International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Inter-Laboratory Agreement of the Multiple Tube Fermentation Technique (MTFT) for Coliform Enumeration in Philippine Drinking Water Laboratories

Nicolle Anne A. Dajang[1], Micah D. Reyes[2], Ma. Mechaella A. Tinio[3], Janiah Y. Allani[4], Camille P. Roz[5]

[1,2,3,4,5]Student, Master of Science in Microbiology, Graduate School, University of Perpetual Help System Laguna

ABSTRACT

Access to safe drinking water depends on reliable microbiological monitoring, with the Multiple Tube Fermentation Technique (MTFT) remaining the standard method for coliform detection in many developing countries, including the Philippines. Despite its widespread regulatory use, evidence regarding its inter-laboratory reproducibility under routine operational conditions remains limited. This study evaluated the inter-laboratory reproducibility of MTFT for enumerating total and thermotolerant coliforms in drinking water. Three accredited Philippine drinking water laboratories independently analyzed three standardized water samples representing low, medium, and high contamination levels using MTFT. Reported Most Probable Number (MPN) values were evaluated using descriptive statistics, coefficients of variation, confidence intervals, range, and pairwise percent difference analyses, trend consistency assessment, Intraclass Correlation Coefficients (ICC), Fleiss' kappa statistics, and exploratory inferential analyses.

All laboratories identified the expected contamination gradient across the standardized samples; however, substantial variability in quantitative measurements was observed. Total coliform measurements demonstrated the greatest variability at low contamination levels (CV = 104.57%), while thermotolerant coliform measurements showed complete agreement only in the low-contamination sample and considerable variability in the remaining samples. Pairwise percent differences reached 176%, and reliability analyses indicated poor inter-laboratory reproducibility for both total coliform (ICC = 0.081) and thermotolerant coliform (ICC = 0.267) measurements. Fleiss' kappa further demonstrated poor agreement for total coliform ($\kappa = -0.125$) and slight agreement for thermotolerant coliform ($\kappa = 0.100$) pass/fail classifications. Although one-way ANOVA and Kruskal–Wallis analyses detected no statistically significant differences among laboratories ($p > 0.05$), descriptive and agreement analyses consistently demonstrated inter-laboratory variability.

These findings indicate that while MTFT remains efficient for detecting microbiological contamination and identifying contamination trends, quantitative reproducibility among participating laboratories remains limited under routine operational conditions. Strengthening quality assurance practices, standardization, and inter-laboratory comparison programs is essential to improve analytical

comparability, support consistent regulatory decision-making, and enhance drinking water quality monitoring in the Philippines.

Keywords: Multiple Tube Fermentation Technique (MTFT), Most Probable Number (MPN), inter-laboratory variability, drinking water quality, coliform bacteria

INTRODUCTION

Access to safe, potable drinking water is a critical global public health priority, directly influencing disease prevention, community well-being, and progress toward the United Nations Sustainable Development Goal 6 (Houck et al., 2019; Rajapakse et al., 2023). Despite sustained global efforts, approximately two billion people continue to rely on drinking water sources contaminated with fecal matter, contributing significantly to the burden of waterborne diseases (Rajapakse et al., 2023; United Nations, 2023). This challenge is particularly pronounced in low- and middle-income countries, where water quality monitoring systems are often constrained by limited infrastructure, technical capacity, and resource availability (Sikder et al., 2021; Zimmer et al., 2025). In such contexts, reliable microbiological assessment is essential to safeguard public health and ensure compliance with water safety standards. Consequently, the detection of total and thermotolerant coliforms, including *Escherichia coli*, remains the gold-standard indicator of fecal contamination and a fundamental component of drinking water monitoring programs (Dincer et al., 2022; Sikder et al., 2021).

Among conventional microbiological methods, the Multiple Tube Fermentation Technique (MTFT) remains widely used for its simplicity, cost-effectiveness, and longstanding regulatory acceptance. The method is embedded in international and national standards, including those aligned with World Health Organization guidelines and standard protocols such as those of the American Public Health Association, and remains a routine procedure in many laboratories worldwide (Dincer et al., 2022; Magwilang et al., 2023). MTFT estimates bacterial density using the Most Probable Number (MPN) approach, which relies on statistical inference based on observed growth patterns across a series of inoculated tubes (Dincer et al., 2022; Ting & Yee, 2020). Despite the emergence of more advanced techniques, including membrane filtration, enzyme-based assays, and molecular methods such as qPCR, MTFT remains the preferred and legally accepted method in many developing settings where access to advanced technologies is limited (Sikder et al., 2021; Ting & Yee, 2020). In some cases, MTFT has demonstrated comparable or even higher detection sensitivity for certain coliform groups than alternative methods, reinforcing its continued relevance in routine water quality monitoring (Saavedra-Ruiz & Resto-Irizarry, 2025; Zimmer et al., 2025).

However, MTFT's reliance on MPN-based statistical estimation introduces inherent variability in reported results. Unlike direct enumeration methods, MPN values are associated with confidence intervals and may be influenced by factors such as analyst interpretation of positive reactions, media preparation, incubation conditions, and procedural consistency (Sikder et al., 2021; Sreya et al., 2024; Ting & Yee, 2020). Consequently, differences in laboratory practices and operational environments may contribute to variability in MTFT outcomes, even when standardized protocols are followed.

Globally, there is increasing recognition of the importance of inter-laboratory comparability in microbiological testing. Reliable water quality monitoring depends not only on the accuracy of individual methods but also on the consistency of results across different testing facilities. Inter-laboratory comparison and proficiency testing are therefore critical components of quality assurance systems that

help identify systematic differences and improve analytical reliability (Molina-Castro et al., 2021; Sreya et al., 2024). Nevertheless, most studies evaluating MTFT focus on method validation under controlled laboratory conditions, with limited attention given to agreement and reproducibility among laboratories operating under routine conditions (Balleste et al., 2019; Sreya et al., 2024). This gap is particularly relevant in developing countries, where differences in laboratory capacity, training, and infrastructure may further influence testing outcomes (Sari et al., 2020; Sikder et al., 2021).

In the Philippine context, drinking water quality is regulated under the Philippine National Standards for Drinking Water of 2017, which establishes strict bacteriological criteria to protect public health (Caneos et al., 2024; Magwilang et al., 2023). A network of accredited laboratories supports compliance across water districts, local government units, and the rapidly expanding drinking water refilling station sector, which serves as a primary water source for many households (Magwilang et al., 2023; Sari et al., 2020). Despite the central role of these laboratories in ensuring water safety, there is a lack of published evidence evaluating the consistency of MTFT results across testing facilities nationwide. In particular, the reproducibility of pass/fail classifications, which directly inform regulatory decisions and public health interventions, has not been systematically assessed (Balleste et al., 2019; Cambarihan et al., 2022). This absence of comparative data raises concerns regarding the reliability and comparability of laboratory findings used to safeguard public health.

Given the critical role of laboratory results in determining water safety, inconsistencies in MTFT outcomes may have significant implications for public health protection and regulatory decision-making. Variability between laboratories could lead to conflicting interpretations of the same water sample, potentially resulting in false assurance of safety or unnecessary corrective actions. Such discrepancies may undermine public confidence in water quality monitoring systems and compromise the effectiveness of regulatory frameworks.

In response to these gaps, this study evaluates the inter-laboratory reproducibility of MTFT results for total and thermotolerant coliform enumeration in drinking water samples within the Philippine context. Specifically, three accredited laboratories independently analyzed a common set of samples representing increasing contamination levels: a low-level, a medium-level, and a high-level *Escherichia coli*-spiked sample.

The study aimed to:

1. Determine the extent of inter-laboratory variability in MTFT results for total and thermotolerant coliform enumeration;
2. Compare the MPN values reported by participating laboratories across different contamination levels;
3. Evaluate the consistency of contamination trends observed among laboratories;
4. Determine the level of agreement in pass/fail classifications based on applicable drinking water standards; and
5. Assess the reproducibility of MTFT under routine testing conditions.

By comparing reported MPN values and corresponding pass/fail classifications across laboratories, this study assesses variability, agreement, trend consistency, and reproducibility under routine testing conditions. By situating the findings within both global standards and local laboratory practices, the study provides critical insight into the reliability of MTFT as applied in real-world settings. The results are expected to strengthen laboratory standardization, enhance quality assurance systems, and improve the consistency of water quality monitoring programs. Ultimately, this work supports broader efforts to ensure

safe drinking water and protect public health, aligning with both national regulatory priorities and global sustainability goals

METHODOLOGY

Study Design

This study employed a multi-laboratory analytical comparative design to evaluate the inter-laboratory consistency of the Multiple Tube Fermentation Technique (MTFT) for total and thermotolerant coliform enumeration in drinking water under routine operational conditions. The design focused on assessing variability in Most Probable Number (MPN) results, adherence to expected contamination trends, reproducibility, and agreement in pass/fail classification across accredited laboratories (Gadrich et al., 2022). By allowing participating laboratories to perform analyses independently using their routine procedures, the study aimed to capture real-world variability reflective of actual monitoring conditions, consistent with established frameworks for the validation of analytical methods (International Organization for Standardization, 1994).

Participating Laboratories

Three Department of Health (DOH)-accredited drinking water laboratories located in CALABARZON and Metro Manila participated in the study: Aquatest Laboratory Water Testing Services, Dasmariñas Water District Water Testing Laboratory, and Wellzone Laboratory Water Testing Services. Laboratories were selected based on the following inclusion criteria: (1) active DOH accreditation for drinking water microbiological analysis; (2) implementation of MTFT as part of routine water quality testing services; (3) availability of qualified microbiology analysts; and (4) willingness to participate in the study under routine operating conditions. All participating laboratories maintained valid DOH accreditation throughout the study and operated under quality management systems aligned with ISO/IEC 17025 requirements.

