

Article

Applied Statistical Validation of GC–FID Acetaldehyde Determination in Extra-Neutral Alcohol Using Recovery, Relative Bias, and Factorial Analysis of Variance

Juanita Castro-Chica ¹, Diógenes de Jesus Ramirez-Ramirez ² and Cristian David Correa-Álvarez ^{2,*}

¹ Laboratorio de Química Analítica y Control de Calidad Industrial, Industria Licorera de Caldas, Zona Industrial Juanchito Kilómetro 10 Vía Magdalena, Manizales 170003, Colombia; jucastroch@unal.edu.co

² Departamento de Matemáticas y Estadística, Universidad Nacional de Colombia, Sede Manizales, Kilómetro 7 Vía al Aeropuerto, Manizales 170003, Colombia; ddramirezra@unal.edu.co

* Correspondence: crdcorreaal@unal.edu.co

Abstract

Acetaldehyde is routinely monitored in neutral alcohol, but pooled precision summaries can obscure changes in performance across the working range. This study evaluated 446 valid GC–FID verification results obtained by three anonymized analysts on two gas chromatographic systems (GC 8890 and GC 7890B) at 0.6, 2.5, and 9.0 mg L^{−1}. The analysis included measured concentration, recovery, relative bias, relative standard deviation (RSD), confidence intervals, and fixed-effect factorial analysis of variance (ANOVA). Mean recoveries for the GC 8890 and GC 7890B systems were 48.7% and 157.2% at 0.6 mg L^{−1}, 92.1% and 124.8% at 2.5 mg L^{−1}, and 92.0% and 103.1% at 9.0 mg L^{−1}, respectively. Recovery RSD decreased from 47.8% and 18.4% at the lowest level to 4.2% and 3.0% at the highest level. Type III ANOVA identified significant verification-level, analyst, GC-system, and interaction effects. A contextual comparison with concentration-matched AOAC Appendix F targets classified both systems as outside the recovery and RSD benchmarks at the lowest level, showed system-specific partial conformity at the middle level, and placed both systems within the screening benchmarks at the highest level. The contribution of the study is an integrated decision framework in which predefined recovery and RSD targets are interpreted together with factorial interactions to determine when results may be pooled and when level- or system-specific control is required.

Keywords: acetaldehyde; extra-neutral alcohol; gas chromatography; analytical method validation; ANOVA; intermediate precision; recovery; bias; quality control



Academic Editor: Takayuki Hibi

Received: 6 July 2026

Revised: 25 July 2026

Accepted: 27 July 2026

Published: 2 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Acetaldehyde is a routinely controlled volatile carbonyl compound in neutral alcohol and alcoholic beverages because its concentration reflects fermentation, distillation, purification, and storage conditions. Its determination becomes more demanding near the lower end of a working range: the analyte is volatile and reactive, and small absolute changes in detector response can produce large relative errors in an ethanol-rich matrix [1,2]. A reliable result must therefore demonstrate not only that different concentration levels are separated, but also that each level is measured with acceptable trueness and precision.

Routine validation reports often summarize accuracy and precision across an entire range. Such pooling is convenient, but it is difficult to justify when recovery or variability

changes with concentration, analyst, or instrument. Analysis of variance (ANOVA) provides a structured way to partition observed variation among predefined factors and their interactions. In a replicated factorial design, ANOVA is particularly useful for determining whether an analyst or instrumental effect is stable across concentration levels or whether its magnitude depends on the level being measured. ANOVA does not establish fitness for purpose by itself; it must be interpreted with analytical performance measures and predefined acceptance criteria.

The scientific contribution of this study lies in the integration of two established practices that are often reported separately. Concentration-dependent recovery and precision criteria are recognized in analytical-validation guidance, while factorial ANOVA is a standard method for evaluating crossed experimental factors. The proposed framework links them through an explicit decision sequence: first, recovery and RSD are evaluated within each verification-level and GC-system combination; second, factorial interactions are used to determine whether those combinations may be pooled; and third, the results are translated into level- and system-specific control actions. The framework is demonstrated with routine GC–FID verification data for acetaldehyde in extra-neutral alcohol.

The objective was to determine whether the chromatographic procedure performed consistently across three verification levels and two GC systems under routine laboratory conditions. The study (i) quantified recovery, relative bias, and RSD by verification level, analyst, and GC system; (ii) fitted factorial ANOVA models for measured concentration and recovery; (iii) compared the observed recovery and RSD values with contextual AOAC Appendix F benchmarks; and (iv) developed a decision framework that connects cell-specific performance and interaction effects to level-specific acceptance and corrective action.

2. Related Work

The literature supporting this study covers two connected areas: chromatographic determination of acetaldehyde and related volatile compounds in alcoholic matrices, and statistical approaches for evaluating concentration-, analyst-, and instrument-dependent analytical performance. The first area establishes the measurement context and the challenges associated with volatile analytes in ethanol-rich matrices. The second provides the basis for integrating recovery, relative standard deviation, and factorial analysis of variance when global performance summaries are insufficient for validation decisions.

2.1. GC–FID Determination of Acetaldehyde in Alcoholic Matrices

Gas chromatography with flame ionization detection (GC–FID) remains widely used for routine volatile-compound analysis because it offers direct quantification, broad availability, and manageable operating costs. Direct GC approaches have been reported for acetaldehyde, methanol, and related volatiles across alcoholic beverages, while fast GC–FID methods have been validated for spirits and wine [1,2]. Related studies have addressed method verification in wine, improvements to volatile-compound GC–FID procedures, and quality-control applications in alcohol-rich matrices [3,4]. Forensic GC–FID validation and advanced GC–MS quality-control studies illustrate the same need for within- and between-run evidence under the conditions of use [5,6].

Alternative approaches include headspace GC, GC–MS/MS, multidimensional GC, microextraction, and rapid colorimetric determination. These methods extend selectivity or sample-handling options, but they do not remove the need to verify performance at the concentrations used for routine decisions [7,8]. Recent studies of acetaldehyde and other volatile risk factors, alcohol-product authenticity, and volatile profiles in specialty beverages reinforce the practical importance of fit-for-purpose quantitative results [9,10].

Headspace applications in forensic alcohol analysis and volatile-impurity studies also show that calibration, storage, and sample handling can influence the reported value [11,12].

Current analytical developments include microextraction-based workflows and comprehensive two-dimensional gas chromatography coupled with chemometric analysis [13,14]. These techniques broaden the analytical toolbox, whereas the present study addresses a different problem: how to extract defensible validation decisions from replicated routine measurements generated by two GC–FID systems. Comparable work on ethanol analysis and chromatographic quality control confirms that method performance must be evaluated under the same operational conditions in which results are reported [15,16]. Chromatographic data combined with chemometric tools have also supported authenticity decisions, providing a related example of applied statistical interpretation [17].

