

A proof-of-concept practical approach for achieving equivalent results from non-harmonized measurement methods while awaiting harmonization: The CA 125 example

Hikmet Can Çubukçu^{a,b,*}, Oğuzhan Zengi^c, Hamit Hakan Alp^d, Murat Cihan^e, Kamil Taha Uçar^c, Marc Thelen^f, Mauro Panteghini^g

^a Rare Diseases Department, General Directorate of Health Services, Turkish Ministry of Health, Ankara, Turkey

^b Department of Medical Biochemistry, Sincan Training and Research Hospital, Ankara, Turkey

^c Department of Medical Biochemistry, Health Science University, Istanbul Basaksehir Cam and Sakura City Hospital, Istanbul, Turkey

^d Department of Biochemistry, Faculty of Medicine, Van Yüziüncü Yıl University, Van, Turkey

^e Department of Medical Biochemistry, Memorial Hospital, Istanbul, Turkey

^f SKML, Nijmegen, The Netherlands, and Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, the Netherlands

^g Department of Laboratory Medicine, Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun, Torun, Poland

ARTICLE INFO

Keywords:

Laboratory medicine
Harmonization
Transformation
Method comparison
Result adjustment formula

ABSTRACT

Background: Laboratory test results are crucial for clinical decisions, yet inconsistencies arise when measurements are not harmonized due to the lack of suitable higher-order references. This study introduces an approach to improve result comparability across different measurement systems, applicable until full metrological harmonization of the measurand is achieved.

Methods: A linear transformation formula was developed, utilizing the 2.5th and 97.5th percentiles of data from source and target methods, to adjust source method results. The feasibility of this formula was tested using carbohydrate antigen 125 (CA 125) data from two commercial assays provided by Roche Diagnostics and Abbott Diagnostics. Method comparison statistics, including difference plots and Passing-Bablok regression, were used to evaluate the transformation's effectiveness before and after adjustment. A web application, "Result Transformer," was developed to facilitate the application of the transformation process.

Results: Prior to transformation, the median relative difference between the Roche and Abbott CA 125 assays was 37.7% (95% CI: 34.5–40.8%), exceeding acceptable bias. Passing-Bablok regression yielded a slope of 1.450 (95% CI: 1.400–1.485) and an intercept of -0.83 kU/L (95% CI: -1.50 to -0.29). After adjustment using the proposed approach, the median relative difference decreased to 6.0% (95% CI: 4.3–7.7%), falling within the desirable acceptable bias goal. The slope and intercept of the regression equation improved to 1.075 (95% CI: 1.039–1.102) and -0.12 kU/L (95% CI: -0.71 to 0.19), respectively.

Conclusion: The proposed transformation method effectively improved the comparability of results from different assays, permitting a more consistent test result interpretation during patient follow-up.

1. Introduction

Laboratory results are indispensable tools in clinical decision-making, providing critical information for diagnosis, treatment, and patient monitoring[1,2]. Their interpretation often relies on comparison to reference intervals, clinical decision limits[3], or previous results[4].

The application of universal clinical decision limits, i.e., thresholds

associated with significantly increased risk of adverse outcome or diagnostic criteria for specific diseases[5], necessitates harmonized results for reliable application[6]. This is because nonequivalent results from different measurement systems can compromise the consistency of clinical decisions based on a single common limit. Similarly, the use of common reference intervals is contingent upon equivalent results from different measurement systems[7].

* Corresponding author at: General Directorate of Health Services, Rare Diseases Department, Turkish Ministry of Health, Bilkent Yerleskesi, 6001 Cadde, Universiteler Mahallesi 06800, Çankaya, Ankara, Turkey.

E-mail address: hikmetcancubukcu@gmail.com (H.C. Çubukçu).

<https://doi.org/10.1016/j.cca.2025.120355>

Received 30 April 2025; Received in revised form 8 May 2025; Accepted 8 May 2025

Available online 15 May 2025

0009-8981/© 2025 Elsevier B.V. All rights reserved, including those for text and data mining, AI training, and similar technologies.

Standardization and harmonization are critical in laboratory medicine to achieve comparable test results across measurement systems. However, both rely on calibration traceability to higher-order references (methods or materials), which are often still unavailable. Consequently, many measurands remain unharmonized, exhibiting non-equivalent results across measurement systems. As an example, a recent article has confirmed inadequate harmonization of some tumor marker measurements, such as those of Carbohydrate Antigen (CA) 125, CA 15-3, and CA 19-9, highlighting significant bias between different measurement methods[8].

CA 125 is the recommended biomarker for the evaluation and management of epithelial ovarian cancer (EOC)[9]. However, measurement procedure-based differences hinder the proper utility of CA 125. These differences in CA 125 assays stem from various factors, including the use of different antibodies, calibration procedures, and assay designs[10]. As a result, as for other non-harmonized measurands, using the same measurement system is the only recommended solution for monitoring biomarker results in the same patient[11]. However, patients should be enabled to receive interchangeable and comparable laboratory results across different providers and locations, eliminating the need to rely on a single laboratory, as a patient-centered approach prioritizes consistency regardless of geographical or logistical constraints[12].

While harmonization of measurement procedures remains the ultimate goal, current non-comparability necessitates interim but immediate solutions. The use of the International Normalized Ratio (INR) is a longstanding example in laboratory medicine of mitigating the non-comparability of results from different measurement systems through transformation, aiming to reduce variability in prothrombin time results [13].

This study proposes a transformation method applicable to wide range of measurands to obtain comparability across different measurement procedures for non-harmonized measurands, facilitating their use in patient result monitoring and clinical decision-making.

To this end, we developed a transformation approach that accounts for the linear relationship between methods and their respective percentiles derived from the data distribution of measurements from each method. The feasibility of this approach was demonstrated using CA 125 as a model measurand.

2. Materials and methods

2.1. Development of the formula

The development of the proposed transformation formula relied on some key premises: a linear relationship between two distinct methods (source and target), the presence of significant bias between these methods, and their intended use on similar patient populations.