Analyses were performed by laboratory personnel authorized to conduct drinking water microbiological examinations under their respective quality management systems. Testing was conducted under routine laboratory environmental conditions, including controlled incubation temperatures, designated microbiological workspaces, and standard aseptic procedures consistent with accreditation requirements.

Test Organism and Inoculum Preparation

A 24-hour broth culture of *Escherichia coli* ATCC 8739 was prepared by inoculating the organism in Tryptic Soy Broth and incubating at 37 °C for 18 h. The culture was maintained within five passages from the original stock to ensure viability and genetic consistency, in accordance with standard microbiological practices for working stocks and reference cultures (ATCC, 2024; Bortolaia et al., 2021; Hendriksen et al., 2012). The stock suspension was serially diluted using a sterile diluent (e.g., phosphate-buffered saline or physiological saline) to obtain target bacterial concentrations. For enumeration, 0.1 mL aliquots from appropriate dilutions were spread-plated onto Tryptic Soy Agar and incubated for 18–24 h at 35 ± 0.5 °C; colony-forming units per milliliter (CFU/mL) were calculated by the spread-plate method. The computed concentrations were used to prepare low (~10^4 CFU/mL), medium (~10^6 CFU/mL), and high (~10^8 CFU/mL) inoculum levels, following the three-level contamination design commonly used in AOAC collaborative validation studies (Crowley et al., 2010; Feldsine et al., 2005).

Sample Preparation and Experimental Setup

All sample preparation was conducted aseptically under a laminar flow hood to prevent contamination. Sterile water was used as the base matrix for all samples. Sample A was prepared by inoculating 125 mL

of sterile water with 0.1 mL of a suspension containing approximately 10^4 CFU/mL to represent low-level contamination. Sample B was prepared similarly using a suspension containing approximately 10^6 CFU/mL to represent medium-level contamination. Sample C was prepared using a suspension containing approximately 10^8 CFU/mL to represent high-level contamination.

Three identical sets of these samples were prepared simultaneously using the same batch of sterile water and inoculum to ensure consistency across all test units. The prepared samples were then distributed to the participating laboratories for analysis. Each laboratory analyzed the samples independently under routine operating conditions without external intervention.

Data Processing and Statistical Analysis

Raw data obtained from the laboratories included reported MPN values and corresponding pass/fail classifications. Values reported as below the detection limit (e.g., <1.1 MPN/100 mL) and above the method's measurable range were treated as censored data and standardized prior to analysis.

Descriptive statistics, including mean, median, standard deviation, coefficient of variation (CV), and range, were used to summarize MPN values across laboratories and sample types. Inter-laboratory variability was assessed using percent-difference calculations and coefficient-of-variation analyses. Trend analysis was performed to evaluate whether the expected contamination gradient (Sample A < Sample B < Sample C) was consistently observed across participating laboratories.

Differences in MPN values among laboratories were evaluated using the Kruskal–Wallis test because of the small sample size and the non-normal distribution of the MPN data (Eissa, 2025; Weiss et al., 2017). Agreement in pass/fail classification was assessed using Fleiss' kappa coefficient (κ), which measures agreement beyond chance among multiple raters and is the appropriate extension of Cohen's kappa for three or more raters (Kılıç, 2015; Zapf et al., 2016). Inter-laboratory reproducibility was further evaluated using the Intraclass Correlation Coefficient, which quantified the consistency of MPN measurements reported by participating laboratories (Liljequist et al., 2019; Nakagawa & Schielzeth, 2010). These statistical approaches are consistent with recognized frameworks for method validation, reproducibility assessment, and inter-laboratory comparison in microbiology (Boubetra et al., 2011; ISO 5725-1, 1994; Sartory, 2007).

To facilitate visualization of agreement patterns and laboratory performance, graphical analyses including box plots, heatmaps, and trend plots were generated. These statistical approaches are consistent with recognized frameworks for method validation, reproducibility assessment, and inter-laboratory comparison (ISO 5725-1, 1994; McBride et al., 2004).

Quality Assurance and Control

Quality assurance measures were implemented throughout sample preparation and analysis to ensure consistency and reliability. All procedures were conducted under aseptic conditions, and a single batch of sterile water and inoculum was used for all samples to minimize variability. Duplicate plating was performed during CFU determination to ensure the accuracy of bacterial concentration estimates. Samples were transported under controlled conditions to prevent contamination or degradation prior to analysis. Each participating laboratory followed its internal quality assurance and quality control protocols as part of DOH accreditation requirements, which are aligned with the ISO/IEC 17025 standard (Syera et al., 2024).

Ethical Considerations

This study did not involve human participants or clinical samples. All water samples were prepared in the laboratory and handled in accordance with standard microbiological safety procedures. Participation of

accredited laboratories was coordinated in advance, and all data were anonymized to maintain institutional confidentiality. The study adhered to ethical standards for laboratory-based research and responsible data reporting.

RESULTS

The raw Most Probable Number (MPN) values generated by the three participating accredited drinking water laboratories are presented in Table 1. Although all laboratories analyzed identical standardized drinking water samples using the Multiple Tube Fermentation Technique (MTFT), differences in quantitative measurements were observed for both total and thermotolerant coliforms.

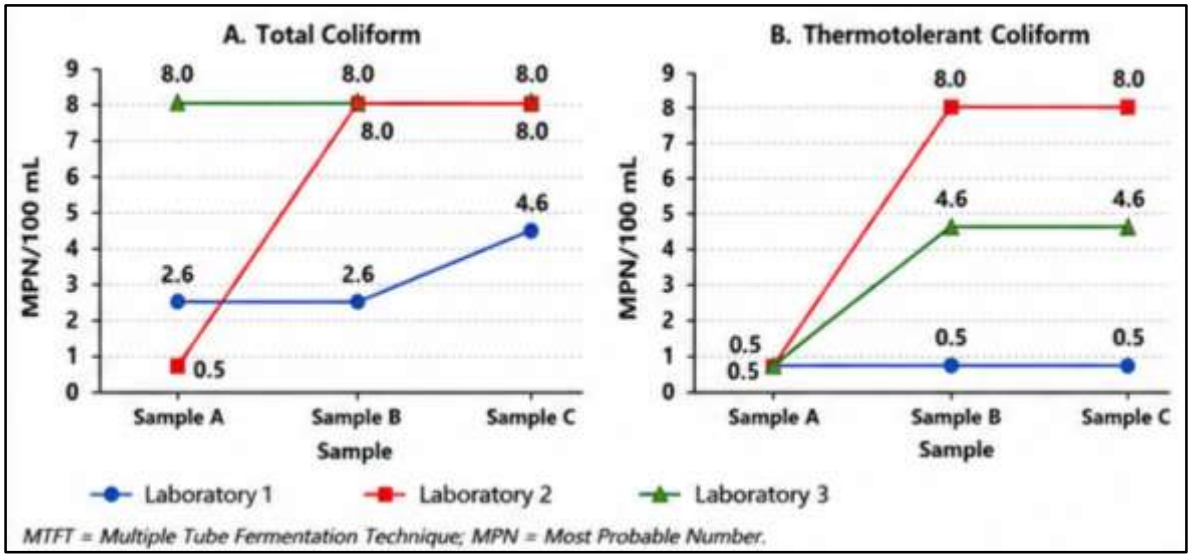
Table 1. Raw Inter-Laboratory MPN (MPN/100 mL) Values for Total and Thermotolerant Coliforms

Parameter	***Lab 1***	***Lab 2***	***Lab 3***
Total Coliform	2.6	0.5	8.0
	2.6	8.0	8.0
	4.6	8.0	8.0
Thermotolerant Coliform	0.5	0.5	0.5
	0.5	8.0	4.6
	0.5	8.0	4.6

For total coliforms, the laboratories generally reflected the expected increase in contamination across the three samples. However, quantitative agreement varied, particularly in Samples B and C, where Laboratories 2 and 3 consistently reported values at the upper detection limit (8.0 MPN/100 mL), while Laboratory 1 reported comparatively lower counts. For thermotolerant coliforms, complete agreement was observed for Sample A, whereas substantial differences emerged for Samples B and C, with the participating laboratories reporting different MPN values despite analyzing identical samples. These observations provide initial evidence of inter-laboratory variability in MTFT-derived quantitative results.

Figure 1.

Laboratory-Specific MTFT Results Across Samples



Note. Graphs (A) and (B) present the laboratory-specific results obtained using the Multiple Tube Fermentation Technique (MTFT) for Total Coliform and Thermotolerant Coliform enumeration, respectively. Measurements are expressed as Most Probable Number (MPN) per 100 mL for Samples A, B, and C. Each line represents the results reported independently by one of the three participating accredited laboratories, illustrating inter-laboratory variation across contamination levels.

Figure 1 illustrates the laboratory-specific MTFT results for total and thermotolerant coliforms across the three standardized samples. All laboratories consistently demonstrated the expected increase in contamination from Sample A to Sample C. However, differences in reported MPN values became more apparent as contamination levels increased.

For total coliforms, Laboratory 3 consistently reported the highest counts across all samples, whereas Laboratory 1 generated comparatively lower values. Laboratory 2 exhibited the greatest increase between Sample A and Samples B and C. For thermotolerant coliforms, identical results were obtained for Sample A, while Samples B and C showed marked variation among laboratories. Overall, the graphical trends show that although MTFT consistently detected contamination, quantitative agreement among laboratories was variable.