2.2. Concentration-Dependent Validation and Factorial Analysis of Variance

ICH Q2(R2), ICH Q14, and Eurachem place the intended use, reportable result, and working range at the center of analytical validation [18,19]. These documents require accuracy and precision to be demonstrated over the relevant range but do not prescribe a universal numerical limit for every analyte and matrix. Eurachem similarly treats fitness for purpose as the governing principle and recommends that acceptance criteria be specified before interpreting validation data [20,21]. Lifecycle case studies show how the same risk-based logic can be used to structure development and subsequent change control [22,23]. For low-level acetaldehyde in extra-neutral alcohol, these principles are directly applicable because the decision risk depends on both the concentration level and the complete analytical procedure.

AOAC Appendix F provides generic, concentration-dependent performance targets that can serve as an external screening reference when method-specific regulatory limits are unavailable. At nominal mass fractions near 1–10 ppm, the expected recovery range is 80–110%, while the predicted repeatability RSD decreases from approximately 11% near 1 ppm to 7.3% near 10 ppm [24]. These values are not adopted here as formal compliance limits, because the present concentrations are expressed in mg L^{-1} and the pooled RSD includes variation among analysts. They are used to place the observed performance on a recognized analytical scale and to show how the decision changes across the evaluated range.

Statistically, replicated factorial designs are appropriate when multiple categorical factors are observed in combination and the study aims to evaluate both main effects and interactions. A significant interaction indicates that the effect of one factor changes across the levels of another; in validation terms, it cautions against reporting one pooled estimate for conditions that behave differently. ISO 5725 describes precision under repeatability, intermediate-precision, and reproducibility conditions, while ISO/IEC 17025 requires laboratories to control factors that affect result validity [25,26]. The present analysis uses factorial ANOVA as an applied model of the observed laboratory design and interprets it jointly with recovery, relative bias, and RSD rather than as a standalone acceptance test. Measurement-uncertainty and sampling guidance further caution against transferring performance estimates across conditions without examining their sources [27,28].

3. Materials and Methods

The methodological workflow comprised four linked components: the GC–FID procedure and standard preparation; the crossed experimental design and data structure; the response variables and contextual validation benchmarks; and the factorial analysis-of-variance implementation. These components are presented in sequence to show how the chromatographic measurements were converted into recovery, relative bias, relative standard deviation, and interaction-based evidence for the final validation decision.

3.1. Chromatographic Procedure and Standard Preparation

Acetaldehyde was determined by GC–FID using the laboratory’s routine quality-control procedure for extra-neutral alcohol. Two systems were evaluated: an Agilent GC 8890 and an Agilent GC 7890B. Both systems used a flame ionization detector and a mid-polar capillary column containing 6% cyanopropylphenyl and 94% polydimethylsiloxane. The column dimensions were 60 m × 0.25 mm i.d., with a 1.4 µm film thickness. Hydrogen was used as the carrier gas at a constant flow of 1.2 mL min^{−1}, and standards were injected in split mode.

Calibration was performed by the external-standard approach over 0.2–12.5 mg L^{−1}. OpenLab CDS generated a linear calibration function relating acetaldehyde peak area to the known standard concentration. Peaks were integrated by the valley-to-valley method within the acetaldehyde retention-time window, and the software calculated the final concentration from the external-standard curve.

The acetaldehyde stock solution had a nominal concentration of 5000 mg L^{−1} and was stored at 2–8 °C according to the manufacturer’s instructions. A 100 mg L^{−1} intermediate solution was prepared from the stock solution. Calibration standards and the three verification standards (PV1 = 0.6 mg L^{−1}, PV2 = 2.5 mg L^{−1}, and PV3 = 9.0 mg L^{−1}) were prepared by a combined volumetric–gravimetric procedure.

For each preparation, the target concentration and final flask volume were used to calculate the required aliquot. The mass of the transferred standard solution was recorded on an analytical balance and used to correct the starting concentration. Working and verification solutions were then prepared according to the dilution relationship referenced in Equation (1).

$$C_{\text{stock}} V_{\text{aliquot}} = C_{\text{target}} V_{\text{final}}, \quad (1)$$

where C_{stock} is the corrected stock or intermediate concentration, V_{aliquot} is the transferred aliquot, C_{target} is the target concentration, and V_{final} is the final volumetric-flask volume. Each sequence began with a blank, followed by calibration standards in ascending order. Verification-standard responses were compared with their theoretical values and with the baseline values obtained after preparation. Re-verification frequency was determined from current response and control-chart behavior. These controls are relevant for low-level acetaldehyde because standard stability, blank contribution, carryover, integration, and sequence conditions can have a disproportionate effect near PV1 [18,29].

3.2. Experimental Design and Data Structure

The experimental structure and missing-data assessment define the observations available for analysis. The study crossed three anonymized analysts, three verification levels, and two GC systems. The laboratory data table contained 225 wide-format records, with nominally 25 replicates for each analyst-by-level combination and paired concentration results from the GC 8890 and GC 7890B systems. Figure 1 and Table 1 summarize the design before the analytical results are presented.

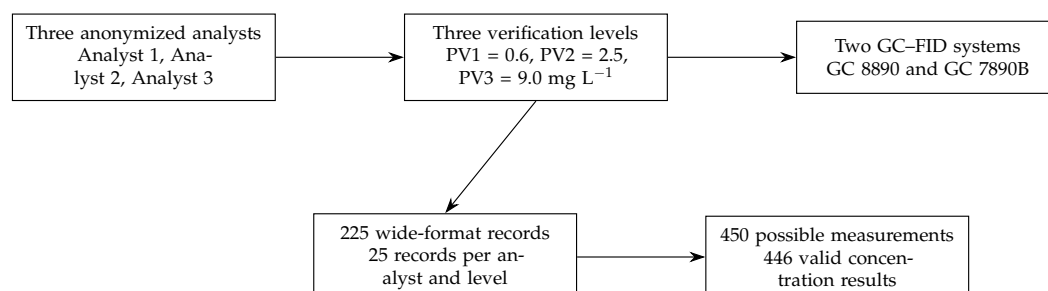


Figure 1. Structure of the verification study and the long-format data set used for analysis.

The experimental factors, replicate counts, and missing observations are detailed in Table 1.

Table 1. Experimental design used for the statistical analysis.

