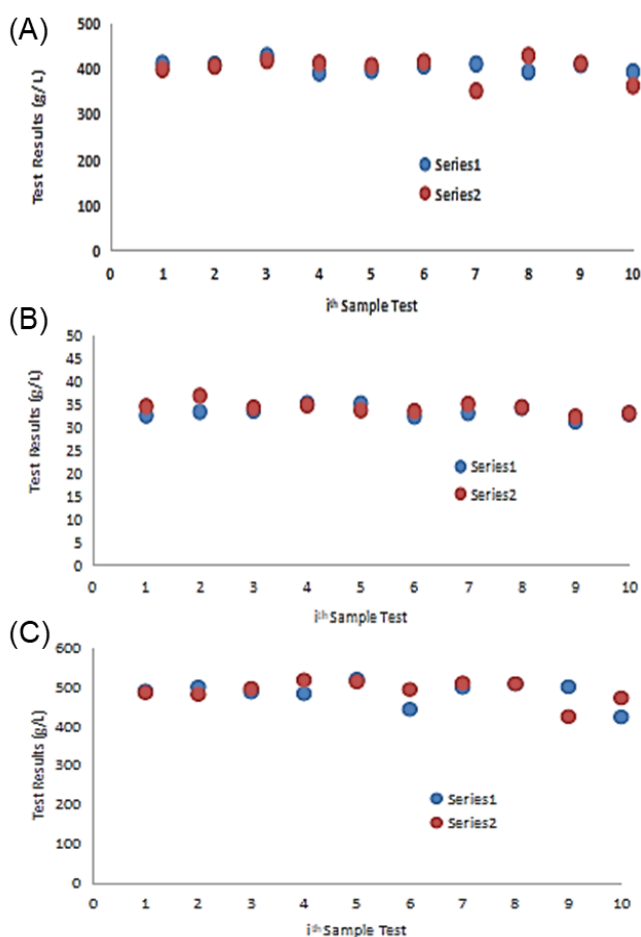
Fig. 2 Homogeneity of (A) beta-cyfluthrin, (B) chlorpyrifos, and (C) profenofos

3.2. Stability Test

Besides the homogeneity test, the stability test (**Table 2**) is also a pivotal factor in determining the success of the comparison program. Many factors can affect the stability of a comparison material, starting from room temperature conditions, storage, and transportation. Therefore, organizers of comparison must ensure that the conditions of the test materials distributed do not change, which can affect the test results of participants [12].

In carrying out this comparison, a stability test is applied to ensure that the quality of the test sample is relatively the same (stable) compared to the quality of

the sample when the test material is being prepared (homogeneity test). If the test substance is stable, then the difference in the participant's test results with the reference value is not caused by differences in the quality of the test material but by the laboratory performance.

Table 2. Summary of stability test statistics

Active Ingredients	$\overline{X_{HM}}$	X_i	$|\overline{X_i} - \overline{\overline{X_{HM}}}|$	$\frac{0.3 \times n}{IQR}$
Beta-cyfluthrin	33.80	33.92	0.123	2.266
Chlorpyrifos	403.21	402.67	0.537	4.873
Profenofos	486.37	481.67	4.69	10.04

The data is said to be stable if the $|\overline{X_i} - \overline{\overline{X_{HM}}}|$ value is smaller than the 0.3 x n/IQR value. Based on the table data above, beta-cyfluthrin, chlorpyrifos, and profenofos active ingredients had $|\overline{X_i} - \overline{\overline{X_{HM}}}|$ value smaller than the 0.3 x n/IQR value; thus, the test sample in this comparison research was declared stable.

3.3. Data Analysis of Participant's Test Result

An easy and internationally accepted statistical method for analyzing the test results of comparison programs and proficiency testing is to calculate the z-score of each participant's test result. Z-score is a normal value that gives a score to each test result and compares it with other test results in the data set for all test results [4].

In this study, the assigned value was determined based on the consensus value of the participant's report. Duplo data of test results sent by each participating laboratory was calculated statistically using the robust z-score statistical calculation algorithm A method referring to SNI ISO 13528: 2016. The result was presented in **Fig. 3**.

The z-score value can be determined with several calculation methods. This research employed the algorithm A method according to SNI ISO 13528:2016 [13]. The preference for the z-score calculation method using algorithm A is based on research conducted by Rosario et al. [7] by comparing four z-score calculation methods where algorithm calculation method A provides a high z-score value and causes many participants to obtain outlier results. This method strictly selected z-score values to increase confidence in the comparison results.