Table 2. Descriptive Statistics of Total and Thermotolerant Coliform Counts (MPN/100 mL) Across Samples

Parameter	***Sample***	***Mean ± SD***	***Minimum***	***Maximum***	***Range***	***CV %***	***95 % CI***
Total Coliform	A	3.70 ± 3.87	0.5	8.0	7.5	104.57	-5.91 to 13.31
	B	6.20 ± 3.12	2.6	8.0	5.4	50.29	-1.54 to 13.94
	C	6.87 ± 1.96	4.6	8.0	3.4	28.59	1.99 to 11.74
Thermotolerant Coliform	A	0.50 ± 0.00	0.5	0.5	0.0	0.00	0.50 to 0.50
	B	4.366 ± 3.76	0.5	8.0	7.5	86.00	-4.96 to 13.70
	C	4.366 ± 3.76	0.5	8.0	7.5	86.00	-4.96 to 13.70

Legend: Values are expressed as mean ± standard deviation (SD). Minimum, maximum, and range represent inter-laboratory variability. Values of 8.0 indicate upper detection limit (>8.0 MPN/100 mL).

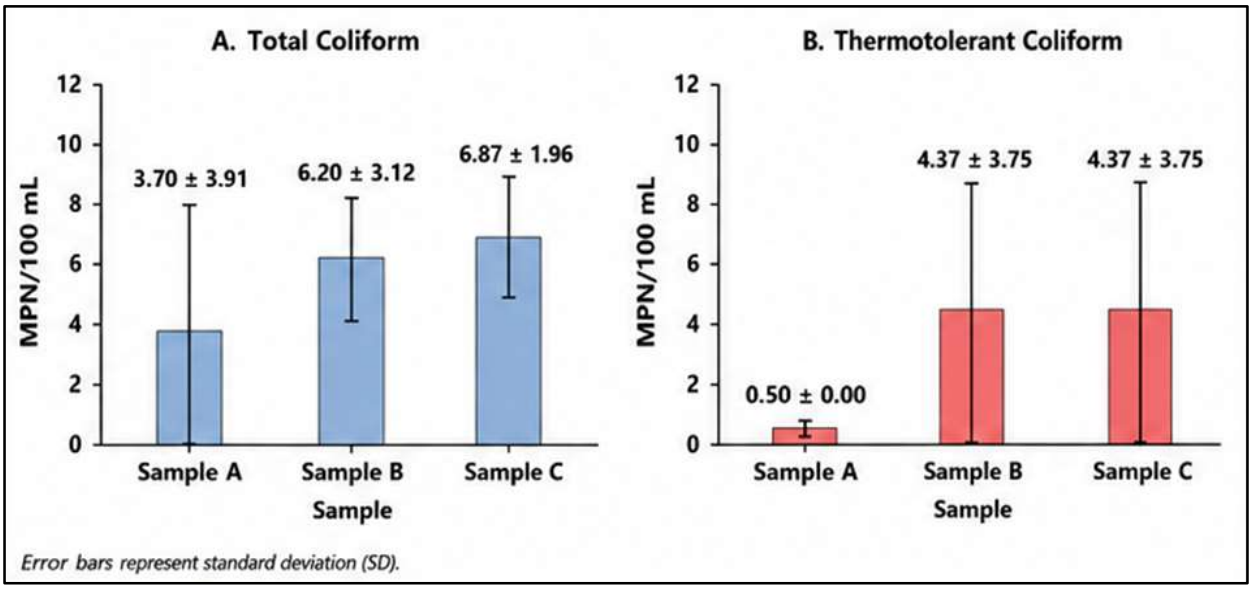
Table 2 shows the descriptive statistics that summarized the central tendency and variability of MTFT-derived MPN values across participating laboratories. For total coliforms, the mean concentration increased from Sample A (3.70 MPN/100 mL) to Sample C (6.87 MPN/100 mL), reflecting the intended contamination gradient of the standardized samples.

Variability among laboratories was greatest for Sample A, which exhibited the highest standard deviation and coefficient of variation (CV = 104.57%). In contrast, Sample C demonstrated the lowest relative variability (CV = 28.59%), indicating comparatively greater consistency among laboratory measurements at higher contamination levels. The corresponding confidence intervals likewise narrowed from Sample A to Sample C, suggesting reduced dispersion of the reported values.

For thermotolerant coliforms, all laboratories reported identical results for Sample A, resulting in a coefficient of variation of 0%. Conversely, Samples B and C exhibited substantial variability (CV = 86.00%), indicating lower consistency in quantitative enumeration among participating laboratories.

Figure 2.

Mean MTFT results across participating laboratories



Note. Graphs (A) and (B) show the mean Total Coliform and Thermotolerant Coliform concentrations, respectively, calculated from the results of the three participating laboratories. Bars represent the mean MPN/100 mL, while error bars indicate one standard deviation (± SD), providing a measure of inter-laboratory variability for each standardized sample.

Figure 2 summarizes the mean MPN values and corresponding standard deviations for each sample. Error bars visually demonstrate the magnitude of inter-laboratory variability.

For total coliforms, variability progressively decreased as contamination levels increased, whereas thermotolerant coliform measurements showed complete agreement only in Sample A and considerably larger dispersion in Samples B and C. These findings are consistent with the descriptive statistics and shows that reproducibility differed across contamination levels and target organisms.

Table 3. Range-Based Variability of Coliform Counts Across Laboratories Per Sample

Parameter	***Sample***	***Range***	***Interpretation***
Total Coliform	A	7.5	Very high variability
	B	5.4	High variability
	C	3.4	Moderate variability
Thermotolerant Coliform	A	0.0	No variability

	B	7.5	Very high variability
	C	7.5	Very high variability

Table 3 shows the range analysis that was performed to evaluate the absolute spread of laboratory measurements. For total coliforms, the largest range was observed in Sample A, indicating very high variability among participating laboratories. Variability gradually decreased in Samples B and C, although differences between laboratory results remained evident.

For thermotolerant coliforms, no variability was observed in Sample A because all laboratories reported identical measurements. In contrast, Samples B and C exhibited equally large ranges, indicating substantial differences in reported MPN values. These findings demonstrate that the magnitude of inter-laboratory variability varied with both contamination level and microbial parameters.

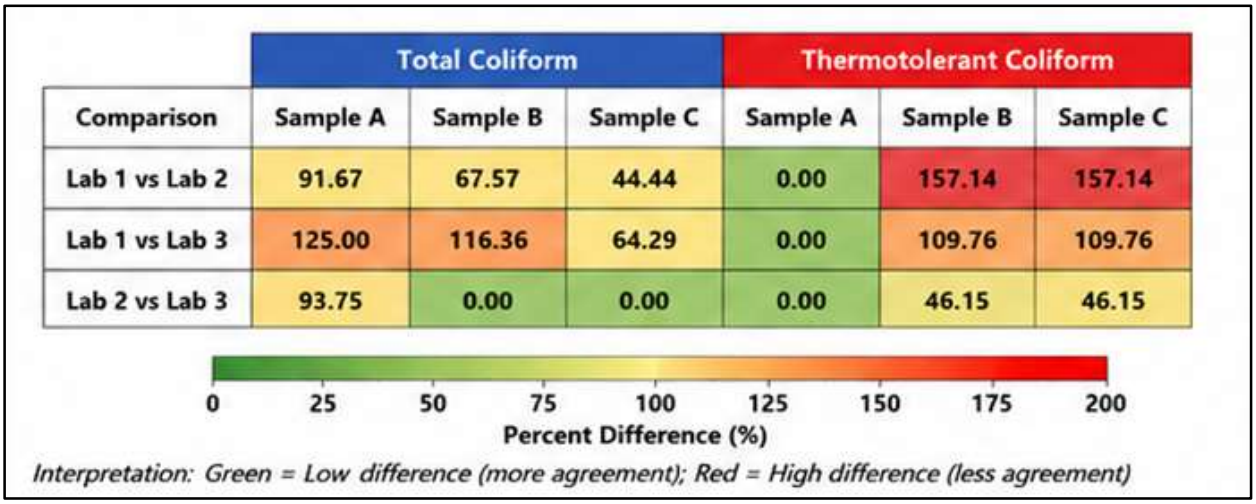
Table 4. Pairwise Percent Difference in Coliform Counts Between Laboratories

Parameter	***Lab 1 vs Lab 2***	***Lab 1 vs Lab 3***	***Lab 2 vs Lab 3***
Total Coliform	135.5%	102%	176%
	102%	102%	0%
	54%	54%	0%
Thermotolerant Coliform	0%	0%	0%
	176%	161%	54%
	176%	161%	54%

Table 4 shows pairwise percent difference analysis that further quantified disagreement among laboratories. Large percent differences were observed for both total and thermotolerant coliform measurements, with several laboratory comparisons exceeding 100%, indicating considerable quantitative variation despite the use of identical samples.

For total coliforms, the greatest disagreement occurred primarily in Sample A, whereas thermotolerant coliform measurements exhibited the largest differences in Samples B and C. These results further support substantial inter-laboratory variability in MTFT-derived quantitative measurements.

Figure 3.
Heatmap of pairwise percent differences among participating laboratories



Note. The heatmap illustrates the pairwise percent differences between laboratory results for Total Coliform and Thermotolerant Coliform measurements. Rows correspond to standardized samples (A–C), while columns represent pairwise laboratory comparisons. Lighter shades indicate lower percent differences and stronger agreement between laboratories, whereas darker shades indicate greater disagreement and increased measurement variability.

Figure 3 shows pairwise percent differences among laboratories presented in a heatmap. Lower percent differences represent stronger agreement, whereas higher percent differences represent greater disagreement.