Element	Description
Matrix	Extra-neutral alcohol
Analyte	Acetaldehyde, reported as mg L ⁻¹
Verification levels	PV1 = 0.6 mg L ⁻¹ ; PV2 = 2.5 mg L ⁻¹ ; PV3 = 9.0 mg L ⁻¹
Analysts	Analyst 1, Analyst 2, and Analyst 3
GC systems	GC 8890 and GC 7890B
Replicates	Nominally 25 per PV-by-analyst cell on each GC system
Potential measurements	450
Valid concentration results	446
Missing results	Four observations for PV3, Analyst 1, GC 7890B, replicates 22–25

The paired system results were converted to long format, and each observation was linked to its nominal PV concentration. Missing observations were retained explicitly and excluded only from calculations requiring a measured concentration; no imputation was performed. The four missing values reduced the PV3 sample size for the GC 7890B from 75 to 71 but did not alter the balance of the other cells.

3.3. Response Variables and Validation Benchmarks

Measured concentration was the direct reportable response. Percentage recovery placed the three verification levels on a common relative scale and was calculated according to Equation (2).

$$\text{Recovery}(\%) = 100 \times \frac{C_{\text{measured}}}{C_{\text{nominal}}}, \quad (2)$$

where C_{measured} is the reported acetaldehyde concentration and C_{nominal} is the assigned PV concentration. Relative bias retained the direction of disagreement with the nominal value and was calculated according to Equation (3).

$$\text{Relative bias}(\%) = \text{Recovery}(\%) - 100. \quad (3)$$

Precision comparisons were based primarily on the relative standard deviation (RSD), which is equivalent to the coefficient of variation when expressed as a percentage. For each PV-by-GC-system or PV-by-analyst-by-GC-system group, recovery RSD was calculated using Equation (4).

$$\text{RSD}(\%) = 100 \times \frac{s_{\text{recovery}}}{\bar{R}_{\text{recovery}}}, \quad (4)$$

where s_{recovery} and $\bar{R}_{\text{recovery}}$ are the sample standard deviation and mean recovery within the relevant group. Ninety-five percent confidence intervals for mean recovery were calculated using the Student t multiplier.

The primary interpretation followed ICH Q2(R2), ICH Q14, and Eurachem: performance was assessed against the intended range and the behavior of the complete analytical procedure [18,20]. AOAC Appendix F was used as a contextual benchmark because it provides explicit concentration-dependent expectations. For nominal levels near 1–10 ppm, the expected recovery range is 80–110%; the approximate predicted repeatability RSD is 11% near 1 ppm and 7.3% near 10 ppm [24]. The comparison is deliberately described as contextual rather than regulatory because mg L⁻¹ is not identical to mg kg⁻¹ and the study RSD includes analyst-related variation rather than repeatability alone.

3.4. Factorial Analysis of Variance and Computational Implementation

Analysis of variance partitions the total variation in a continuous response into components associated with categorical factors and residual variation. A factorial ANOVA is applicable when observations are replicated across combinations of two or more factors and the purpose is to estimate both main effects and interactions. The present design met those conditions because verification level, analyst, and GC system were crossed factors with repeated measurements in each cell. Two fixed-effect models were fitted: one for measured concentration and one for recovery. The model is defined in Equation (5), where G_k denotes the GC-system factor.

$$Y_{ijkl} = \mu + PV_i + Analyst_j + G_k + (PV \times Analyst)_{ij} + (PV \times G)_{ik} + (Analyst \times G)_{jk} + (PV \times Analyst \times G)_{ijk} + \epsilon_{ijkl}. \quad (5)$$

Type III sums of squares were used because four missing PV3 observations made the design slightly unbalanced. Sum-to-zero contrasts were specified so that each Type III test evaluated a factor after accounting for the remaining terms. Main effects describe average differences, whereas interaction terms test whether analyst or GC-system differences change across verification levels. The ANOVA results were not treated as direct acceptance criteria; they were interpreted with cell-specific recovery, relative bias, RSD, and external benchmarks. This use of factor effects to support method understanding is consistent with practical analytical-quality-by-design tools [30].

All calculations were performed in R version 4.5.2 [31]. Data restructuring and graphics used the tidyverse collection [32]; variable-name cleaning used janitor [33]; and Type III tests were obtained with the car package, following the implementation described by Fox and Weisberg [34]. The workflow retained missing records explicitly, calculated the validation summaries, fitted the two ANOVA models, and generated the figures. Appendix A summarizes the computational sequence.

4. Results

The results are organized around the validation decision. Recovery and RSD are first compared by verification level and GC system, followed by a contextual benchmark assessment, analyst-specific behavior, and the factorial ANOVA results. The analysis used the 446 valid observations described in Section 3.2.

4.1. Recovery and Precision by Verification Level and GC System

Table 2 summarizes measured concentration, recovery, RSD, and relative bias by PV and GC system. The system difference was largest at PV1: mean recovery was 48.7% on the GC 8890 and 157.2% on the GC 7890B, corresponding to relative biases of −51.3% and 57.2%. At PV2, the GC 8890 result was closer to nominal (92.1%), whereas the GC 7890B remained high (124.8%). Both systems were closer to the nominal value at PV3, with recoveries of 92.0% and 103.1%. Recovery RSD also decreased with concentration, from 47.8% and 18.4% at PV1 to 4.2% and 3.0% at PV3.

Figure 2 shows the same pattern with 95% confidence intervals. The PV1 intervals are far from 100% in opposite directions, while the PV2 interval for the GC 7890B remains above 100%. PV3 provides the closest agreement on both systems. The narrow confidence intervals indicate that these mean shifts were reproducible within the observed data rather than being driven by a few isolated measurements.

Table 2. Descriptive validation summaries by verification level and GC system.

PV	GC System	<i>n</i>	Mean Measured Concentration (mg L ⁻¹)	Mean Recovery (%)	Recovery RSD (%)	Relative Bias (%)
PV1	GC 8890	75	0.292	48.7	47.8	−51.3
PV1	GC 7890B	75	0.943	157.2	18.4	57.2
PV2	GC 8890	75	2.303	92.1	16.4	−7.9
PV2	GC 7890B	75	3.120	124.8	7.1	24.8
PV3	GC 8890	75	8.276	92.0	4.2	−8.0
PV3	GC 7890B	71	9.279	103.1	3.0	3.1