The “source method” was identified as the measurement procedure whose results necessitated adjustment to facilitate comparison or consistent monitoring alongside results from the target method. Conversely, the “target method” served as the reference to which the source method results were adjusted to ensure consistency in inter-method comparison. Notably, the term “target” did not denote superiority but rather indicated the method whose reference intervals or result scale was used as the basis for consistent monitoring. For example, if a patient’s previous CA 125 result was obtained using Method A and her current result using Method B, and these methods yield non-equivalent results, a transformation can be applied. In this scenario, we may choose Method A as the target and then convert the results from Method B (the source) to align with Method A’s scale. Designations are arbitrary; the roles of ‘target’ and ‘source’ could be reversed without affecting the principle of the transformation.

The formula’s objective was to transform the source method’s results to achieve enough comparability with those of the target method, thereby ensuring consistent interpretation in result monitoring and

clinical decision-making. This transformation formula employed a linear approach to account for both constant and proportional bias, utilizing the 2.5th and 97.5th percentiles of the population data distributions obtained using the target and source methods.

The formula is based on two assumptions:

- I. **Linear Relationship:** A linear relationship exists between the results of the target and source methods, allowing one method’s results to be transformed into the other using a linear equation. This relationship can be confirmed either indirectly through existing literature or directly via a method comparison study.
- II. **Similar Distribution of Population Data Across Methods:** Although the 2.5th and 97.5th percentiles may differ between methods, the underlying population data for the measurand exhibit a similar shape of distribution. The similarity can be shown statistically (Kolmogorov–Smirnov test, etc.). This similarity ensures that a linear adjustment based on these percentiles can effectively align the source method’s data with that of the target method.

The linear relationship between source and target methods was represented by the following equation:

$$\text{AdjustedResult}(\text{Equivalent to target method result}) = a + b \times \text{Result}_{\text{source}} \quad (1)$$

Where a and b were the constant bias (intercept) and the proportional bias (slope), respectively. $\text{Result}_{\text{source}}$ stand for source method result.

Equation (1) was applied to the 2.5th and 97.5th percentiles of the same population data of concern for each method.

At the lower percentile (2.5th percentile):

$$L_{\text{target}} = a + b \times L_{\text{source}} \quad (2)$$

At the upper percentile (97.5th percentile):

$$U_{\text{target}} = a + b \times U_{\text{source}} \quad (3)$$

Where:

- L_{target} : Lower percentile (2.5th percentile) of the data from target method
- U_{target} : Upper percentile (97.5th percentile) of the data from target method
- L_{source} : Lower percentile (2.5th percentile) of the data from source method
- U_{source} : Upper percentile (97.5th percentile) of the data from source method

Step 1: Calculation of the proportional bias (b).

The **equation (2)** was subtracted from the **equation (3)**:

$$U_{\text{target}} - L_{\text{target}} = b \times (U_{\text{source}} - L_{\text{source}}) \quad (4)$$

b was isolated on the left side of the equation:

$$b = \frac{U_{\text{target}} - L_{\text{target}}}{U_{\text{source}} - L_{\text{source}}} \quad (5)$$

Step 2: Calculation of the constant bias (a).

The **equation (2)** was rearranged to isolate a on the left side:

$$a = L_{\text{target}} - b \times L_{\text{source}} \quad (6)$$

Step 3: Transformation of equation (1):

a) Adding constant bias formula (Equation (4)) to Equation (1):

$$\text{Adjusted Result} = (L_{\text{target}} - b \times L_{\text{source}}) + b \times \text{Result}_{\text{source}} \quad (7)$$

b) Group the terms involving b:

$$\text{Adjusted Result} = L_{\text{target}} + b \times (\text{Result}_{\text{source}} - L_{\text{source}}) \quad (8)$$

c) Adding proportional bias formula (Equation (5)) to Equation (8) to

achieve the final formula

$$AdjustedResult = L_{target} + \frac{U_{target} - L_{target}}{U_{source} - L_{source}} \times (Result_{source} - L_{source}) \quad (9)$$

2.2. Demonstration of the use of the adjustment formula

To demonstrate the efficacy of the adjustment formula, a two-stage evaluation was performed (Fig. 1). First, we conducted a simulation study to showcase the transformation method’s applicability in scenarios involving varying measurement methods on the same patient population. Second, a real-world method comparison study was undertaken to demonstrate the formula’s use across two different laboratories serving two diverse patient populations.

2.2.1. Proof-of-concept simulation study

This simulation aims to replicate the common scenario of different measurement methods yielding non-comparable data for the same patient population, to test the transformation formula when a linear relationship exists between hypothetical methods. The simulation study utilized a large dataset of serum CA 125 measurements from Istanbul Cam Sakura City Hospital (n = 19,096), obtained with the Roche Cobas e801 electrochemiluminescence immunoassay, as the target method. To simulate biased source methods, both negatively and positively biased results were generated using specific formulas, incorporating a 5 % standard measurement uncertainty, which is below minimum analytical performance specification for relative standard measurement uncertainty (6.5 %). Subsequently, the 2.5th and 97.5th percentiles were calculated for distribution of all datasets, and the simulated biased results were transformed using an adjustment formula to align with the target method. Finally, 120 paired samples were randomly selected from both the target and simulated datasets for method comparison analysis, as detailed in the following sections.

2.2.1.1. Simulation data. Two sets of source method results were

simulated to exemplify methods with negative and positive biases. The following formulas were used to generate the biased source method results:

$$Result_{negativebias} = 0.80 \times Result_{target} - 0.5$$

$$Result_{positivebias} = 1.20 \times Result_{target} + 10$$

Where:

- Result_{target}: Target method result
- Result_{negative bias}: Negatively biased source method result
- Result_{positive bias}: Positively biased source method result

To take into account the method standard measurement uncertainty, a 5 % relative variation was applied to both positively and negatively biased source method results, resulting in the final source method results, as previously reported[14]:

$$Result_{MU+negativebias} = Result_{negativebias} \times [1 + n(0,1) \times u_{rel}]$$

$$Result_{MU+positivebias} = Result_{positivebias} \times [1 + n(0,1) \times u_{rel}]$$

Where:

- Result_{MU+negative bias}: Negatively biased result with measurement uncertainty.
- Result_{MU+positive bias}: Positively biased result with measurement uncertainty.
- n(0,1): A random number generated with a normal distribution having a mean of 0 and a standard deviation of 1.
- u_{rel}: Relative standard measurement uncertainty assigned as 5 % (in the formula, expressed as fraction as 0.05)