The heatmap demonstrates that disagreement was not uniformly distributed across all samples. Complete agreement was observed for thermotolerant coliform measurements in Sample A, while greater disagreement was concentrated in higher contamination samples, particularly for thermotolerant coliform analyses. Overall, the heatmap confirms that inter-laboratory variability varied according to both sample and parameter.

Table 5. Trend Consistency of Coliform Counts Across Laboratories

Laboratory Pair	***Total Coliform Trend (A < B < C)***	***Thermotolerant Trend***	***Interpretation***
Lab1 vs Lab2	Partial	No	Inconsistent ordering
Lab1 vs Lab3	Yes	Partial	Moderate agreement
Lab2 vs Lab3	Yes	Yes	Strong agreement

Legend: Yes = trend strictly followed, Partial = trend partially followed, and No = trend not followed.

Table 5 shows the trend consistency analysis which evaluated whether laboratories consistently identified the expected increase in contamination across the standardized samples. Laboratory pair comparisons demonstrated varying levels of agreement, with Laboratories 2 and 3 consistently reproducing the expected contamination trend, whereas comparisons involving Laboratory 1 showed only partial consistency.

These findings shows that although laboratories collectively recognized increasing contamination across samples, consistency in reproducing the complete contamination gradient differed among laboratory pairs.

Laboratory Pair	***ICC (2,1)***	***Interpretation***
Total Coliform	0.081	Poor reliability
Thermotolerant Coliform	0.267	Poor reliability

Table 6. Intraclass Correlation Coefficient for Inter-Laboratory Reliability

Inter-laboratory reliability was evaluated using the Intraclass Correlation Coefficient (ICC). Total coliform measurements demonstrated poor reliability (ICC = 0.081), while thermotolerant coliform measurements showed slightly higher but still poor reliability (ICC = 0.267).

These results display limited agreement in quantitative MTFT measurements among participating laboratories and are consistent with the variability observed in the descriptive, range, and pairwise comp-

arison analyses.

Because this validation study involved only three participating laboratories analyzing three standardized samples, Bland–Altman analysis was not included. The limited sample size would not provide sufficiently stable estimates of bias and limits of agreement; therefore, agreement was evaluated using descriptive variability measures, pairwise comparisons, ICC, and Fleiss' kappa.

Table 7. Inter-Laboratory Agreement in Pass/Fail Classification and Fleiss' Kappa Values

Parameter	***Sample***	***Lab 1***	***Lab 2***	***Lab 3***	***Fleiss' κ***	***Agreement Level***
Total Coliform	A	Fail	Pass	Fail	−0.125	Poor agreement
	B	Fail	Fail	Fail		
	C	Fail	Fail	Fail		
Thermotolerant	A	Pass	Pass	Pass	0.100	Slight agreement
Coliform	B	Pass	Fail	Fail		
	C	Pass	Fail	Fail		

Table 7 shows the agreement in regulatory pass/fail classification which was evaluated using Fleiss' kappa statistics. Total coliform classification demonstrated poor agreement ($\kappa = -0.125$), whereas thermotolerant coliform classification showed slight agreement ($\kappa = 0.100$).

Although the participating laboratories generally detected contamination, differences in quantitative measurements resulted in inconsistent pass/fail classifications for some samples. These findings suggest that inter-laboratory agreement may influence regulatory interpretation when MTFT-derived MPN values are used for compliance assessment.

Table 8. Inferential Analysis of Inter-Laboratory Differences

One-Way ANOVA			
Parameter	***F-value***	***p-value***	***Interpretation***
Total Coliform	2.513	0.161	No significant difference among laboratories
Thermotolerant Coliform	2.316	0.180	No significant difference among laboratories
Kruskal–Wallis			
Parameter	***H-statistic***	***p-value***	***Interpretation***
Total Coliform	3.879	0.144	No significant difference among laboratories

Thermotolerant Coliform	3.374	0.185	No significant difference among laboratories

Exploratory inferential analyses were performed using one-way ANOVA and the Kruskal–Wallis test to evaluate differences among participating laboratories. Neither analysis detected statistically significant differences for total coliform (ANOVA: p = 0.161; Kruskal–Wallis: p = 0.144) nor thermotolerant coliform measurements (ANOVA: p = 0.180; Kruskal–Wallis: p = 0.185).

Although statistical significance was not observed, descriptive statistics, agreement analyses, and reliability measures consistently demonstrated variability among laboratories. The absence of significant inferential results should therefore be interpreted cautiously, considering the limited sample size and exploratory nature of the study.

DISCUSSION

This study evaluated the inter-laboratory reproducibility of the Multiple Tube Fermentation Technique for the enumeration of total and thermotolerant coliforms in standardized drinking water samples analyzed by three accredited drinking water laboratories. Although the participating laboratories collectively identified the expected contamination gradient across the three samples, the findings demonstrated considerable variability in quantitative MTFT-derived Most Probable Number measurements. This variability was reflected in the descriptive statistics, pairwise percent differences, trend consistency analysis, Intraclass Correlation Coefficients, and Fleiss' kappa statistics. Conversely, exploratory inferential analyses using one-way ANOVA and the Kruskal–Wallis test did not detect statistically significant differences among laboratories (Bartlett & Frost, 2008). Taken together, these findings suggest that while MTFT remains effective for detecting the presence of microbiological contamination, reproducibility of quantitative results under routine laboratory conditions remains limited.

The absence of statistically significant differences among laboratories should be interpreted alongside the agreement analyses rather than in isolation. Statistical significance evaluates whether observed differences are unlikely to have occurred by chance, whereas agreement analyses determine the consistency of measurements obtained from identical samples. In the present study, descriptive measures consistently demonstrated substantial variability despite non-significant ANOVA and Kruskal–Wallis results. This apparent discrepancy is likely influenced by the small number of participating laboratories and standardized samples, which limited statistical power (Serdar et al., 2020). Consequently, the descriptive and agreement analyses provide a more informative assessment of inter-laboratory reproducibility than hypothesis testing alone. Similar recommendations have been made in laboratory validation studies, where agreement statistics are considered more appropriate than significance testing when evaluating analytical comparability among laboratories (Ballesté et al., 2019)

These findings address an important evidence gap in drinking water microbiology within the Philippine setting. Although the Multiple Tube Fermentation Technique remains the reference method for regulatory monitoring under the Philippine National Standards for Drinking Water, published evidence evaluating the reproducibility of MTFT among accredited laboratories under routine operational conditions remains limited. Previous local studies have primarily focused on microbial contamination of drinking water and compliance with national standards rather than on analytical agreement between testing laboratories (Bagsik et al., 2019; Lomboy et al., 2016). By evaluating inter-laboratory variability using standardized

samples, the present study provides baseline evidence regarding the consistency of MTFT-derived measurements among accredited laboratories and highlights aspects of the method that may benefit from strengthened quality assurance practices (Sreya et al., 2024).

The descriptive analyses demonstrated that total coliform measurements followed the expected contamination gradient, with mean MPN values increasing progressively from Sample A to Sample C. This shows that the participating laboratories consistently detected increasing levels of bacterial contamination in the standardized samples. Nevertheless, the degree of agreement among laboratories varied considerably. Sample A exhibited the highest coefficient of variation (104.57%), whereas Sample C demonstrated substantially lower relative variability (28.59%). These findings suggest that inter-laboratory reproducibility was poorest at lower bacterial concentrations and improved with increasing contamination levels.

Greater variability at low bacterial densities is consistent with the statistical characteristics of the Most Probable Number method. Unlike direct enumeration techniques, MTFT estimates bacterial concentration using probability models based on patterns of positive and negative fermentation tubes. Consequently, small differences in the number of positive tubes may produce disproportionately large differences in calculated MPN values, particularly when bacterial concentrations approach the lower detection range (Martini et al., 2024; Schmidt et al., 2019). This probabilistic nature of the method contributes to wider confidence intervals and increased uncertainty at low concentrations, even when identical samples are analyzed (Tambi et al., 2023). The wide confidence intervals observed for Sample A in the present study therefore, likely reflect an inherent limitation of MPN estimation rather than analytical error alone.

As contamination levels increased, inter-laboratory variability decreased for total coliform measurements. Higher bacterial concentrations typically produce more pronounced fermentation reactions, including clearer gas production and turbidity, thereby reducing ambiguity during result interpretation. Similar observations have been reported in previous evaluations of culture-based water quality methods, where stronger microbial growth improves consistency among analysts by minimizing subjective interpretation of fermentation endpoints (Gunawardane & Priyadarshani, 2025). Although reproducibility improved for Sample C, interpretation remained constrained by the MTFT procedure's upper detection limit. Several laboratories reported identical maximum MPN values (>8.0 MPN/100 mL), suggesting that bacterial concentrations exceeded the quantifiable range of the dilution scheme employed. Such censoring effects may reduce the apparent variability among highly contaminated samples because true differences beyond the upper detection limit cannot be distinguished, an issue documented in culture-based environmental microbiology where counts beyond the upper reporting limit are conventionally reported as too numerous to count (Chik et al., 2018).

A different pattern was observed for thermotolerant coliform measurements. Complete agreement among all participating laboratories was obtained for Sample A, resulting in a coefficient of variation of 0%. However, substantial variability emerged in Samples B and C, both of which exhibited coefficients of variation of approximately 86%. These findings indicate that reproducibility varied not only with contamination level but also with the microbial parameter measured. The greater variability observed for thermotolerant coliforms at higher concentrations may reflect cumulative effects of procedural differences during incubation, media preparation, inoculum handling, and interpretation of positive reactions. Although all participating laboratories followed standardized MTFT procedures, minor operational differences can contribute to variation in MPN-based measurements, as the technique relies on sequential

presumptive and confirmatory reactions that are sensitive to incubation conditions and analyst interpretation (Schmidt et al., 2010; Tambi et al., 2023).