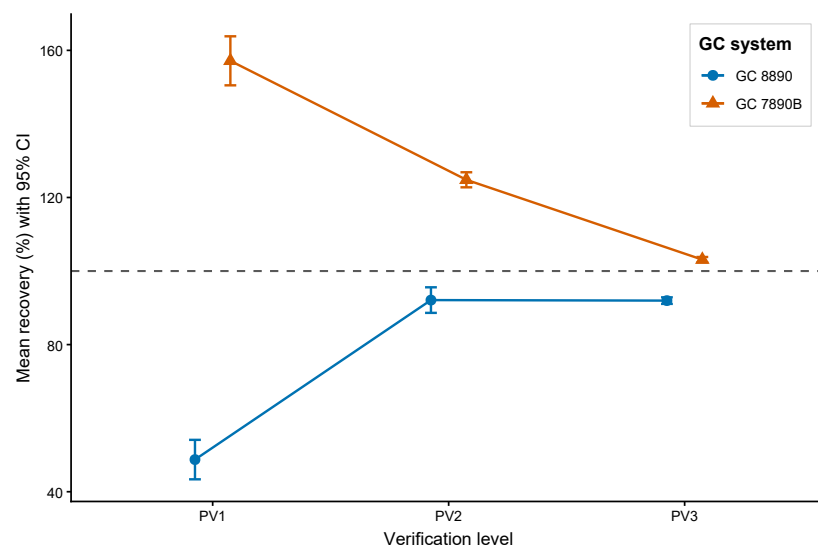
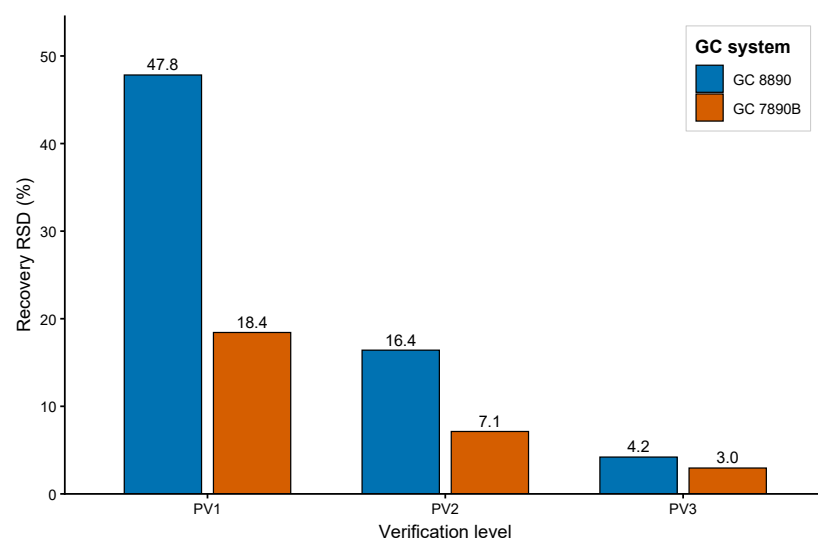
**Figure 2.** Mean recovery and 95% confidence intervals by verification level and GC system. The dashed horizontal line indicates 100% recovery.

Figure 3 compares recovery RSD across the three verification levels and the two GC systems. Relative dispersion was highest at PV1, reaching 47.8% for the GC 8890 and 18.4% for the GC 7890B; it decreased at PV2 to 16.4% and 7.1%, and at PV3 to 4.2% and 3.0%, respectively. This concentration-dependent decline shows that a single pooled precision estimate would not adequately represent the working range and supports level-specific precision assessment.

**Figure 3.** Recovery relative standard deviation by verification level and GC system.

4.2. Contextual Comparison with Concentration-Dependent Benchmarks

Table 3 compares the observed recovery and RSD values with the nearest concentration-dependent AOAC Appendix F screening targets. At PV1, both systems were outside both benchmarks. At PV2, the GC 8890 met the recovery range but exceeded the approximate RSD target, while the GC 7890B had an acceptable RSD but excessive recovery. At PV3, both systems were within the recovery and RSD screening ranges. This comparison identifies three different decision states: broad nonconformity at PV1, system-specific partial conformity at PV2, and provisional conformity at PV3.

Table 3. Contextual screening of observed recovery and RSD against concentration-matched AOAC Appendix F targets.

PV	GC System	Recovery (%)	RSD (%)	AOAC Recovery Target (%)	Approx. AOAC RSD Target (%)
PV1	GC 8890	48.7	47.8	80–110	≤11
PV1	GC 7890B	157.2	18.4	80–110	≤11
PV2	GC 8890	92.1	16.4	80–110	7.3–11
PV2	GC 7890B	124.8	7.1	80–110	7.3–11
PV3	GC 8890	92.0	4.2	80–110	≤7.3
PV3	GC 7890B	103.1	3.0	80–110	≤7.3

Note: AOAC Appendix F targets are based on nominal mass fractions. The present concentrations are reported in mg L⁻¹, and the study RSD pools analysts within each PV-by-system group. The comparison is therefore a screening reference rather than a formal AOAC conformity assessment.

The PV2 contrast further shows that acceptable relative precision does not compensate for biased recovery, and that acceptable recovery does not compensate for excessive dispersion.

4.3. Analyst-Specific Recovery and Relative Bias

Table 4 reports analyst-specific recovery and RSD for each PV and GC system. The largest analyst contrast occurred at PV1 on the GC 8890: mean recovery ranged from 70.9% for Analyst 1 to 19.4% for Analyst 3. On the GC 7890B, all three PV1 means were high, ranging from 121.8% to 184.3%. The PV2 system difference persisted for every analyst. At PV3, recovery and RSD were narrower, although the GC 8890 means remained slightly low for Analysts 2 and 3.

Table 4. Analyst-specific recovery and RSD by verification level and GC system.

PV	Analyst	GC System	n	Mean Recovery (%)	Recovery RSD (%)
PV1	Analyst 1	GC 8890	25	70.91	13.00
PV1	Analyst 1	GC 7890B	25	165.36	7.62
PV1	Analyst 2	GC 8890	25	55.87	15.83
PV1	Analyst 2	GC 7890B	25	184.35	7.44
PV1	Analyst 3	GC 8890	25	19.44	37.57
PV1	Analyst 3	GC 7890B	25	121.76	7.98
PV2	Analyst 1	GC 8890	25	95.34	21.44
PV2	Analyst 1	GC 7890B	25	128.76	4.17
PV2	Analyst 2	GC 8890	25	91.11	4.14
PV2	Analyst 2	GC 7890B	25	128.40	8.06
PV2	Analyst 3	GC 8890	25	89.86	17.78
PV2	Analyst 3	GC 7890B	25	117.27	3.72
PV3	Analyst 1	GC 8890	25	96.18	3.63
PV3	Analyst 1	GC 7890B	21	105.91	2.01
PV3	Analyst 2	GC 8890	25	88.38	1.36
PV3	Analyst 2	GC 7890B	25	100.50	2.71
PV3	Analyst 3	GC 8890	25	91.32	0.53
PV3	Analyst 3	GC 7890B	25	103.34	1.32

Figure 4 preserves the direction of the analyst-specific bias. At PV1, the most negative bias was -80.56% for Analyst 3 on the GC 8890, and the largest positive bias was 84.35% for Analyst 2 on the GC 7890B. Biases at PV3 were closer to zero for all analyst–system combinations. The figure shows why analyst effects cannot be interpreted independently of PV and GC system.