2.2.1.2. Determination of the 2.5th and 97.5th percentiles. The 2.5th and 97.5th percentiles were calculated from the distribution of the original

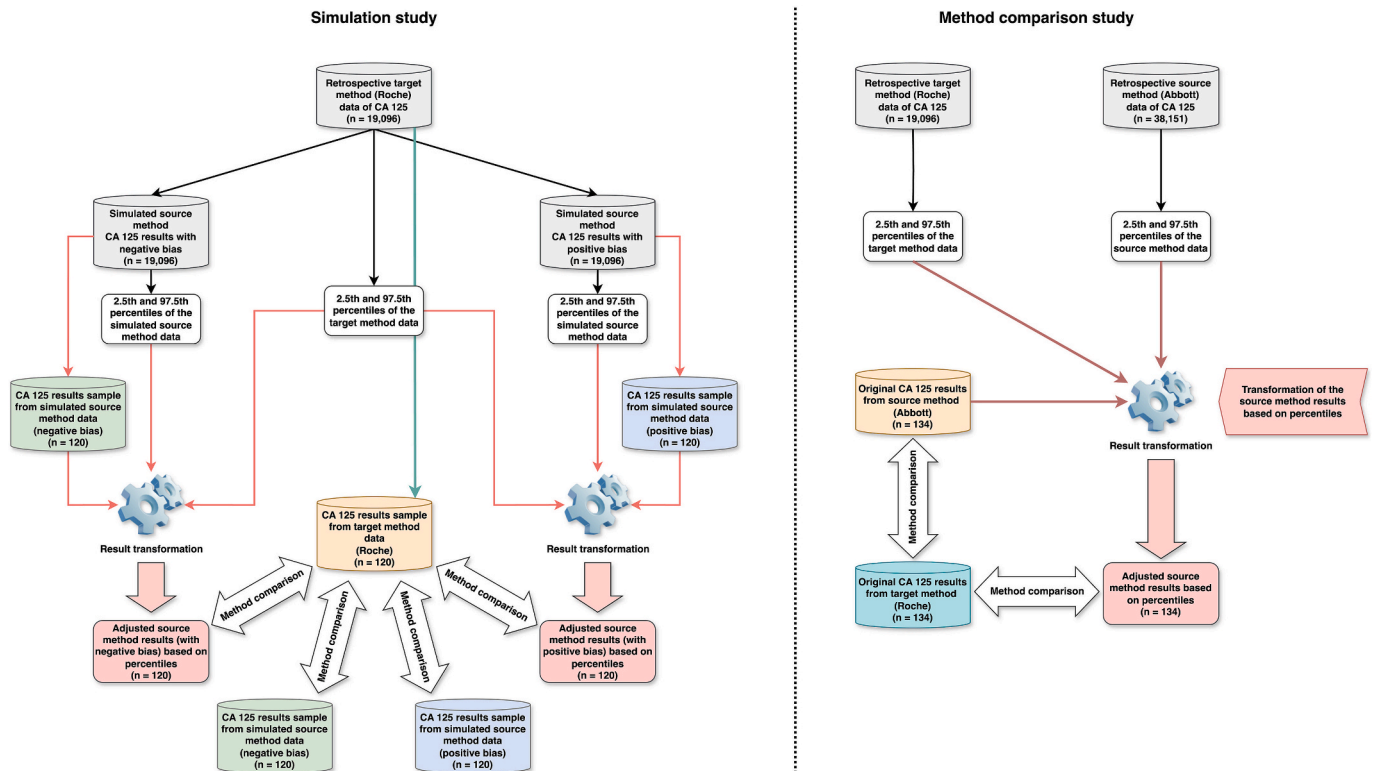


Fig. 1. Study design for transformation method evaluation.

population data, as well as for the simulated positively and negatively biased source method results.

2.2.1.3. Transformation of biased results using adjustment formula. The simulated positively and negatively biased source method results were transformed to align with the target method results using the adjustment formula (Equation (9)). This transformation was performed using the 2.5th and 97.5th percentiles of both the target and source methods.

2.2.1.4. Sampling from population and simulation data sets. A set of 120 paired results was randomly selected from the target and simulated source method datasets. The sampling was stratified based on the target method results: 60 results were selected with target method values at or below 35 kU/L (i.e., the upper reference limit for the Roche CA 125 II assay) and 60 results were selected with target method values between 35 kU/L and 1,000 kU/L. These paired samples were then used for method comparison analysis (Fig. 1).

2.2.2. Real-world method comparison study

The real-world method comparison study was conducted to evaluate the use of the formula across two laboratories measuring their patient populations by different non-harmonized assays for CA 125. The study was based on the presumption of a linear relationship between the results of these different CA 125 measurement methods, as supported by previous research [10,15].

Two datasets from laboratories from different cities using different measurement methods (one designated as the target and the other as the source for demonstration purpose) were obtained. We then confirmed that the two distributions exhibit similar patterns statistically.

The 2.5th and 97.5th percentiles were calculated for each original dataset, and these were used for result transformation. After transformation, 134 fresh serum samples were analyzed for CA 125 using both target and source methods. Using Roche assay, the biomarker concentrations ranged between 2.8 and 472 kU/L. The previously determined percentiles were then used to transform the source method CA 125 results. Finally, method comparison statistics were applied to assess the results of the transformation approach, comparing the target method results to both the original source method results and the transformed source method results.

2.2.2.1. CA 125 measurement methods and the study data. For the demonstration of the transformation method, the Roche electrochemiluminescence immunoassay for CA 125 measurement (Elecsys CA 125 II assay) was selected as the target method, while the Abbott chemiluminescent microparticle immunoassay for CA 125 measurement (Architect CA 125 II assay) was designated as the source method.

While both assays leverage OC125 and M11 antibodies, they differ in their format and detection principles. The Abbott assay uses OC125 as the capture antibody, while the Roche assay uses M11 as the capture antibody. The Abbott assay uses chemiluminescence for detection, while the Roche assay uses electrochemiluminescence.