The use of the algorithm A method is also mentioned by Aryana [14] in her research, which compares three calculation methods, namely the algorithm A method, the Hampel method, and the M estimator method, where the three methods provide similar effectiveness in assessing participant performance. Rosario et al. [7] also state that the algorithm A method is the second best of the four research methods after the fit-to-purpose method.
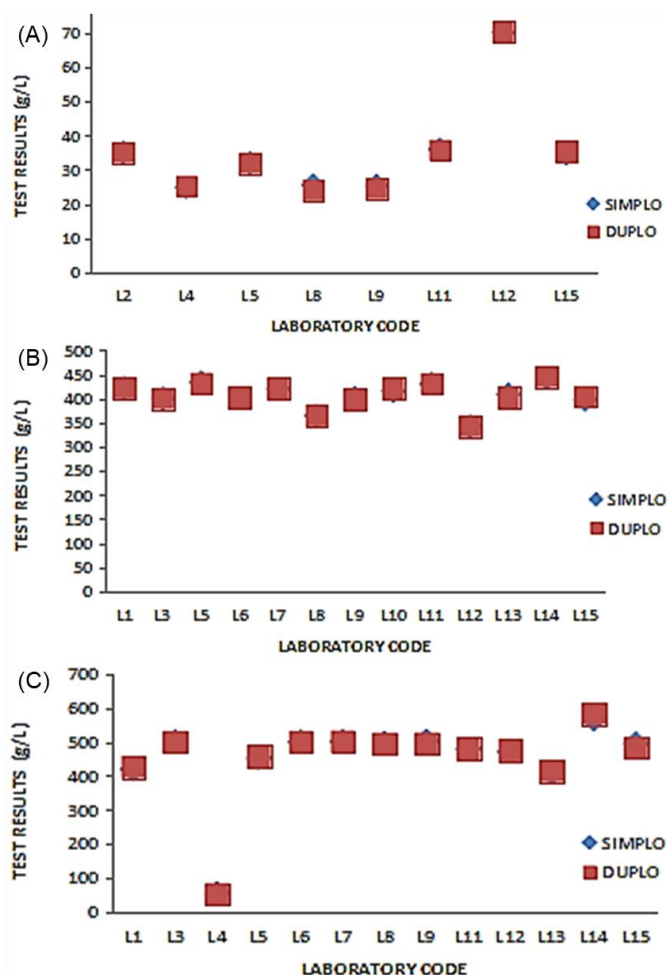
Fig. 3 Data distribution of (A) beta-cyfluthrin, (B) chlorpyrifos, and (C) profenofos test result

Non-homogeneous data (L12 in **Fig. 3(A)** and L4 in **Fig. 3(C)**) were obtained in the results of the beta-cyfluthrin and profenofos data tests. This data must be removed from the data. Dixon test can be used to remove the data. The Dixon test is another well-known test for outliers, which is popular because it's easy to calculate. In the case of small samples, Dixon tests a suspect measurement by finding the difference between that measurement and the nearest one in size with respect to the range of the measurements [15].

The Dixon test is employed to select data from proficiency tests or comparisons if it turns out that there are different data from the other participants' data, such as the data on the distribution of the profenofos test results from the participating laboratories. There was data from the L4 code, which differed greatly from the data from most other participating laboratories. Likewise, the data from the L12 code on the results of the beta-cyfluthrin test were much different from the laboratory results of the other participants.

To process the data using the Dixon test, the data must first be arranged starting from the smallest data. Data declared outliers by the Dixon test were excluded from the Robust z-score test.

Table 3. Dixon test result of beta-cyfluthrin and profenofos

Active Ingredients	Total Data	Lowest Data	Highest Data	D-table Dn	95%
Beta-	8	0.051	0.763	D8	0.608
cyfluthrin	7	–	0.082	D7	0.569
Profenofos	13	0.818	0.448	D13	0.611
	12	0.098	–	D12	0.479

Based on the Dixon test (**Table 3**), the lowest data was smaller than D8; thus, the code data from L4 was not discarded. Meanwhile, the highest data was greater than D8; thus, the L12 code data was discarded. Furthermore, the L11 code is not discarded because the highest data is still smaller than D8. Hence, the L12 code has been declared an outlier by the Dixon test and is not included in the Robust z-score test.