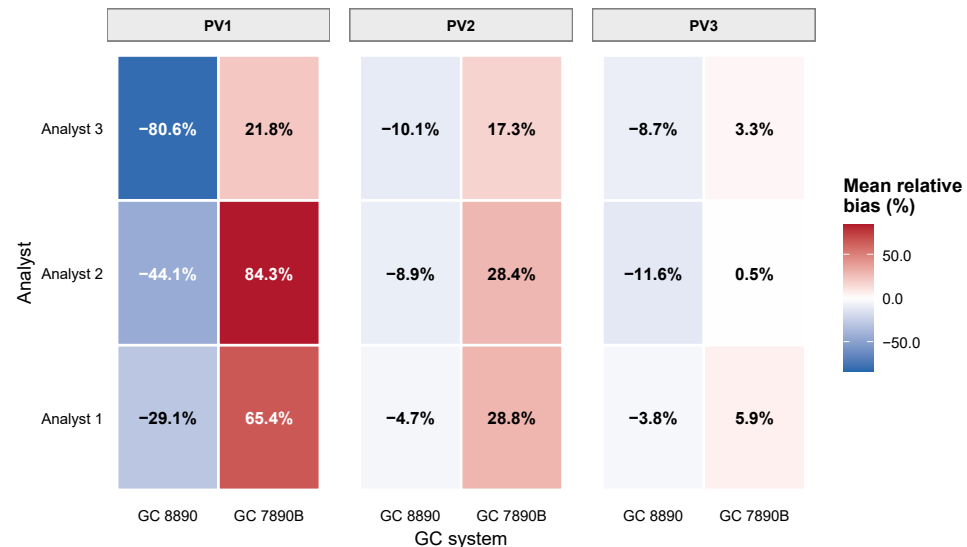


Figure 4. Mean relative bias by verification level, analyst, and GC system. Negative values indicate underestimation; positive values indicate overestimation relative to the nominal concentration.

4.4. Factorial Analysis-of-Variance Results

Type III ANOVA results for measured concentration are presented in Table 5. Verification level dominated the concentration model because the nominal levels were widely separated. Analyst, GC system, and all interaction terms were also significant. The three-way interaction was smaller than the other effects but remained significant at $p = 0.048$.

Table 5. Type III analysis of variance for measured concentration.

Term	Sum of Squares	df	F	p Value
PV	5292.994	2	61,445.03	<0.001
Analyst	6.199	2	71.97	<0.001
GC system	76.333	1	1772.27	<0.001
PV × Analyst	6.489	4	37.66	<0.001
PV × GC system	2.470	2	28.68	<0.001
Analyst × GC system	0.614	2	7.12	<0.001
PV × Analyst × GC system	0.416	4	2.41	0.048
Residual	18.434	428	–	–

The recovery model in Table 6 is more directly connected to validation decisions because all PV levels are expressed on the same relative scale. Every main effect and interaction was significant at $p < 0.001$. The large PV-by-GC-system interaction confirms that the system difference changed across the working range, while the PV-by-analyst interaction shows that analyst differences were also concentration-dependent. These interactions rule out one pooled recovery or precision statement as an adequate description of the observed design.

The recovery model explained 94.1% of the observed recovery variation, whereas the concentration model explained 99.7% of measured-concentration variation. The latter value mainly reflects the separation among nominal levels and should not be interpreted

as evidence of analytical fitness. In the recovery model, the largest sums of squares were associated with GC system and the PV-by-GC-system interaction, followed by PV-by-analyst variation. This ordering is consistent with the descriptive and benchmark results.

Table 6. Type III analysis of variance for recovery.

Term	Sum of Squares	df	F	p Value
PV	8695.3	2	50.78	<0.001
Analyst	35,326.9	2	206.29	<0.001
GC system	287,333.6	1	3355.70	<0.001
PV × Analyst	45,936.4	4	134.12	<0.001
PV × GC system	193,492.2	2	1129.87	<0.001
Analyst × GC system	4062.3	2	23.72	<0.001
PV × Analyst × GC system	4495.2	4	13.12	<0.001
Residual	36,647.7	428	–	–

5. Discussion

The discussion interprets the descriptive and factorial results as components of a single validation decision rather than as independent statistical findings. It first defines the contribution of combining concentration-specific recovery and relative standard deviation with factorial interactions, and then explains how this evidence supports level- and system-specific acceptance decisions. The subsequent subsections examine plausible analytical causes, implications for routine control, consistency with current validation practice, and the limitations of the available evidence.

5.1. Primary Contribution and Decision Implications

The contribution of this work is an integrated decision process that combines established analytical criteria with factorial modeling. Existing guidance already recognizes concentration-dependent performance and fitness for purpose, but cell-specific analytical performance and factorial interactions are often reported separately. In the proposed framework, recovery and RSD determine whether each PV-by-system condition meets a predefined target, whereas interaction terms determine whether results across verification levels, analysts, or GC systems can be pooled. This integration is important because a global concentration model can appear satisfactory even when low-level recovery is unacceptable.

The practical risk is clearest at PV1. Underestimation on the GC 8890 could cause a sample near an internal action limit to be reported too low, while overestimation on the GC 7890B could produce unnecessary investigation or product hold. PV2 presents a different problem: one system showed acceptable mean recovery but excessive relative variability, whereas the other showed low RSD but excessive positive bias. PV3 was the only level at which both systems satisfied the contextual recovery and RSD benchmarks. The data therefore support different actions by level rather than one statement of overall method acceptance.

5.2. From Factorial Interactions to a Level-Specific Acceptance Framework

The logical connection between ANOVA and acceptance is summarized in Figure 5. ANOVA does not generate acceptance limits. The limits must be defined from the intended use, applicable guidance, product requirements, and measurement risk. ANOVA addresses a separate question: whether the same limit and summary can be applied across the observed factors. A material or statistically significant PV-by-system or PV-by-analyst interaction indicates that performance is conditional on concentration and that pooled estimates can conceal relevant differences.