The Abbott Architect CA 125 II assay calibrators are referenced, through an intermediate step of value transfer by Abbott AxSYM measuring system, to the Fujirebio Diagnostic standard OC 125, with a stock antigen solution value assigned using the Fujirebio CA 125 II radioimmunoassay (RIA). The Roche Elecsys CA 125 II assay is standardized against the Enzymun-Test CA 125 II method, which itself is standardized against the Fujirebio Diagnostics CA 125 II RIA. Therefore, although both manufacturers state to use the same higher-order material at the top of the traceability chain, there are some aspects that can blame the unsatisfactory harmonization between the two assays. Recent papers have reported examples showing that referring the IVD-MD calibration to the same metrologically higher-order references does not automatically lead to equivalence of results by different IVD-MDs [16,17]. The metrology approach theoretically allows for multiple traceability

schemes and leaves it up to the IVD manufacturer to develop and validate the calibration scheme. Since there is no coordination among different IVD manufacturers, they use different internal standardization procedures (meaning traceability paths) in which each step of the calibration hierarchy should be governed to obtain final unbiased results on clinical samples. In the case of CA 125 assays employed in this study, the assigned values for the two commercial calibrators, although both derived from “metrologically legitimate” traceability chains, produce nonequivalent results as probably in some steps of the traceability implementation some important source of bias has not been removed. Information sharing of employed practices would be very helpful to understand the correctness of the traceability implementation. However, exhaustive information on how IVD manufacturers have procedurally implemented the calibrator traceability to the selected higher-order reference is lacking.

Study data were collected from two laboratories located in different cities in Türkiye (Istanbul and Van), using these two CA 125 measurement methods. CA 125 serum data obtained using the target method ($n = 19,096$) were sourced from Istanbul Cam Sakura City Hospital, where the Roche Cobas 8000 e801 immunoassay module was used. Meanwhile, CA 125 serum data for the source method ($n = 38,151$) were collected from Van Yüzüncü Yıl University Hospital, utilizing the Abbott Architect i4000SR immunoassay analyzer.

Both laboratories perform daily internal quality control to check the system alignment using PreciControl Tumor Marker for Roche assay and Technopath Multichem IA Plus for Abbott assay, at two and three concentration levels, respectively. Over the past three months, Cam Sakura City Hospital (Roche) reported a CV of 4.03 % (at 28.2 kU/L) and 3.63 % (at 85.8 kU/L), while Van Yüzüncü Yıl University Hospital (Abbott) reported CVs of 1.94 % (at 22.3 kU/L), 2.77 % (at 40.1 kU/L), and 2.27 % (at 70.2 kU/L), respectively. Both laboratories participate in monthly external quality assessment (EQA) programs for CA 125. All EQA results covering the period of patient laboratory result collection were within the permissible limits of the EQA programs.

2.2.2.2. Assessing distributional similarity across data sets. The two data sets exhibited comparable histogram distribution patterns (see [Supplementary Material](#)). To further assess the similarity in distribution shapes, we randomly sampled 2,000 results from each dataset and rescaled them using the formula $((\text{result} - \text{median}) / \text{interquartilerange})$ which aligns each dataset on a common scale—centering the median at zero and setting the interquartile range to one—while preserving the original shape of the distributions. Finally, we confirmed that the two distributions of rescaled data sets were similar by performing a Kolmogorov–Smirnov test, considering p value higher than 0.05 insignificant (Kolmogorov–Smirnov statistic: 0.03, p-value = 0.43).

2.2.2.3. Determination of the 2.5th and 97.5th percentiles. The 2.5th and 97.5th percentiles were calculated for both original target (Roche) and source (Abbott) method data.

2.2.2.4. Transformation of source method results using percentiles. After obtaining percentiles for transformation, 134 fresh serum samples were analyzed for CA 125 using both target (Roche Elecsys CA 125 II assay) and source (Abbott Architect CA 125 II assay) methods. Samples were selected based on the following criteria: women aged between 18 and 90 years, CA 125 concentrations from 2 kU/L (i.e., the limit of quantification for the Roche assay) to 1,000 kU/L (i.e., the Architect dilution threshold), and absence of significant hemolysis, icterus, and lipemia.

The source method results were transformed to align with the target method results using the adjustment formula (Equation (9)). This transformation was performed using both the 2.5th and 97.5th percentiles of the target and source methods. Finally, method comparison analysis was performed to evaluate the performance of the transformation formula.

2.2.3. Evaluation of conformance between target and source method results

Method comparison was performed using difference plots and Passing-Bablok regression analysis both prior to and following the adjustment of source method results. In the simulation study, the target method served as the comparative method across all analyses, while the simulated biased results ($\text{Result}_{\text{MU}+\text{negative bias}}$, $\text{Result}_{\text{MU}+\text{positive bias}}$) and their adjusted counterparts were considered candidate method results. In the real-world method comparison study, the target method results ($\text{Result}_{\text{target}}$) were consistently used as the comparative method. The source method results ($\text{Result}_{\text{source}}$) and their adjusted values were designated as candidate method results.

The Passing-Bablok regression plots and difference plots were generated using Pandas version 1.3.5[18] and NumPy version 1.21.2 [19] and Plotly version 5.15.0[20] through the VerifyMyLab software [21]. As analytical performance specifications for bias between CA 125 assays, $\pm 10.1\%$ (minimum quality level) and $\pm 6.7\%$ (desirable quality level) were employed [22]. In difference plots, the minimum acceptable bias was depicted to evaluate assay conformance. For the Passing-Bablok regression analysis, clinical equivalence was deemed acceptable if the 95% confidence interval (CI) of the intercept included 0, and the slope fell within the relative acceptable bias range of 10.1% (corresponding to slope values between 0.899 and 1.101).

2.2.4. Development of the web application

A web application, "Result Transformer," was developed to streamline the transformation and calculation process for future applications. The application was built using Python 3.12[23], Pandas 2.2.3[18]; NumPy 2.2.3[19]; Plotly 6.0.0[20], and Streamlit 1.29.0[24], with Microsoft Visual Studio Code version 1.77.3 as the development environment. Both the source code and the application's URL are publicly available on GitHub at https://github.com/hikmetc/result_transformer. The web application itself is freely accessible at <https://resulttransformer.streamlit.app>.

3. Results

3.1. Simulation study

The original population data had 2.5th and 97.5th percentiles of 4.1–263.6 kU/L. For the positively biased source method results ($\text{Result}_{\text{MU}+\text{positive bias}}$), the percentiles were found to be 14.7–329.7 kU/L, while the negatively biased source method results ($\text{Result}_{\text{MU}+\text{negative bias}}$)

exhibited percentiles of 2.8–212.7 kU/L.