The Dixon test treatment was also applied to the distribution of data from the profenofos test results, which demonstrated that the data were highly different from other participants' data.
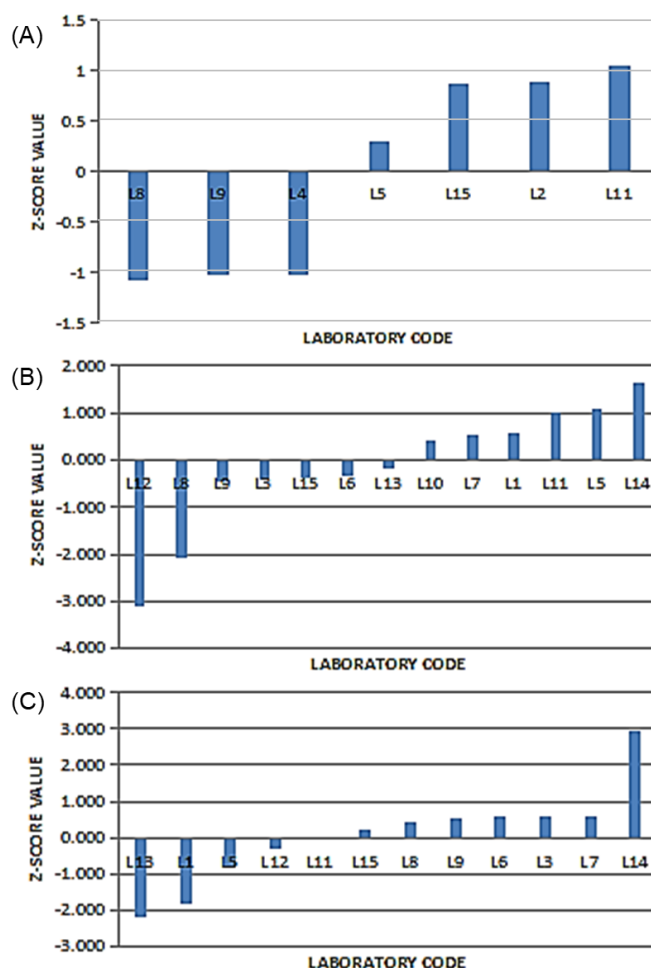
Fig. 4 Z-score graph value of the (A) beta-cyfluthrin, (B) chlorpyrifos, and (C) profenofos

Based on **Table 3**, it can be implied that the lowest data had a more significant value than D13; thus, the L4 code data was discarded, while the highest data had a smaller value than D13, so the L14 code data was not

discarded. The data from the L4 code were declared an outlier using the Dixon test; hence, it was not included in the Robust z-score test.

After all the data from the participant's laboratory was uniform, the data were analyzed using the Robust z-score test with the algorithm A method. The algorithm A method explains that the robust average value and robust standard deviation are calculated based on the algorithm A with an iterative (repeated) process. This method has been mentioned in SNI ISO 13528:2016, complementing ISO/IEC Guide 43, which does not detail the use of statistical methods in proficiency testing [7]. The z-score values from the results of algorithm A calculations are illustrated in **Fig. 4.**

The z-score value obtained can be used to conclude whether or not the participants' laboratory test results are accepted. The acceptance criterion of the z-score obtained in each participating laboratory was that the laboratory was included in the criteria for acceptance (inlier) if the |z-score| ≤ 2.0. The laboratory is included in the criteria for being questionable if the z-score value of the score is in the range of 2.0 < |z-score| < 3.0, and the laboratory is in the criteria of not being accepted (outlier) if the |z-score| ≥ 3.0. The complete results of the evaluation of participant acceptance based on the z-score value criteria are shown in **Table 5**.

Table 5. Complete participant acceptance evaluation results

Participant Lab Code	Beta-cyfluthrin	Chlorpyrifos	Profenofos
L1	*	Inlier	Inlier
L2	Inlier	*	*
L3	*	Inlier	Inlier
L4	Inlier	*	*
L5	Inlier	Inlier	Inlier
L6	*	Inlier	Inlier
L7	*	Inlier	Inlier
L8	Inlier	Questionable	Inlier
L9	Inlier	Inlier	Inlier
L10	*	Inlier	*
L11	Inlier	Inlier	Inlier
L12	*	Outlier	Inlier
L13	*	Inlier	Questionable
L14	*	Inlier	Questionable
L15	Inlier	Inlier	Inlier

*did not follow the test parameters

Table 5 reveals that there are three participating laboratories in the outlier category. Participating laboratories that received questionable criteria and were outliers must conduct investigations and corrective actions on their test results. According to KAN U-08 Rev. 1 [16], if the results are unsatisfactory, the laboratory must immediately investigate to review the laboratory's technical competence and quality system. Laboratories must analyze the root cause, take appropriate corrective action, and preserve evidence recording of actions taken. Corrective action was carried out by determining the root cause of the problem. In determining the root of the problem, it is necessary to carefully analyze all potential causes, including requirements of the proficiency test providers, samples, methods, and procedures, personnel skills and training, consumables, or equipment and their calibration. Using fishbone diagrams to determine the root causes of non-conformities can be an option for comparison participants. The fishbone diagram can identify all potential causes. Moreover, determining the root cause of the non-conformity can be carried out so that the next step in the form of corrective action can be considered [17].