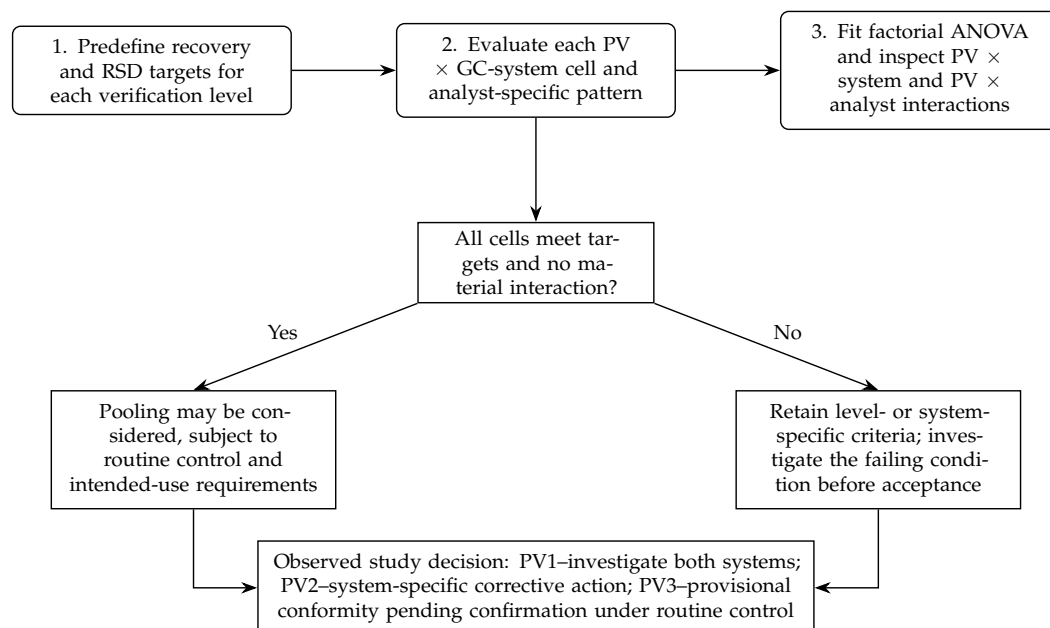


Figure 5. Decision framework linking level-specific analytical targets and factorial interactions to pooling, stratification, and corrective action.

Applied to the present data, the framework produces a clear sequence. PV1 fails both recovery and RSD screening on both systems, so neither pooling nor routine acceptance is supported. At PV2, the two systems fail different criteria, and the large PV-by-system interaction indicates that system-specific investigation is required. At PV3, both systems meet the contextual benchmarks; however, provisional acceptance should still be confirmed with predefined laboratory criteria and routine-control evidence. The framework therefore translates the statistical model into an operational rule without treating a p value as a specification.

5.3. Likely Analytical Causes and Routine Control

The data do not identify the physical cause of the system-dependent bias. Plausible sources include calibration preparation, standard stability, injection-system condition, response-factor estimation, peak integration, blank contribution, carryover, and sequence-level drift. Acetaldehyde volatility makes low-level standards especially sensitive to vial sealing, handling time, storage, and headspace conditions [12,29]. The opposite PV1 biases on the two GC systems are unlikely to be resolved by adding more pooled replicates; they require targeted review of chromatograms, calibration residuals, standard-preparation records, blank responses, maintenance history, and integration settings.

Routine control should track PV1, PV2, and PV3 separately. Each control chart should use limits established after corrective action and confirmation of stable performance. Analyst checks should also be stratified by level because the heatmap and PV-by-analyst interaction show that analyst effects were greatest at PV1. If the two GC systems are intended to be interchangeable, equivalence should be demonstrated prospectively with common standards, balanced sequences, and explicit system identifiers.

5.4. Relation to Current Validation Practice

The proposed framework extends current practice by specifying how established guidance and factorial evidence are combined. ICH Q2(R2) and Eurachem require predefined criteria across the reportable range but leave the numerical limits to the intended use [18,20]. AOAC Appendix F supplies concentration-dependent screening values that make the range effect visible, while factorial ANOVA determines whether analyst and

GC-system effects are sufficiently stable to support pooled reporting. The framework does not replace method-specific criteria; it provides a reproducible rule for deciding the scale at which those criteria should be applied.

This interpretation is consistent with the analytical-procedure lifecycle described in ICH Q14 and recent implementation work [19,35]. Validation is not complete when a single favorable statistic is obtained. It requires knowledge of where the method is reliable, where additional control is needed, and which operational factors must remain monitored. Measurement-uncertainty guidance leads to a similar conclusion: uncertainty should reflect the conditions relevant to the reported result rather than borrow low-level behavior from a more stable high-level condition [36,37]. Improved approaches to expanded uncertainty likewise depend on a defensible characterization of the underlying variability [38].

5.5. Limitations

The study has five principal limitations. First, it evaluates one analyte in one matrix and one laboratory. Second, the design treats the three analysts and two GC systems as fixed observed factors; broader generalization would require a prospective mixed-effects design with days, calibration batches, and independently prepared standards. Third, the pooled PV-by-system RSD includes between-analyst variation and is therefore not identical to AOAC repeatability RSD. Fourth, AOAC concentrations are expressed as mass fractions, whereas the present study reports mg L^{-1} ; the benchmark comparison is contextual. Fifth, the available results did not include the numerical blank, certified-reference-material, and spiked-sample data needed for a complete validation and uncertainty budget.

The known chromatographic conditions support interpretation of the verification results, but independent reproduction would also require full reporting of injection volume, split ratio, inlet and detector temperatures, oven program, and retention-time window. Future work should use independently prepared low-level standards, balanced runs across GC systems and analysts, and predefined acceptance criteria linked to the laboratory's decision limits. A mixed model would then permit estimation of variance components for analyst, system, day, calibration, and preparation effects in addition to the fixed-effect comparisons reported here [21,25].

6. Conclusions

The GC-FID procedure did not perform uniformly across the evaluated range. At PV1, both GC systems were outside the contextual recovery and RSD benchmarks and showed large biases in opposite directions. At PV2, the GC 8890 showed acceptable mean recovery but excessive RSD, while the GC 7890B showed acceptable RSD but excessive positive bias. At PV3, both systems were within the screening benchmarks and had RSD values below 5%.

Factorial ANOVA showed that verification level, analyst, GC system, and their interactions affected recovery. The PV-by-system and PV-by-analyst interactions are the statistical basis for rejecting one pooled acceptance statement: performance depended on the concentration level and operating condition. The integrated framework developed here combines predefined recovery and RSD targets with factorial interactions to decide whether results may be pooled or require level- and system-specific control. For the present procedure, corrective investigation is required at PV1 and PV2 before routine low-level acceptance can be supported, whereas PV3 provides the strongest basis for provisional routine use.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/appliedmath6080124/s1>: Table S1. Anonymized acetaldehyde verification data used for the validation summaries, recovery and bias calculations, and factorial ANOVA models.