For the positively biased source method results, the adjustment formula with percentiles was applied as follows:

$$\text{AdjustedResult} = 4.1 + \frac{263.6 - 4.1}{329.7 - 14.7} \times (\text{Result}_{\text{MU}+\text{positive bias}} - 14.7) \quad (9)$$

For the negatively biased source method results, the adjustment formula with percentiles was applied as follows:

$$\text{AdjustedResult} = 4.1 + \frac{263.6 - 4.1}{212.7 - 2.8} \times (\text{Result}_{\text{MU}+\text{positive bias}} - 2.76) \quad (9)$$

Due to the non-parametric distribution of data, as confirmed by the Kolmogorov-Smirnov test, the median relative difference was used as the difference plot metric. Both simulated positive ($y = 1.2x + 10$) and negative ($y = 0.8x - 0.5$) biased source methods showed significant deviations from the target method, failing acceptability criteria for median relative difference, slope, and intercept (Table 1).

The relative difference plots (Figs. 2A and 3A) and regression analyses (Fig. 2C and Fig. 3C) illustrate these deviations. Following adjustment, both the simulated positive and negative biased source methods met acceptability criteria at the desirable quality level (Table 1), as shown in the difference plots (Figs. 2B and 3B) and the Passing-Bablok regression analyses (Figs. 2D and 3D).

3.2. Real-world method comparison study

For the target method, the 2.5th and 97.5th percentiles were determined to be 4.1 kU/L and 263.6 kU/L, respectively. For the source method, the corresponding percentiles were 5.0 kU/L and 354.4 kU/L.

The percentile-based transformation was applied as follows:

$$\text{AdjustedSourceMethodResult} = 4.1 + \frac{263.6 - 4.1}{354.4 - 5.0} \times (\text{Result}_{\text{source}} - 5.0)$$

In the real-world comparison of Roche (target) and Abbott (source) CA 125 methods, the unadjusted Abbott method exhibited significant bias (Table 1), with a 37.7% median relative difference (Fig. 4A) and a slope of 1.450 (Fig. 4C). After adjustment, the Abbott CA 125 method demonstrated acceptable equivalence, with a 6.0% median relative difference, meeting the acceptability criteria at the desirable quality level (Fig. 4B). The adjusted method also achieved a slope of 1.075, meeting the minimum quality level criteria, and an intercept of -0.12

Table 1
Method comparison results for CA 125.

	N	Difference plot	Passing-Bablok regression equation		Clinical Equivalence Acceptability*		
		Median relative difference (95% CI)	Slope (95% CI)	Intercept, kU/L (95% CI)	Relative difference	Slope	Intercept
Simulation study							
Target method vs. simulated positive biased method ($y = 1.2x + 10$)	120	50.3% (40.1, 60.5)	1.184 (1.165, 1.209)	10.0 (9.7, 10.4)	No	No	No
Target method vs. adjusted simulated positive biased method ($y = 1.2x + 10$)	120	-0.6% ($-1.4, -0.3$)	0.975 (0.959, 0.996)	0.3 (0.0, 0.5)	Yes	Yes	Yes
Target method vs. simulated negative biased method ($y = 0.8x - 0.5$)	120	-23.8% ($-24.6, -23.0$)	0.789 (0.778, 0.804)	-0.4 ($-0.6, -0.3$)	No	No	No
Target method vs. adjusted simulated negative biased method ($y = 0.8x - 0.5$)	120	-1.1% ($-1.6, -0.5$)	0.976 (0.962, 0.994)	0.1 ($-0.1, 0.3$)	Yes	Yes	Yes
Real-world method comparison study							
Target method (Roche) vs. source method (Abbott)	134	37.7% (34.5–40.8)	1.450 (1.400, 1.485)	-0.8 ($-1.5, -0.3$)	No	No	No
Target method (Roche) vs. adjusted source method (Abbott)	134	6.0% (4.3, 7.7)	1.075 (1.039, 1.102)	-0.1 ($-0.7, 0.2$)	Yes	Yes**	Yes

*Defined as: median relative difference within $\pm 10.1\%$ for the minimum quality level and within $\pm 6.7\%$ for the desirable quality level; slope between 0.889 and 1.101 (corresponding to $\pm 10.1\%$ proportional bias) for the minimum quality level, and between 0.933 and 1.067 (corresponding to $\pm 6.7\%$ proportional bias) for the desirable quality level; confidence interval of the intercept including zero.