The range analysis and pairwise percent difference analysis further demonstrated the magnitude of inter-laboratory variability. Percent differences reached as high as 176% between laboratory pairs, indicating that identical standardized samples occasionally produced markedly different quantitative estimates. Although the contamination trend was generally consistent, these discrepancies demonstrate that laboratories may produce substantially different MPN values when evaluating the same water sample. Fleiss' kappa statistics, applied to categorical classifications, further quantified the extent of agreement beyond chance, with kappa values indicating only modest concordance that was sensitive to the magnitude of MPN category differences (Hallgren, 2012; McHugh, 2012). Similar findings have been reported in international inter-laboratory proficiency testing programs, where methodological implementation, analyst experience, reagent quality, and interpretation criteria contributed to measurable differences in reported microbiological results despite the use of standardized testing procedures (Ballesté et al., 2019; Sreya et al., 2024)

Collectively, these findings indicate that the Multiple Tube Fermentation Technique remains efficient for detecting microbiological contamination and identifying general contamination trends. However, the quantitative reproducibility of MTFT-derived MPN values appears to be influenced by both the method's probabilistic characteristics and routine operational factors inherent to laboratory testing (Martini et al., 2024; Schmidt et al., 2010). Consequently, quantitative MTFT measurements should be interpreted with appropriate consideration of analytical variability, particularly when comparing results generated by different laboratories or when measured concentrations are close to regulatory decision thresholds (Bartlett & Frost, 2008; Chik et al., 2018).

Beyond descriptive measures of variability, the reliability analyses provided further evidence regarding the consistency of MTFT-derived measurements among participating laboratories. The Intraclass Correlation Coefficients demonstrated poor reliability for both total coliform (ICC = 0.081) and thermotolerant coliform (ICC = 0.267) measurements, demonstrating that variability among laboratories exceeded the consistency expected from repeated analyses of identical samples (Liljequist et al., 2019). Although the ICC for thermotolerant coliforms was slightly higher than that for total coliforms, both values remained below the 0.50 threshold conventionally used to denote poor reliability, indicating limited reproducibility under routine laboratory conditions (Liljequist et al., 2019).

Poor ICC values do not necessarily mean that laboratories failed to detect contamination; rather, they suggest that laboratories differed in the quantitative values assigned to the same samples. This distinction is particularly important when interpreting MTFT-derived MPN estimates. The technique is inherently based on probabilistic estimation rather than direct microbial enumeration, meaning that relatively small differences in fermentation tube outcomes can produce substantial differences in calculated MPN values (Martini et al., 2024). Consequently, analytical variability may accumulate throughout the sequential stages of presumptive and confirmatory testing, resulting in reduced agreement among laboratories despite adherence to standardized procedures (Ballesté et al., 2019; Tambi et al., 2023).

The trend consistency analysis provided additional insight into laboratory performance. Most laboratory comparisons successfully identified the expected increase in contamination across the standardized samples, with the strongest agreement observed between Laboratories 2 and 3. Comparisons involving Laboratory 1, however, demonstrated only partial consistency. These findings suggest that MTFT remains reliable for identifying the general presence and progression of microbiological contamination but may be

less consistent when ranking contamination severity based on quantitative MPN estimates (Tambi et al., 2023). For routine surveillance, this distinction is important because public health interventions often depend primarily on identifying contamination rather than obtaining identical numerical counts across laboratories (Gunnarsdottir et al., 2019). Nevertheless, quantitative consistency remains essential for evaluating treatment efficiency, monitoring temporal trends, and comparing results generated by different laboratories (Tsaridou & Karabelas, 2021).

Agreement analysis using Fleiss' kappa further demonstrated that quantitative variability extended to regulatory interpretation. Poor agreement was observed for total coliform pass/fail classification ($\kappa = -0.125$), while thermotolerant coliform classification demonstrated only slight agreement ($\kappa = 0.100$). On the conventional Landis and Koch interpretive scale, κ values below 0 indicate poor agreement, 0.00–0.20 indicate slight agreement, and 0.21–0.40 indicate fair agreement, placing both observed values within the poor-to-slight agreement range. These findings indicate that identical water samples may receive different regulatory classifications depending on the laboratory performing the analysis. Although such discrepancies were most apparent for samples near regulatory decision thresholds, they highlight the practical implications of analytical variability because regulatory compliance, corrective actions, and public health responses frequently rely on categorical rather than continuous microbiological results (Borchardt et al., 2021; Pluym et al., 2024).

The combination of poor ICC values and limited Fleiss' kappa agreement suggests that inter-laboratory variability influenced both quantitative enumeration and regulatory interpretation. These findings are consistent with international proficiency testing studies reporting that differences in analyst interpretation, incubation conditions, media quality, and implementation of standardized protocols contribute to measurable differences among accredited laboratories (Ballesté et al., 2019; Sreya et al., 2024). Accreditation therefore, represents an important foundation for quality assurance but does not completely eliminate analytical variability, particularly for probabilistic microbiological methods such as MTFT (Borchardt et al., 2021).

The findings of this study have important implications for drinking water quality monitoring and regulatory decision-making in the Philippines. Under the Philippine National Standards for Drinking Water, accredited laboratories play a central role in determining compliance with microbiological standards for public water systems, water refilling stations, and other drinking water providers (Alcalde & Cataytay, 2025; Magwilang et al., 2023). Consequently, consistency among laboratories is essential to ensure that regulatory decisions are based on comparable analytical results.

Although the participating laboratories recognized the expected contamination gradient, the observed variability in quantitative MPN values and pass/fail classifications suggests that laboratory-specific differences may influence regulatory outcomes, particularly for samples with bacterial concentrations near compliance thresholds (Tambi et al., 2023). Under these circumstances, relatively small analytical differences may alter whether a water sample is classified as compliant or non-compliant, potentially affecting public health interventions, corrective actions, and regulatory enforcement (Rodríguez et al., 2025; Sreya et al., 2024). Similar concerns have been reported internationally, where inconsistent microbiological classifications have influenced drinking water surveillance and risk management programs (Gunnarsdottir et al., 2019; Tsaridou & Karabelas, 2021).

These findings should not be interpreted as evidence that MTFT is unsuitable for regulatory monitoring. Rather, they highlight the importance of strengthening quality assurance systems to minimize avoidable analytical variability (Gunawardane & Priyadarshani, 2025). The MTFT remains the internationally

recognized reference method in many regulatory frameworks because of its ability to detect low concentrations of indicator organisms using relatively simple laboratory infrastructure (Pluym et al., 2024). However, because the method relies on probability-based estimation, maintaining consistency requires rigorous implementation of standardized operating procedures, regular equipment calibration, careful media preparation, controlled incubation conditions, and continuous competency assessment of laboratory personnel (Borchardt et al., 2021; Tsaridou & Karabelas, 2021).

Within the Philippine context, these findings support continued investment in inter-laboratory comparison programs and external quality assessment schemes. Participation in proficiency testing enables laboratories to identify systematic analytical differences, evaluate long-term performance, and improve comparability across institutions (Kostyanev et al., 2025; Molina-Castro et al., 2021). Such programs are particularly important for decentralized laboratory networks, where differences in personnel experience, laboratory infrastructure, and routine operational practices may contribute to analytical variability despite accreditation (Magwilang et al., 2023). Strengthening these quality assurance initiatives, through the systematic adoption of standards such as the EMMI guidelines for environmental microbiology reporting (Borchardt et al., 2021), would improve confidence in microbiological monitoring and support more consistent implementation of national drinking water standards (Alcalde & Cataytay, 2025).

The present findings are consistent with broader international observations regarding the strengths and limitations of culture-based microbiological methods for drinking water monitoring. Despite the development of molecular and instrument-based techniques, the Multiple Tube Fermentation Technique remains widely used because it is inexpensive, well-standardized, and compatible with routine laboratory infrastructure in many low- and middle-income countries (Brown et al., 2020). Its continued regulatory acceptance reflects its practicality rather than absolute analytical precision, and the same rationale underpins its retention as a benchmark against which newer enzyme-based and chromogenic methods, such as Colilert, continue to be evaluated (Tambi et al., 2023).

Nevertheless, the probabilistic nature of the MPN approach presents an inherent limitation for quantitative reproducibility. Unlike direct enumeration methods, MTFT estimates bacterial concentration from statistical probability distributions derived from tube positivity patterns (Dinser et al., 2022). Consequently, some degree of measurement uncertainty is unavoidable, particularly at low bacterial concentrations or when measured values approach the upper detection limit of the selected dilution series (Martini et al., 2024). The present study reinforces these theoretical limitations by demonstrating that quantitative variability persisted despite the use of standardized samples analyzed by accredited laboratories.

Recent advances in microbiological water quality monitoring have introduced alternative methods, including membrane filtration with chromogenic media, quantitative polymerase chain reaction, and other rapid molecular approaches that provide improved analytical precision and shorter turnaround times (Pluym et al., 2024; Tambi et al., 2023). However, widespread implementation of these technologies remains challenging in many resource-limited settings because of higher equipment costs, specialized technical expertise, and regulatory requirements that continue to recognize classical culture-based methods as reference standards (Pluym et al., 2024). Consequently, MTFT is expected to remain an essential component of drinking water monitoring programs for the foreseeable future (Tsaridou & Karabelas, 2021).