Author Contributions: Conceptualization, J.C.-C. and D.d.J.R.-R.; methodology, J.C.-C.; formal analysis, C.D.C.-Á.; data curation, C.D.C.-Á.; writing—original draft preparation, J.C.-C. and D.d.J.R.-R.; writing—review and editing, C.D.C.-Á.; supervision, J.C.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The anonymized experimental data supporting the results of this study are provided as Supplementary Table S1. The computational workflow used to organize the data, calculate validation summaries, fit the factorial ANOVA models, and generate the manuscript figures is described in Appendix A.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (version 5.5, OpenAI) for limited language editing and stylistic refinement. No generative AI tools were used for study design, data collection, data analysis, or interpretation of results. The authors reviewed and edited all AI-assisted outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Computational Workflow and Pseudocode

This appendix summarizes the computational workflow used to organize the acetaldehyde verification data, calculate validation summaries, fit the factorial ANOVA models, and generate the manuscript figures. Data manipulation and graphics used `tidyverse` [32]; variable-name cleaning used `janitor` [33]; Type III ANOVA used `car` [34]; and linear models used the base R `stats` package [31]. The appendix documents the analytical logic of the models described in Section 3.4 rather than serving as a software supplement. The anonymized experimental data are provided as Supplementary Table S1.

1. Define $PV1 = 0.6 \text{ mg L}^{-1}$, $PV2 = 2.5 \text{ mg L}^{-1}$, and $PV3 = 9.0 \text{ mg L}^{-1}$.
2. Import the wide-format table containing analyst, PV, replicate, area, and paired GC-system concentration results.
3. Anonymize analysts and map the paired results to the GC 8890 and GC 7890B systems.
4. Convert the table to long format and retain missing concentration entries explicitly.
5. Join each observation to its nominal PV concentration.
6. Calculate recovery and relative bias according to Equations (2) and (3).
7. Calculate recovery RSD according to Equation (4) for each required grouping.
8. Exclude missing values only from calculations that require a measured concentration; do not impute.
9. Summarize n , mean concentration, mean recovery, relative bias, recovery RSD, and 95% confidence intervals by PV, GC system, and analyst.
10. Set sum-to-zero contrasts and fit the two fixed-effect models:

$$C_{\text{measured}} \sim PV \times Analyst \times GCSystem,$$

$$Recovery \sim PV \times Analyst \times GCSystem.$$

11. Estimate Type III ANOVA tables and report sum of squares, degrees of freedom, F statistic, and p value.
12. Compare PV-by-system recovery and RSD with the contextual AOAC Appendix F benchmarks.
13. Generate the recovery confidence-interval figure, recovery-RSD figure, relative-bias heatmap, and decision-framework diagram.
14. Interpret the ANOVA interactions jointly with the cell-specific performance results; do not use the p values as acceptance limits.

References

- Charapitsa, S.; Sytova, S.; Kavalenka, A.; Sobolenko, L.; Shauchenka, Y.; Kostyuk, N.; Egorov, V.; Leschev, S.; Vetokhin, S.; Zayats, N.; et al. The method for direct gas chromatographic determination of acetaldehyde, methanol, and other volatiles using ethanol as a reference substance: Application for a wide range of alcoholic beverages. *Food Anal. Methods* **2021**, *14*, 2088–2100. [CrossRef]
- Paolini, M.; Tonidandel, L.; Larcher, R. Development, validation and application of a fast GC–FID method for the analysis of volatile compounds in spirit drinks and wine. *Food Control* **2022**, *136*, 108873. [CrossRef]
- Šorgić, S.; Sredović Ignjatović, I.; Antić, M.; Šaćirović, S.; Pezo, L.; Čejčić, V.; Đurović, S. Monitoring of the wines' quality by gas chromatography: HSS–GC/FID method development, validation, verification, for analysis of volatile compounds. *Fermentation* **2022**, *8*, 38. [CrossRef]
- Sabalenska, L.; Charapitsa, S.; Sytova, S.; Kavalenka, A.; Yushkevitch, D.; Zayats, M.; Leschev, S. Proposals for the improving of the existing GC–FID methods for determination of methanol and volatile compounds in alcoholic beverages. *J. Food Meas. Charact.* **2023**, *17*, 3207–3217. [CrossRef]
- Cerioni, A.; Mietti, G.; Cippitelli, M.; Ricchezze, G.; Buratti, E.; Frolidi, R.; Cingolani, M.; Scendoni, R. Validation of a headspace gas chromatography with flame ionization detection method to quantify blood alcohol concentration (BAC) for forensic practice. *Chemosensors* **2024**, *12*, 133. [CrossRef]
- Korban, A.; Charapitsa, S.; Čabala, R.; Sobolenko, L.; Egorov, V.; Sytova, S. Advanced GC–MS method for quality and safety control of alcoholic products. *Food Chem.* **2021**, *338*, 128107. [CrossRef] [PubMed]
- Oh, H.-A.; Park, M.K. Determination of a robust headspace GC–MS/MS method for analysis of ethyl alcohol and acetaldehyde: Clinical application to pharmacokinetics study. *Transl. Clin. Pharmacol.* **2024**, *32*, 159–172. [CrossRef] [PubMed]
- Hatta-Sakoda, B.; Giusti, M.M.; Rodriguez-Saona, L.E.; Condezo-Hoyos, L. Digital colorimetric approach for rapid determination of acetaldehyde in pisco head during distillation. *Chemosensors* **2026**, *14*, 84. [CrossRef]
- Wu, Z.; Lv, S.; Xiao, P.; Salentijn, G.; Cheng, H.; Li, H.; Sun, J.; Ye, X.; Sun, B. Comprehensive analysis of risk factors (methanol, acetaldehyde and higher alcohols) in alcoholic beverages and their reduction strategies: GC–MS analysis and modified activated carbon adsorption and characterization. *Food Chem.* **2024**, *460*, 140461. [CrossRef] [PubMed]
- Jeong, D.; Kim, Y.-S. Analysis of methanol, fusel alcohols, and other volatile compounds of local specialty alcoholic beverages (wine, beer, and soju) in Korea. *Appl. Biol. Chem.* **2025**, *68*, 11. [CrossRef]
- Ndikumana, E.; Niyonizera, E.; Kabera, J.N.; Pandey, A. Analysis of methanol and ethanol content in illegal alcoholic beverages using headspace gas chromatography: Case studies at Rwanda Forensic Institute. *Braz. J. Anal. Chem.* **2024**, *11*, 18–33. [CrossRef]
- To, N.D.K.; Theruvathu, J.A. Determination and quantification of acetaldehyde, acetone, and methanol in hand sanitizers using headspace GC/MS: Effect of storage time and temperature. *Int. J. Environ. Res. Public Health* **2024**, *21*, 74. [CrossRef] [PubMed]
- Qiu, Y.; Deng, Q.; Zhang, Y.; Sun, B.; Li, W.; Dong, W.; Sun, X. Applications of microextraction technology for the analysis of alcoholic beverages quality: Current perspectives and future directions. *Foods* **2025**, *14*, 1152. [CrossRef] [PubMed]
- He, X.; Jelen, H.H. Comprehensive two dimensional gas chromatography–time of flight mass spectrometry (GC×GC–TOFMS) for the investigation of botanical origin of raw spirits. *Food Chem.* **2025**, *465*, 142004. [CrossRef] [PubMed]
- Farine, G.; Scilinguo, M.; Calvagna, C.; Romano, S.; Polito, M.; Sgro, S. Determination of ethanol content in alcoholic products by LF-NMR. *Analytica* **2024**, *5*, 458–470. [CrossRef]
- Costa, N.G.; Freitas, D.S.; Barros, A.; Silva, C.; Antunes, J.C.; Rocha, A.M. Development and optimization of a SPME–GC–FID method for ethanol detection. *Processes* **2024**, *12*, 247. [CrossRef]
- do Nascimento, H.O.; da Silva, M.Z.F.; Alexandre, J.B.; Oliveira, T.I.S.; de Lima, A.C.A.; do Nascimento, R.F. Chemometric tools associated to chromatographic data to assess the quality and authenticity whiskeys. *Microchem. J.* **2023**, *189*, 108542. [CrossRef]
- International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. *ICH Harmonised Guideline Q2(R2): Validation of Analytical Procedures*; ICH: Geneva, Switzerland, 2023. Available online: [https://database.ich.org/sites/default/files/ICH_Q2\(R2\)_Guideline_2023_1130.pdf](https://database.ich.org/sites/default/files/ICH_Q2(R2)_Guideline_2023_1130.pdf) (accessed on 10 December 2025).
- International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. *ICH Harmonised Guideline Q14: Analytical Procedure Development*; ICH: Geneva, Switzerland, 2023. Available online: https://database.ich.org/sites/default/files/ICH_Q14_Guideline_2023_1116_1.pdf (accessed on 10 December 2025).
- Cantwell, H. (Ed.) *Eurachem Guide: The Fitness for Purpose of Analytical Methods—A Laboratory Guide to Method Validation and Related Topics*, 3rd ed.; Eurachem: Bucharest, Romania, 2025. Available online: <https://www.eurachem.org/index.php/publications/guides/mv> (accessed on 18 December 2025).
- Barwick, V. (Ed.) *Planning and Reporting Method Validation Studies—Supplement to Eurachem Guide on the Fitness for Purpose of Analytical Methods*, 2nd ed.; Eurachem: Bucharest, Romania, 2025. Available online: <https://eurachem.org/index.php/publications/guides/planning-validation-studies> (accessed on 18 December 2025).
- Zhang, J.; Raghavachari, R.; Kirkpatrick, D.C.; Keire, D.A.; Xu, X.; Faustino, P.J. Analytical procedure development and proposed established conditions: A case study of a mass spectrometry based NDSRI analytical procedure. *J. Pharm. Sci.* **2024**, *113*, 3028–3033. [CrossRef] [PubMed]