** At minimum quality level.

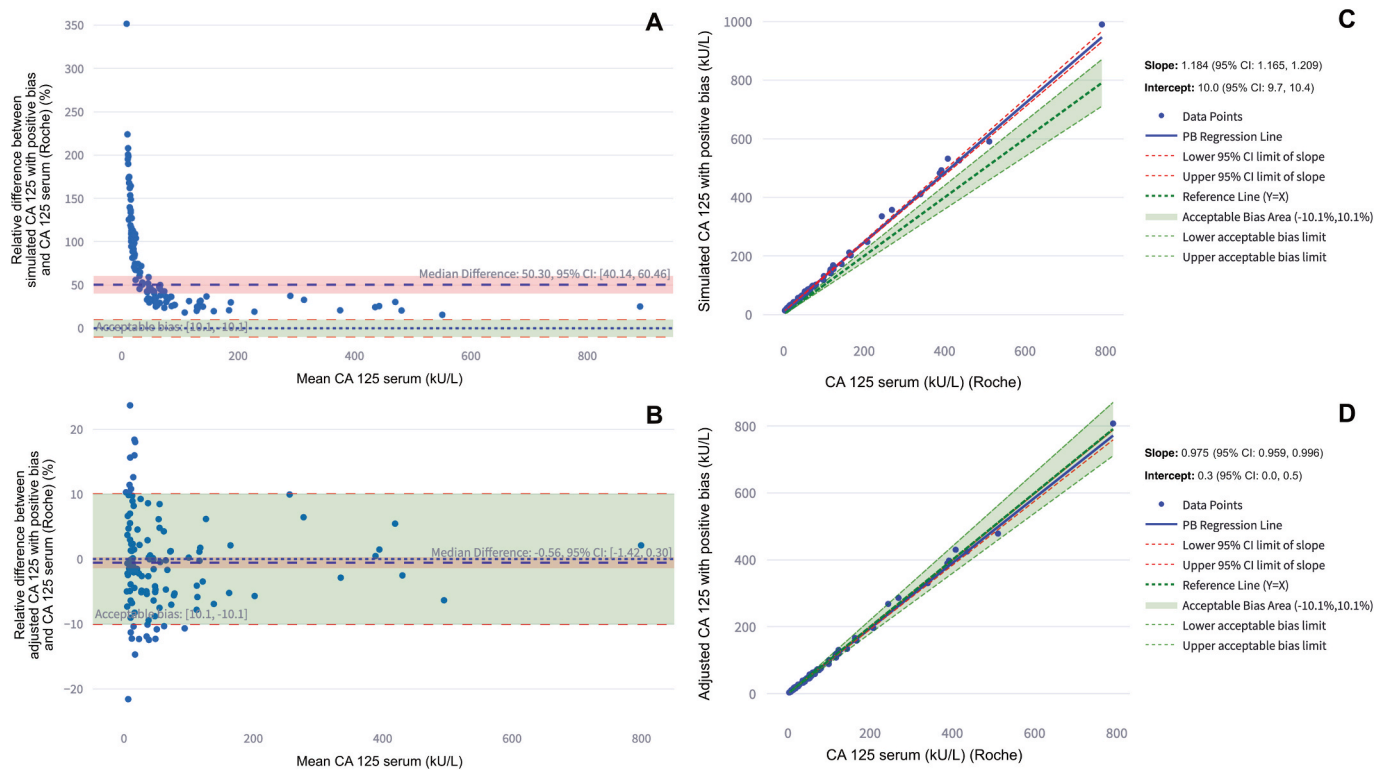


Fig. 2. Left: Relative difference plots of Roche immunoassay for measured and simulated CA 125 results with positive bias (A) and simulated CA 125 results adjusted using the 2.5–97.5 percentile limits (B), plotted against the mean of the two sets of results. Right: Passing-Bablok regression plots comparing Roche immunoassay for measured and simulated CA 125 results with positive bias (C) and simulated CA 125 results adjusted using the 2.5th and 97.5th percentile limits (D).

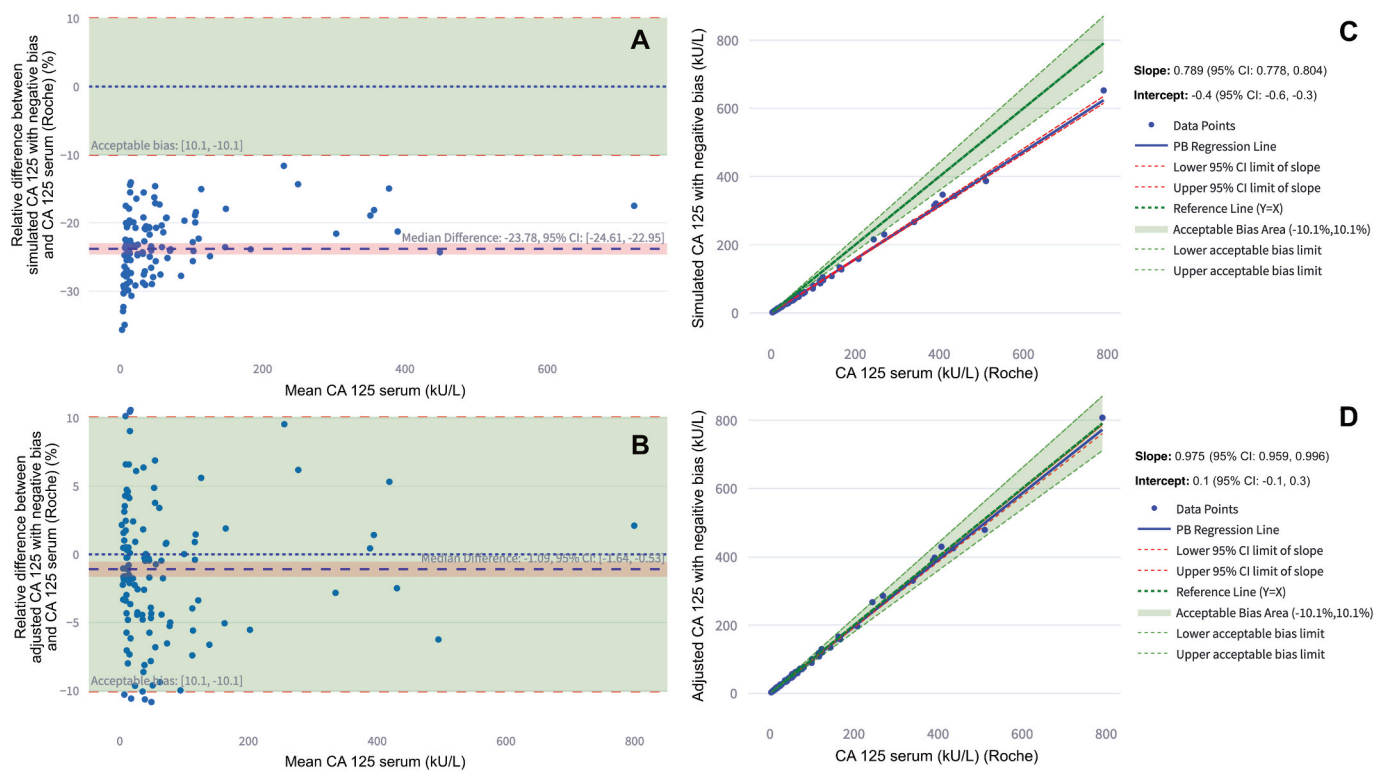


Fig. 3. Left: Relative difference plots of Roche immunoassay for measured and simulated CA 125 results with negative bias (A) and simulated CA 125 results adjusted using the 2.5–97.5 percentile limits (B), plotted against the mean of the two sets of results. Right: Passing-Bablok regression plots comparing Roche immunoassay for measured and simulated CA 125 results with negative bias (C) and simulated CA 125 results adjusted using the 2.5th and 97.5th percentile limits (D).