CONCLUSION

Statistical calculation method robust z-score algorithm A could be used to evaluate the acceptability of the results of comparison participants testing the quality of pesticides beta-cyfluthrin, chlorpyrifos, and profenofos. There were three participating laboratories in the questionable category and one participating laboratory in the outlier category, which must carry out investigations and corrective actions.

SUPPORTING INFORMATION

There is no supporting information of this paper. The data that support the findings of this research are available on request from the corresponding author (BR. Fitriadi).

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

All authors do not have a conflict of interest with any party in this research.

AUTHOR CONTRIBUTIONS

A. C. Putri conducted the experiment and calculations, B. R. Fitriadi wrote and revised the manuscript. All authors agreed to the final version of this manuscript.

REFERENCES

[1] Komite Akreditasi Nasional. 2017. *SNI ISO/IEC 17025:2017: Persyaratan Umum Kompetensi Laboratorium Pengujian dan Laboratorium Kalibrasi.* Jakarta.

[2] Komite Akreditasi Nasional. 2015. *SNI ISO/IEC 17043:2015: Persyaratan Umum Kompetensi Laboratorium Pengujian dan Laboratorium Kalibrasi.* Jakarta.

[3] Khattab, G.M., El-Sayed, W.M., Attalla, I.M., and Megeed, A. 2020. Monitoring of the counterfeit Abamectin in Pesticide Product in Egypt. *Arab Univ. J. Agric. Sci.* 82(2), 601-615. doi: 10.21608/ajs.2020.21756.1146

[4]   Sanyal D., and Rani A. 2009. Proficiency test for chemical laboratories for the analysis of a pesticide in a formulated product: interlaboratory study. *J. AOAC Int.* 92(1), 271-278. PMID: 19382586.

[5]   Brookman, B. and Mann, I. 2021. *Eurachem Guide: Selection, Use and Interpretation of Proficiency Testing (PT) Schemes (3rd edition).* Belgium: Eurachem.

[6]   Krismastuti, F.S.H. 2021. Proficiency testing scheme based on ISO/IEC 17043 in Indonesia: Strategy in complying the technical requirements for the proficiency testing providers. *J. Kim. Terap. Indones.* 23(1), 17–23. doi: 10.14203/inajac.v23i1.476.

[7]   Rosario, P., Martinez, J.L., and Silvan, J.M. 2008. Comparison of different statistical methods for evaluation of proficiency test data. *Accreditation Qual. Assur.* 13(9), 493-499. doi: 10.1007/s00769-008-0413-7.

[8]   Setiawati, I. 2018. Analisis robust untuk evaluasi hasil uji banding antar laboratorium produk asam formiat teknis. *Portal Media Ilmiah Bidang Kimia dan Kemasan.* 5(1), 208-214.

[9]   Hardjito, 2019. Evaluasi uji banding antar laboratorium untuk mengukur kompetensi personil. *Indones. J. Lab.* 2(1), 29-33. doi: 10.22146/ijl.v1i4.52990.

[10] International Standarization Organization. 2015. *ISO 13528 Statistical methods for use in proficiency testing by interlaboratory comparison.* Switzerland.

[11] Carobbi. 2017. A modified ISO 13528 robust analysis (Algorithm A) that takes measurement uncertainty into account. *Measurement.* 110, 96-306. doi: 10.1016/j.measurement.2017.07.006.

[12] Sykes, M. and Thompson, M. 2015. Assessing the stability of a proficiency test material by participant-blind re-use after a period of storage. *Analytical Methods.* 7(23), 9753–9755. doi: 10.1039/C5AY02255J.

[13] Komite Akreditasi Nasional. 2016. *SNI ISO 13528:2016, Penggunaan Metode Statistik pada Uji Profosiensi melalui Uji Banding antar Laboratorium.* Jakarta

[14] Aryana, N. 2020. Proficiency testing for determination of pesticides residues in black tea: Comparison of three robust statistical approaches based on ISO 13528 to estimate the consensus values for small number of participants. *IOP Conf. Ser. Mater. Sci. Eng.* 1011, 012029. doi: 10.1088/1757-899X/1011/1/012029.

[15] Miller, J.N. and Miller, J.C. 2018. *Statistics and Chemometrics for Analytical Chemistry, 7th edition.* London: Coronet Books Inc.

[16] Komite Akreditasi Nasional, 2022. *Kebijakan Uji Profisiensi: KAN U-08 Rev. 1.* KAN: Jakarta.

[17] Hadi, A. 2015. Investigasi uji profisiensi klorida dengan hasil 'outlier'. https://www.infolabling.com/2015/02/investigasi-uji-profisiensi-klorida.html#.YyfOknZBzIU accessed on August 2, 2023.