Rather than replacing MTFT, future research should focus on improving its reproducibility through enhanced quality management and methodological optimization. Multi-center validation studies involving

larger numbers of accredited laboratories, broader concentration ranges, and repeated testing over multiple sampling periods would provide more robust estimates of inter-laboratory agreement (Allkja et al., 2021; Sreya et al., 2024. Further investigations comparing MTFT with membrane filtration and molecular detection methods under routine operational conditions would also help identify complementary approaches that improve microbiological surveillance while remaining feasible for implementation within existing regulatory frameworks (Tambi et al., 2023). Such efforts would contribute to strengthening evidence-based drinking water monitoring and support progress toward Sustainable Development Goal 6, which emphasizes universal access to safe and reliable drinking water; as Bain et al. note, SDG indicator 6.1.1 specifically tracks the use of safely managed drinking water services that are free from faecal contamination (Bain et al., 2020), and recent global assessments confirm that approximately two billion people still lacked access to safely managed drinking water in 2020, with countries off-track to meet SDG 6.1 by 2030 (Arora & Mishra, 2022; (Rajapakse et al., 2023).

This study provides one of the first evaluations of inter-laboratory reproducibility of the Multiple Tube Fermentation Technique among accredited drinking water laboratories in the Philippines using standardized drinking water samples. By integrating descriptive statistics, range and pairwise variability analyses, trend consistency, Intraclass Correlation Coefficients, Fleiss' kappa statistics, and exploratory inferential analyses, the study offers a comprehensive assessment of MTFT performance under routine operational conditions (Liljequist et al., 2019). Collectively, these complementary analytical approaches provide a broader understanding of laboratory agreement than reliance on a single statistical measure and establish baseline evidence that may guide future laboratory quality assurance initiatives within the Philippine drinking water monitoring system (Borchardt et al., 2021).

Despite the strengths of this study, several limitations should be considered when interpreting the findings. First, the investigation included only three Department of Health-accredited drinking water laboratories and three standardized drinking water samples representing low, medium, and high contamination levels. Although this design allowed direct comparison of laboratory performance under routine operational conditions, the relatively small number of participating laboratories and samples reduced the statistical power of the inferential analyses and limited the precision of inter-laboratory agreement estimates (Serdar et al., 2020). Consequently, the absence of statistically significant differences among laboratories should not be interpreted as evidence of complete analytical equivalence but rather as reflecting the exploratory nature of the investigation and the limited sample size (Serdar et al., 2020). In addition, the use of standardized *Escherichia coli*-inoculated sterile water minimized extraneous sources of variation and enhanced experimental control; however, it does not fully represent the complexity of natural drinking water matrices, where background microorganisms, suspended solids, organic matter, and other physicochemical characteristics may influence MTFT performance.

Second, several measurements reached the upper reporting limit of the dilution scheme (>8.0 MPN/100 mL), resulting in a ceiling effect that limited the ability to distinguish between highly contaminated samples. This constraint may have underestimated the true magnitude of inter-laboratory variability at higher bacterial concentrations and is an inherent limitation of probability-based enumeration methods (Martini et al., 2024; Takada & Kato, 2023). As emphasized by Takada and Kato (2023), censored microbiological water quality data require specialized statistical treatment because measurements reported only above or below analytical detection limits contain important uncertainty that may influence subsequent statistical interpretation.

Third, although all participating laboratories were accredited and followed standardized MTFT procedures, the study did not independently evaluate laboratory-specific operational factors, including analyst experience, media preparation techniques, incubation equipment performance, environmental conditions, or internal quality control practices. These variables are recognized contributors to inter-laboratory variability and may partially explain the differences observed among laboratories despite the implementation of standardized protocols (Allkja et al., 2021; Tambi et al., 2023). Likewise, Allkja et al. (2021) demonstrated through an international ring-trial evaluation that between-laboratory operational differences remain an important source of reproducibility variation even when harmonized analytical procedures are employed. Future investigations incorporating laboratory process audits or controlled evaluations of procedural variables would provide a more comprehensive understanding of the sources of analytical variation and support the development of targeted quality improvement strategies (Borchardt et al., 2021).

Finally, Bland–Altman analysis was not included because the limited number of participating laboratories and standardized samples would not provide sufficiently robust estimates of bias and limits of agreement. Instead, inter-laboratory reproducibility was evaluated using complementary statistical approaches, including descriptive statistics, pairwise percent difference analysis, trend consistency assessment, Intraclass Correlation Coefficients (ICC), Fleiss' kappa statistics, and exploratory inferential analyses. Although these complementary methods provided a comprehensive assessment of reproducibility, future studies involving a larger number of laboratories, repeated analyses across multiple testing periods, broader contamination ranges, and more diverse drinking water matrices would improve the generalizability of the findings and provide more robust estimates of inter-laboratory agreement (Molina-Castro et al., 2021; Sreya et al., 2024). Nevertheless, this study provides valuable baseline evidence regarding the reproducibility of the Multiple Tube Fermentation Technique among accredited Philippine drinking water laboratories operating under routine conditions and contributes to the limited body of local evidence supporting the continuous improvement of microbiological drinking water monitoring.

Even so, it is important to note that the findings of this study highlight several opportunities for future research to improve the reproducibility and reliability of microbiological drinking water analysis. Larger multi-center validation studies involving a broader network of accredited laboratories should be conducted to establish national benchmarks for MTFT reproducibility and evaluate analytical performance across different geographic regions and laboratory settings (Kostyanev et al., 2025; Sreya et al., 2024). Expanding the number of participating laboratories, increasing the range of contamination levels, and incorporating repeated testing over multiple sampling periods would provide more robust estimates of inter-laboratory reproducibility and strengthen the evidence supporting routine microbiological water quality monitoring. In addition, evaluating laboratories from different regions of the Philippines would help identify geographic or operational factors that may contribute to analytical variability and provide stronger evidence regarding the consistency of MTFT implementation within the national drinking water monitoring system (Alcalde & Cataytay, 2025).

Future investigations should also evaluate the influence of specific analytical and operational factors, including media preparation, incubation conditions, analyst interpretation, laboratory quality assurance practices, and quality control procedures, on measurement variability. Understanding the relative contribution of these factors would facilitate the development of targeted interventions to improve inter-laboratory consistency and strengthen laboratory standardization (Borchardt et al., 2021; Tambi et al., 2023). Furthermore, adopting standardized reporting frameworks, such as the Environmental

Microbiology Minimum Information (EMMI) guidelines, may improve the consistency and completeness of methodological reporting across laboratories and proficiency testing programs (Borchardt et al., 2021). Comparative evaluations of the Multiple Tube Fermentation Technique with alternative microbiological methods, including membrane filtration using chromogenic media, enzyme-substrate assays, and molecular approaches such as quantitative polymerase chain reaction (qPCR), are likewise recommended (Brown et al., 2020; Tambi et al., 2023). Future studies should include a wider range of drinking water matrices, including groundwater, surface water, treated distribution water, and water from refilling stations, to determine whether these methods provide improved analytical precision and reproducibility under diverse environmental conditions. While these alternative techniques offer advantages in analytical sensitivity and turnaround time, their feasibility, cost-effectiveness, and regulatory applicability should be evaluated within the context of routine drinking water surveillance in resource-limited settings (Pluym et al., 2024; Tsaridou & Karabelas, 2021). Findings from such comparative investigations may help inform future revisions of national monitoring guidelines while maintaining compatibility with existing regulatory frameworks (Thavarajah et al., 2020).

Future research may also benefit from applying advanced statistical approaches specifically developed for probability-based microbiological enumeration, including Bayesian estimation methods, hierarchical modeling, and improved maximum-likelihood estimators (Martini et al., 2024). Recent work has demonstrated that maximum-likelihood and exact Poisson-based estimators can improve the precision of bacterial concentration estimates by integrating information across multiple dilution levels, thereby providing a stronger statistical foundation for uncertainty quantification in MPN-based analyses (Martini et al., 2024). The application of these approaches may improve the interpretation of MTFT-derived MPN estimates, particularly for measurements obtained near analytical detection limits or regulatory decision thresholds (Martini et al., 2024; Serdar et al., 2020). Collectively, these future research directions will contribute to improving the reproducibility, comparability, and regulatory utility of microbiological drinking water monitoring while strengthening evidence-based public health decision-making in the Philippines.

CONCLUSION

This study evaluated the inter-laboratory reproducibility of the Multiple Tube Fermentation Technique (MTFT) for enumerating total and thermotolerant coliforms in standardized drinking water samples, analyzed by three accredited drinking water laboratories. The findings demonstrated that although MTFT consistently identified the expected contamination gradient across samples, quantitative reproducibility among laboratories was limited under routine operational conditions. Substantial inter-laboratory variability was observed through descriptive statistics, range and pairwise percent difference analyses, trend consistency assessment, Intraclass Correlation Coefficients (ICC), and Fleiss' kappa statistics, indicating that identical samples may yield different quantitative measurements and, in some cases, different regulatory classifications depending on the laboratory performing the analysis.

While exploratory inferential analyses using one-way ANOVA and the Kruskal–Wallis test did not demonstrate statistically significant differences among laboratories, the descriptive and agreement analyses consistently show a pattern that suggests variability in MTFT-derived Most Probable Number (MPN) measurements. These findings suggest that the lack of statistical significance should be interpreted with caution given the limited sample size and should not be taken as evidence of complete analytical equivalence among participating laboratories.