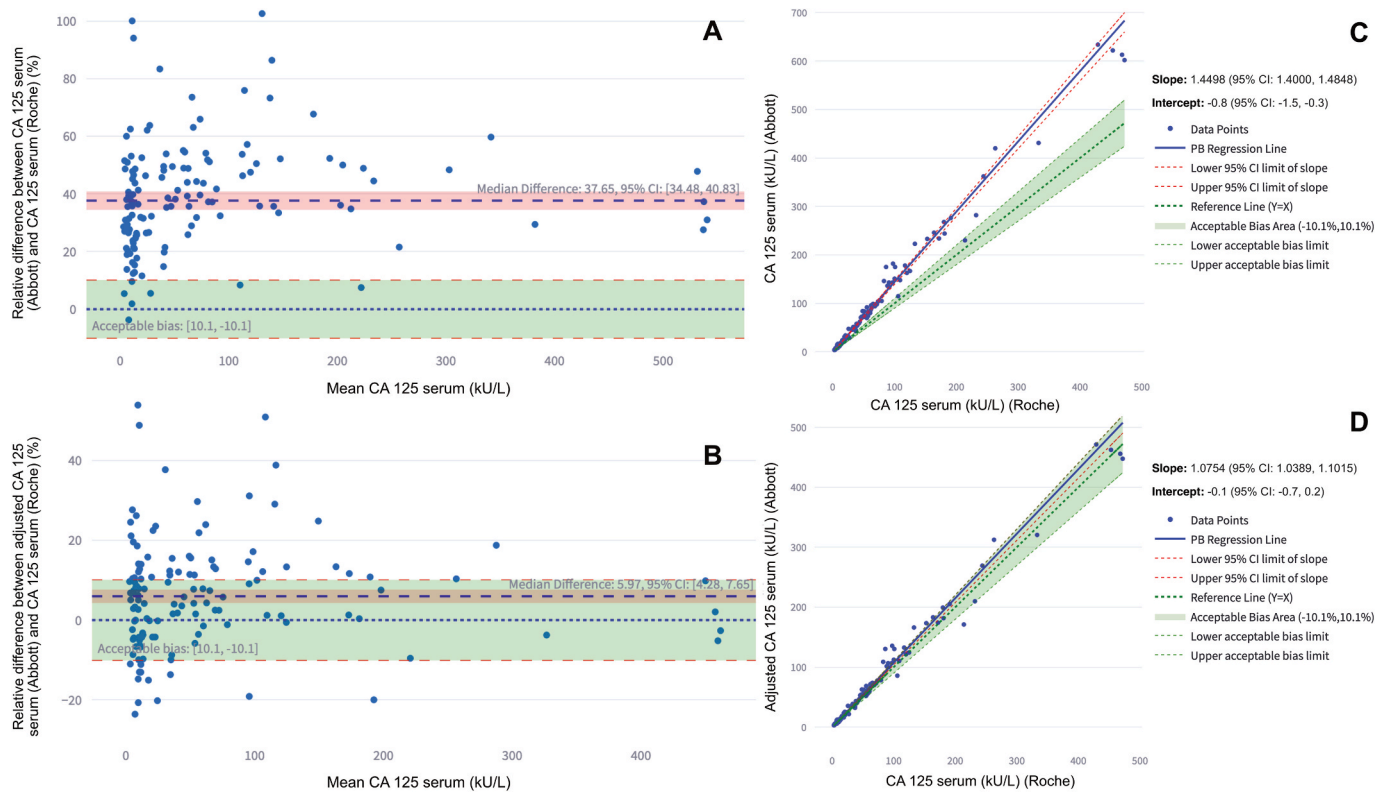


Fig. 4. Left: Relative difference plots for Roche and Abbott immunoassays for CA 125 measurements in serum, plotted against the mean of the two sets of results. (A) Difference between Roche and Abbott results before any adjustment. (B) Difference between Roche and Abbott results adjusted using the proposed approach. Right: Passing-Bablok regression plots comparing serum CA 125 results by Roche and Abbott immunoassays: (C) before any adjustment; (D) after adjustment using the proposed approach.

kU/L, with a confidence interval including zero (Fig. 4D).

4. Discussion

In laboratory medicine, achieving equivalent results from different measurement systems can theoretically be attained through standardization and harmonization procedures. However, for many analytes, standardization remains challenging due to factors such as the heterogeneity of complex molecules[25] and the lack of higher-order reference materials and procedures[26]. Even when reference materials are available, their non-commutability with patient samples can lead to inaccurate results when used for calibration[27]. In such cases, corrective actions, such as developing correction factors and protocols, are necessary to reduce non-commutability bias[28].

When standardization to SI (International System of Units) is not possible, harmonization can be employed to achieve agreement between measurement procedures. This involves using the results of a single measurement procedure or the trimmed mean of multiple procedures to establish consensus values for a set of commutable materials[26,29]. Subsequent recalibration using various statistical approaches is then applied to harmonize the results[29,30].

Despite ongoing harmonization efforts, many measurands are still non-harmonized, and even for those that have been theoretically harmonized, non-equivalent results are often observed across different measurement systems[16,17,31]. Clinicians must be aware of these limitations and interpret results cautiously while considering the specific assay used.

First-generation CA 125 assays utilized the OC125 monoclonal antibody to detect the respective antigen. To improve sensitivity and specificity, second-generation assays introduced a double-determinant monoclonal antibody targeting the OC125 and M11 antigenic domains [32]. Although studies have demonstrated a linear relationship between

methods[10,15], CA 125 assays still exhibit a significant inter-method bias[32].

In this study, we compared two second-generation CA 125 assays that employ different measurement principles, but are created traceable to the same higher-order reference. The Roche electrochemiluminescent immunoassay utilizes M11 as the capture antibody, whereas the Abbott chemiluminescent microparticle immunoassay uses OC125 as the capture antibody. These differences in measurement principles and some issues in the implementation of the calibration traceability likely contribute to discrepancies between the two methods, as evidenced by a median relative difference of 37.7%. However, the commonality in their assay format likely contributes to the strong linearity between results, enabling effective transformation.