23. Kirkpatrick, D.; Gupta, N.; Gopaldaswamy, R.; Kolhatkar, R.; Hudson-Davis, M.; Keire, D. The role of enhanced analytical procedure development in facilitating post-approval changes via established conditions. *AAPS J.* **2025**, *27*, 61. [CrossRef] [PubMed]
24. Latimer, G.W. Appendix F: Guidelines for Standard Method Performance Requirements. In *Official Methods of Analysis of AOAC INTERNATIONAL*, 22nd ed.; Latimer, G.W., Jr., Ed.; Oxford University Press: New York, NY, USA, 2023; pp. AF-1–AF-18. [CrossRef]
25. ISO 5725-3:2025; Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method. International Organization for Standardization: Geneva, Switzerland, 2025. Available online: <https://www.iso.org/standard/90054.html> (accessed on 18 December 2025).
26. ISO/IEC 17025:2017; General Requirements for the Competence of Testing and Calibration Laboratories. International Organization for Standardization: Geneva, Switzerland, 2017. Available online: <https://www.iso.org/standard/66912.html> (accessed on 5 December 2025).
27. Joint Committee for Guides in Metrology. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement*; JCGM 100:2008; BIPM: Sevres, France, 2008. Available online: <https://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e.html/C045315e.html?csnumber=50461> (accessed on 10 December 2025).
28. Ramsey, M.H.; Rostron, P.D.; Raposo, F.C. *Validation of Measurement Procedures That Include Sampling (VaMPIS)*; Eurachem: Bucharest, Romania, 2024. Available online: https://sussex.figshare.com/articles/report/Validation_of_Measurement_Procedures_that_Include_Sampling_VaMPIS_/28204742 (accessed on 18 December 2025).
29. Cantwell, H. (Ed.) *Blanks in Method Validation—Supplement to Eurachem Guide The Fitness for Purpose of Analytical Methods*, 2nd ed.; Eurachem: Bucharest, Romania, 2025. Available online: <https://www.eurachem.org/index.php/publications/guides/blanks-in-method-validation> (accessed on 18 December 2025).
30. van Tricht, E.; Sanger-van de Griend, C.E. A practical approach to implementing ICH Q14: Tools for analytical quality by design in capillary electrophoresis method development. *Electrophoresis* **2025**, *46*, 933–947. [CrossRef] [PubMed]
31. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2025. Available online: <https://www.R-project.org/> (accessed on 12 May 2026).
32. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; Franois, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]
33. Firke, S. *janitor: Simple Tools for Examining and Cleaning Dirty Data*; R Package Version 2.2.1; R Foundation for Statistical Computing: Vienna, Austria, 2024. [CrossRef]
34. Fox, J.; Weisberg, S. *An R Companion to Applied Regression*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2018. Available online: <https://books.google.com/books?id=SfNrDwAAQBAJ> (accessed on 12 May 2026).
35. Eberle, M.; Wasylenko, J.T.; Kostelac, D.; Kiehna, S.; Schellinger, A.; ZhangJason, Z.; Ehrick, D. A modern framework for analytical procedure development and lifecycle management based on ICH Q14 principles. *Anal. Chem.* **2024**, *97*, 12–26. [CrossRef] [PubMed]
36. Raposo, F.C.; Ramsey, M.H. The role of measurement uncertainty in the validation of a measurement procedure. *Accred. Qual. Assur.* **2024**, *29*, 267. [CrossRef]
37. Ramsey, M.H.; Rostron, P.D. Measurement uncertainty from sampling and its role in validation of measurement procedures. *Accred. Qual. Assur.* **2024**, *29*, 153–162. [CrossRef]
38. Analytical Methods Committee. Improved estimates of expanded measurement uncertainty. *Anal. Methods* **2024**, *16*, 5482–5485. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.