The study further demonstrated that MTFT remains an effective qualitative tool for detecting microbiological contamination and identifying general contamination trends. However, its quantitative reproducibility is influenced by the probabilistic nature of MPN estimation and routine operational factors that contribute to inter-laboratory variation. Consequently, quantitative MTFT results should be interpreted with appropriate consideration of analytical uncertainty, particularly when measurements are close to regulatory decision thresholds or when results generated by different laboratories are compared.

These findings underscore the importance of strengthening quality assurance practices within accredited drinking water laboratories. Continued participation in proficiency testing and inter-laboratory comparison programs, harmonization of analytical procedures, rigorous quality control, and ongoing competency development of laboratory personnel are essential to improving inter-laboratory comparability and confidence in microbiological water quality assessments.

Although this study was limited to three accredited laboratories and three standardized drinking water samples, it provides baseline evidence on the inter-laboratory reproducibility of MTFT under routine operational conditions in the Philippines. The findings contribute to the limited body of local evidence on laboratory agreement and provide a foundation for future multi-center validation studies involving more laboratories, broader contamination ranges, and comparative evaluations with alternative microbiological methods. Strengthening the reproducibility and comparability of microbiological water testing will support more consistent implementation of the Philippine National Standards for Drinking Water, improve evidence-based regulatory decision-making, and contribute to the broader goal of ensuring safe drinking water and advancing Sustainable Development Goal 6.

REFERENCES

1. Alcalde, A. L., & Cataytay, R. (2025). Assessment of bacterial contamination in freshwater stations: Implications for Drinking Water Safety based on PNSDW 2017 standards. *Psychology and Education a Multidisciplinary Journal*, *38*(7), 695–700. https://doi.org/10.70838/pemj.380701
2. Allkja, J., Van Charante, F., Aizawa, J., Reigada, I., Guarch-Pérez, C., Vazquez-Rodriguez, J. A., Cos, P., Coenye, T., Fallarero, A., Zaat, S. a. J., Felici, A., Ferrari, L., Azevedo, N. F., Parker, A. E., & Goeres, D. M. (2021). Interlaboratory study for the evaluation of three microtiter plate-based biofilm quantification methods. *Scientific Reports*, *11*(1), 13779. https://doi.org/10.1038/s41598-021-93115-w
3. Arora, N. K., & Mishra, I. (2022). Sustainable development goal 6: Global Water Security. *Environmental Sustainability*, *5*(3), 271–275. https://doi.org/10.1007/s42398-022-00246-5
4. Bagsik, C. T., Reyes, G. C. P., & Sulibit, K. E. (2019). Novel automated microbiological analysis in meeting Philippine drinking water standards. *Water Practice & Technology*, *14*(2), 444–448. https://doi.org/10.2166/wpt.2019.033
5. Bain, R., Johnston, R., & Slaymaker, T. (2020). Drinking water quality and the SDGs. *Npj Clean Water*, *3*(1). https://doi.org/10.1038/s41545-020-00085-z
6. Ballesté, E., Belanche-Muñoz, L. A., Farnleitner, A. H., Linke, R., Sommer, R., Santos, R., Monteiro, S., Maunula, L., Oristo, S., A, A. T., Stange, C., & Blanch, A. R. (2019). Improving the identification of the source of faecal pollution in water using a modelling approach: From multi-source to aged and diluted samples. *Water Research*, *171*, 115392. https://doi.org/10.1016/j.watres.2019.115392

7. Bartlett, J. W., & Frost, C. (2008). Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound in Obstetrics and Gynecology*, *31*(4), 466–475. https://doi.org/10.1002/uog.5256
8. Borchardt, M. A., Boehm, A. B., Salit, M., Spencer, S. K., Wigginton, K. R., & Noble, R. T. (2021). The Environmental Microbiology Minimum Information (EMMI) Guidelines: QPCR and DPCR Quality and Reporting for Environmental Microbiology. *Environmental Science & Technology*, *55*(15), 10210–10223. https://doi.org/10.1021/acs.est.1c01767
9. Bortolaia, V., Pedersen, S. K., & Hendriksen, R. S. (2021). *The 26th Proficiency Testing, 2019 – Enterococci, Staphylococci and Escherichia coli*.
10. Boubetra, A., Nestour, F. L., Allaert, C., & Feinberg, M. (2011). Validation of Alternative Methods for the Analysis of Drinking Water and Their Application to Escherichia coli. *Applied and Environmental Microbiology*, *77*(10), 3360–3367. https://doi.org/10.1128/aem.00020-11
11. Brown, J., Bir, A., & Bain, R. E. S. (2020). Novel methods for global water safety monitoring: comparative analysis of low-cost, field-ready E. coli assays. *Npj Clean Water*, *3*(1). https://doi.org/10.1038/s41545-020-0056-8
12. Cambarihan, S. D., Patricio, E. R. P., & Lumogdang, L. P. (2022). Detection and enumeration of coliforms in drinking water sources in the selected barangay in Santa Maria, Davao Occidental Philippines. *Asian Journal of Biology*, 1–7. https://doi.org/10.9734/ajob/2022/v15i130227
13. Caneos, W. G., Macapando, S., Clavano, C. J., & Abragan, D. F. (2024). Assessment of Physico-chemical and Microbiological Characteristics of Water from Lake Dapao, Lanao del Sur, Philippines. *Research Square*. https://doi.org/10.21203/rs.3.rs-3994511/v1
14. Chik, A. H. S., Schmidt, P. J., & Emelko, M. B. (2018). Learning something from nothing: the critical importance of rethinking microbial non-detects. *Frontiers in Microbiology*, *9*, 2304. https://doi.org/10.3389/fmicb.2018.02304
15. Dinser, S., Aysun, M. T. H., & Özdenefe, M. S. (2022). Water quality - new perspectives. In *IntechOpen eBooks*. https://doi.org/10.5772/intechopen.100757
16. Gadrich, T., Kuselman, I., Pennecchi, F. R., Hibbert, D. B., Semenova, A. A., Cheow, P. S., & Naidenko, V. N. (2022). Interlaboratory comparison of the intensity of drinking water odor and taste by two-way ordinal analysis of variation without replication. *Journal of Water and Health*, *20*(6), 1005–1016. https://doi.org/10.2166/wh.2022.060
17. Gómez, L., Brandão, J., Navarro, A., Martines, A., Sanseverino, I., Porcel-Rodríguez, E., Marinov, D., Ableitner, O., Allerberger, F., Atsma, A., Briancesco, R., Cenov, A., Číchová, M., Coccia, A. M., Della Libera, S., De Maeyer, K., De Vries, H. J., Elsinga, G., Heijnen, L., . . . Lettieri, T. (2025). Application of a real-time reverse transcription polymerase chain reaction for rapid detection of Escherichia coli in drinking water: an EU representative study. *Environmental Research*, *279*(Pt 2), 121786. https://doi.org/10.1016/j.envres.2025.121786
18. Gunawardane, M. M., & Priyadarshani, K. P. R. (2025). Decoding the multiple tube fermentation method used in coliform testing. *JMTR.*, *10*(1), 42–49. https://doi.org/10.4038/jmtr.v10i1.79
19. Gunnarsdottir, M. J., Gardarsson, S. M., Figueras, M. J., Puigdomènech, C., Juárez, R., Saucedo, G., Arnedo, M. J., Santos, R., Monteiro, S., Avery, L., Pagaling, E., Allan, R., Abel, C., Eglitis, J., Hambsch, B., Hügler, M., Rajkovic, A., Smigic, N., Udovicki, B., . . . Hunter, P. (2019). Water safety plan enhancements with improved drinking water quality detection techniques. *The Science of the Total Environment*, *698*, 134185. https://doi.org/10.1016/j.scitotenv.2019.134185

20. Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023
21. Hendriksen, R. S., Karlsmose, S., Jensen, A. B., & Aarestrup, F. M. (2014). *The External Quality Assurance System of the WHO Global Foodborne Infections Network: Year 2012*. Welcome to DTU Research Database. https://orbit.dtu.dk/en/publications/76a2ccd7-d7c8-4ab8-bc03-e32af6490ab0
22. Hiura, S., Abe, H., Koyama, K., & Koseki, S. (2021). Bayesian Generalized Linear Model for Simulating Bacterial Inactivation/Growth considering variability and uncertainty. *Frontiers in Microbiology*, *12*, 674364. https://doi.org/10.3389/fmicb.2021.674364
23. Houck, K. M., Terán, E., Ochoa, J., Zapata, G. N., Gomez, A. M., Parra, R., Dvorquez, D., Stewart, J. R., Bentley, M. E., & Thompson, A. L. (2019). Drinking water improvements and rates of urinary and gastrointestinal infections in Galápagos, Ecuador: Assessing household and community factors. *American Journal of Human Biology*, *32*(1), e23358. https://doi.org/10.1002/ajhb.23358
24. International Organization for Standardization. (1994). *ISO 5725-1:1994*. ISO. https://www.iso.org/standard/11833.html
25. Kato, H., Maruyama, S., Merino, G. J., & Juban, N. R. (2021). Community-Driven Evaluation of E. coli Levels in Domestic-Use Underground Water Sources in Angono, Philippines. *Acta Medica Philippina*. https://doi.org/10.47895/amp.vi0.3088
26. Kilic, S. (2015). Kappa test. *Journal of Mood Disorders*, *5*(3), 142. https://doi.org/10.5455/jmood.20150920115439
27. Kostyanev, T., Al-Mir, H., Prifti, K., Rasmussen, L. B., Abegaz, F. A., Santos, P. T. D., Kwon, S., Prathan, R., Tongkum, T. L., Chanchaithong, P., Santanirand, P., Kamjumphol, W., Mogeni, O. D., Guarnacci, T., Poudyal, N., Holm, M., Paveenkittiporn, W., Chuanchuen, R., & Hendriksen, R. S. (2025). External quality assurance (EQA) network in South and South-East Asia: experience and results from an international EQA programme in One Health sector reference laboratories. *Journal of Antimicrobial Chemotherapy*, *80*(4), 1037–1046. https://doi.org/10.1093/jac/dkaf032
28. Lee, C., Kim, J. Y., Lee, W. I., Nelson, K. L., Yoon, J., & Sedlak, D. L. (2008). Bactericidal Effect of Zero-Valent Iron Nanoparticles on Escherichia coli. *Environmental Science & Technology*, *42*(13), 4927–4933. https://doi.org/10.1021/es800408u
29. Liljequist, D., Elfving, B., & Roaldsen, K. S. (2019). Intraclass correlation – A discussion and demonstration of basic features. *PLoS ONE*, *14*(7), e0219854. https://doi.org/10.1371/journal.pone.0219854
30. Lomboy, M., De Dios, J. R., Magtibay, B., Quizon, R., Molina, V., Fadrilan-Camacho, V., See, J., Enoveso, A., Barbosa, L., & Agravante, A. (2016). Updating national standards for drinking-water: a Philippine experience. *Journal of Water and Health*, *15*(2), 288–295. https://doi.org/10.2166/wh.2016.177
31. Magwilang, E. B., Paredes, A. L. Y., Armas, F. C., Bugnay, H. G. P., & Dagupen, R. D. (2023). Physico-chemical and microbiological assessment of domestic water supply in Bontoc, Mountain Province, Philippines. *H2Open Journal*, *6*(2), 242–255. https://doi.org/10.2166/h2oj.2023.019
32. Martini, K. M., Boddu, S. S., Nemenman, I., & Vega, N. M. (2024). Maximum likelihood estimators for colony-forming units. *Microbiology Spectrum*, *12*(9), e0394623. https://doi.org/10.1128/spectrum.03946-23

33. McBride, G. B., McWhirter, J. L., & Dalgety, M. H. (2004). Uncertainty in most probable number calculations for microbiological assays. *PubMed*, *86*(5), 1084–1088. https://pubmed.ncbi.nlm.nih.gov/14632414
34. McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 276–282. https://doi.org/10.11613/bm.2012.031
35. Molina-Castro, G., Venegas-Padilla, J., Molina-Marcia, J., Scarioni, L., & Calderón-Jiménez, B. (2021). Improving the quality control of drinking water in Nicaragua through proficiency testing in a metrological multilateral cooperation project. *Scientific Reports*, *11*(1), 16853. https://doi.org/10.1038/s41598-021-96230-w
36. Pichel, N., De Souza, F. H., Sabogal-Paz, L., Shah, P., Adhikari, N., Pandey, S., Shrestha, B., Gaihre, S., Pineda-Marulanda, D., Hincapie, M., Luwe, K., Kumwenda, S., Aguilar-Conde, J., Cortes, M., Hamilton, J., Byrne, J., & Fernandez-Ibañez, P. (2023). Field-testing solutions for drinking water quality monitoring in low- and middle-income regions and case studies from Latin American, African and Asian countries. *Journal of Environmental Chemical Engineering*, *11*(6), 111180. https://doi.org/10.1016/j.jece.2023.111180
37. Pluym, T., Waegenaar, F., De Gusseme, B., & Boon, N. (2024). Microbial drinking water monitoring now and in the future. *Microbial Biotechnology*, *17*(7), e14532. https://doi.org/10.1111/1751-7915.14532
38. Qiu, Y., Han, C., & Dong, T. (2024). Comparison of methods between the multitube fermentation method and the filter membrane method for Escherichia coli in drinking water. *Academic Journal of Science and Technology*, *9*(3), 134–137. https://doi.org/10.54097/dgs74108
39. Rajapakse, J., Otoo, M., & Danso, G. (2023). Progress in delivering SDG6: Safe water and sanitation. *Cambridge Prisms Water*, *1*. https://doi.org/10.1017/wat.2023.5
40. Rodríguez, F. A., Vargas, M. N. G., & Riquelme, S. S. (2025). Microbiological proficiency testing: Establishing a fixed standard deviation for proficiency assessment. *Measurement Sensors*, *38*, 101855. https://doi.org/10.1016/j.measen.2025.101855
41. Saavedra-Ruiz, A., & Resto-Irizarry, P. J. (2025). A portable UV-LED/RGB sensor for Real-Time bacteriological water quality monitoring using ML-Based MPN estimation. *Biosensors*, *15*(5), 284. https://doi.org/10.3390/bios15050284
42. Sari, S. Y. I., Faisal, M., Raksanagara, A. S., Agustian, D., & Rusmil, K. (2020). Water Quality and Factors Associated with Compliance of Drinking Water Refilling Stations as a Choice for Middle–Low Urban Households in Developing Countries. *Journal of Water and Environment Technology*, *18*(1), 27–36. https://doi.org/10.2965/jwet.19-037
43. Sartory, D. (2007). Review Validation, verification and comparison: Adopting new methods in water microbiology[#] *Water SA*, *31*(3). https://doi.org/10.4314/wsa.v31i3.5211
44. Schmidt, P. J., Emelko, M. B., & Reilly, P. M. (2010). Quantification of analytical recovery in particle and microorganism enumeration methods. *Environmental Science & Technology*, *44*(5), 1705–1712. https://doi.org/10.1021/es902237f
45. Schmidt, P. J., Emelko, M. B., & Thompson, M. E. (2019). Recognizing structural nonidentifiability: when experiments do not provide information about important parameters and misleading models can still have great fit. *Risk Analysis*, *40*(2), 352–369. https://doi.org/10.1111/risa.13386

46. Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2020). Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia Medica*, *31*(1), 27–53. https://doi.org/10.11613/bm.2021.010502
47. Sikder, M., Naumova, E. N., Ogudipe, A. O., Gomez, M., & Lantagne, D. (2021). Fecal Indicator Bacteria Data to Characterize Drinking Water Quality in Low-Resource Settings: Summary of current practices and recommendations for improving validity. *International Journal of Environmental Research and Public Health*, *18*(5), 2353. https://doi.org/10.3390/ijerph18052353
48. Sosa-Moreno, A., Lee, G. O., Wu, Z., Fanny, S. A., Trueba, G., Cooper, P. J., Levy, K., & Eisenberg, J. N. S. (2025). Comparing the Power of Low vs High-Precision Methods for Measuring E. coli in Drinking Water in Low-Resource Settings. *ACS ES&T Water*, *5*(5), 2244–2254. https://doi.org/10.1021/acsestwater.4c01117
49. Sreya, M., Alam, M. S., Daula, S., Lee, C., Restelli, V., Middlebrook, K., Noble, M. A., & Perrone, L. A. (2024). Improving drinking water quality through proficiency testing—the impact of testing method and accreditation status on Escherichia coli detection by Canadian environmental testing laboratories. *Frontiers in Molecular Biosciences*, *11*, 1338549. https://doi.org/10.3389/fmolb.2024.1338549
50. Sylvestre, É., Rhoads, W. J., Julian, T. R., & Hammes, F. (2024). Quantification of Legionella pneumophila in building potable water systems: a meta-analysis comparing qPCR and culture-based detection methods. *medRxiv*. https://doi.org/10.1101/2024.05.02.24306716
51. Takada, Y., & Kato, T. (2023, February 21). *Multi-Target Tobit models for completing water quality data*. arXiv.org. http://arxiv.org/abs/2302.10648
52. Tambi, A., Brighu, U., & Gupta, A. B. (2023). Methods for detection and enumeration of coliforms in drinking water: a review. *Water Science & Technology Water Supply*, *23*(10), 4047–4058. https://doi.org/10.2166/ws.2023.247
53. Thavarajah, W., Verosloff, M. S., Jung, J. K., Alam, K. K., Miller, J. D., Jewett, M. C., Young, S. L., & Lucks, J. B. (2020). A primer on emerging field-deployable synthetic biology tools for global water quality monitoring. *Npj Clean Water*, *3*(1). https://doi.org/10.1038/s41545-020-0064-8
54. Ting, J. O. N., & Yee, S. K. (2020). Review on water quality monitoring technologies. *Indonesian Journal of Electrical Engineering and Computer Science*, *18*(3), 1416. https://doi.org/10.11591/ijeecs.v18.i3.pp1416-1423
55. Tsaridou, C., & Karabelas, A. J. (2021). Drinking Water Standards and Their Implementation—A Critical Assessment. *Water*, *13*(20), 2918. https://doi.org/10.3390/w13202918
56. United Nations. (2023). *The Sustainable Development Goals Report 2023*. https://unstats.un.org/sdgs/report/2023/
57. Yazawa, T., Juanites, S., Baring, K. M., Andres, R., Peña, A. M. D., & Macabata-Rubite, P. E. (2026). Unequal water quality under decentralization: A case study of rural barangays in Antique, Philippines. *Journal of Water and Health*, *24*(3), 402–414. https://doi.org/10.2166/wh.2026.211
58. Zapf, A., Castell, S., Morawietz, L., & Karch, A. (2016). Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Medical Research Methodology*, *16*(1), 93. https://doi.org/10.1186/s12874-016-0200-9
59. Zimmer, C., Boyer, R., Cassivi, A., Tilley, E., Bain, R., Johnston, R., & Dorea, C. C. (2025). Reducing water quality data inequities: A low-cost to a membrane filtration technique for the quantification of

Escherichia coli in drinking water in low-resource contexts. *PLOS Water*, *4*(5), e0000367. https://doi.org/10.1371/journal.pwat.0000367