According to CLSI EP09c and EP28 guidelines, when a candidate measurement procedure produces non-equivalent results compared to a comparative method, separate reference intervals should be established for the candidate method[33,34]. Non-equivalent results are commonly observed for measurands that are neither harmonized nor standardized. In such cases, distinct reference intervals must be determined for each measurement procedure rather than relying on a common reference interval. For example, given the lack of result comparability mentioned above, CA 125 reference intervals vary depending on the specific assay used. The 97.5th percentile upper reference limits, determined using the same cohort of healthy women, have been reported to range from 22 kU/L to 45 kU/L[32]. Similarly, in our study, in addition to the non-comparable CA 125 results observed between Roche and Abbott CA 125 assays, we identified different 97.5th percentile upper reference limits using RefineR: 36.5 kU/L for the Roche assay and 42.8 kU/L for the Abbott assay (data not shown). Our findings highlighted the need for harmonization efforts for CA 125 to reduce variability across assays to support the effective use of common reference intervals and decision levels, and ensure consistent interpretation of results across different

measurement systems[8].

Although recalibration or other harmonization efforts remain the definitive solutions, such endeavors are often reliant on stakeholder engagement and the availability of commutable reference materials. In practical settings where harmonization is difficult and slow to be implemented, the transformation formula we presented in this study may offer an interim, patient-centered approach that can be readily deployed to immediately improve the clinical utility of non-harmonized measurands, such as CA 125 and other tumor markers[8]. Our approach takes inspiration from the well-established INR system, which has successfully harmonized prothrombin time measurements for patients undergoing vitamin K antagonist treatment [13]. Our proposed method can be applied to a broader range of measurands with a demonstrable linear relationship between measurement systems, as exemplified for CA 125 in our study.

The simulation study replicated scenarios where different measurement methods yield non-equivalent results for the same patient population. By introducing both positively and negatively biased results and subsequently applying the adjustment formulas, significant improvements in agreement with the target method were observed. Both biased datasets initially failed the acceptability criteria for bias, but after transformation, they met all predefined criteria, suggesting that the adjustment formula effectively minimizes systematic bias. The real-world method comparison further validated the formula's applicability across laboratories serving different patient populations and utilizing distinct measurement systems but having comparable histogram patterns of data distribution. Despite initial significant bias between the unadjusted Abbott (source) and Roche (target) methods, the transformation, utilizing 2.5th and 97.5th percentiles, resulted in acceptable comparability. This was reflected by the substantial reduction in median relative difference and the improved alignment of slope and intercept parameters (Table 1). The ability to achieve acceptable method comparison results across different laboratory settings and patient populations underscores the practical utility of the proposed adjustment formula for medical laboratories from different locations. The ability to align diverse measurement systems fosters consistency in clinical decision-making and patient monitoring, crucial when a measurand, such as CA 125, serves as a biomarker in detecting and monitoring disease.

While the premises of the adjustment formula can be easily met in many cases, some important factors should be carefully considered. Linearity is a key assumption that must be satisfied for the proper utility of our approach. However, linearity is not always present and therefore necessitates indirect confirmation from published literature (if available) or direct confirmation through a prior method comparison study, as mentioned in this study. Inherent differences in assay design and analytical principles, such as variations in antibody selectivity, can contribute to non-linear relationships between measurement systems [35]. Therefore, even if proportional or constant bias exists, a linear relationship between the measurement systems is essential for proper use of the adjustment formula. Furthermore, the adjustment formula is expected to be more effective for laboratories serving similar populations, as this ensures that proper transformation can be achieved using results corresponding to 2.5th and 97.5th percentiles obtained in similar populations. Laboratories serving different populations may encounter datasets with varying distributions for the same measurands, which can hinder proper adjustment. In such cases, it is essential for laboratories to ensure that their datasets exhibit a similar distribution pattern to those of their counterparts before to apply our approach. Moreover, in case of reagent/calibrator lot changes causing a significant bias, the transformation percentiles shall be re-established. Finally, while this approach can momentarily improve result comparability, it should not replace harmonization efforts, which remain fundamental for standardize laboratory results globally[17].

In support of applying our proposed transformation method into routine practice, we developed the “Result Transformer” web

application, providing a user-friendly platform that enables rapid, automated calculation of adjusted results. By integrating established percentiles, as well as input values for individual patient samples, the application streamlines the transformation process without necessitating in-depth statistical expertise, thus accelerating the use of this interim harmonization strategy.

5. Conclusion

In conclusion, this study introduced and validated a simple, percentile-based transformation formula to address the persistent challenge of non-equivalent results from different measurement procedures, exemplified by the CA 125 assays. By leveraging the 2.5th and 97.5th percentiles, our method effectively aligned source method results with target method results for CA 125. This approach offers a practical, interim solution for enhancing result comparability, particularly for measurands where standardization and harmonization efforts are slow and challenging. The development of the “Result Transformer” web application further facilitates the adoption of this method into clinical practice, ensuring consistent interpretation, monitoring, and clinical decision-making across different non-harmonized measurement procedures. Although this approach does not replace the need for harmonization initiatives, it provides a valuable tool for improving patient care by minimizing measurement-dependent differences until harmonization of test results is achieved.

Research ethics

This study was approved by the Ordu University Non-Interventional Scientific Research Ethics Committee with the decision number 2024/221 on December 20, 2024.

Author contributions

The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Research funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Hikmet Can Çubukçu: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Oğuzhan Zengi:** Resources, Investigation, Data curation. **Hamit Hakan Alp:** Investigation, Data curation. **Murat Cihan:** Project administration, Investigation, Data curation. **Kamil Taha Uçar:** Investigation, Data curation. **Marc Thelen:** Writing – review & editing, Supervision, Project administration. **Mauro Panteghini:** Writing – review & editing, Supervision, Project administration, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cca.2025.120355>.

Data availability

The authors do not have permission to share data.

References

- [1] S. Ferraro, F. Braga, M. Panteghini, Laboratory medicine in the new healthcare environment, *Clin Chem Lab Med.* 54 (4) (2016) 523–533.
- [2] U.P. Rohr, C. Binder, T. Dieterle, F. Giusti, C.G. Messina, E. Toerien, et al., The value of in vitro diagnostic testing in medical practice: a status report, *PLoS One* 11 (3) (2016) e0149856.
- [3] K. Doyle, D.R. Bunch, Reference intervals: past, present, and future, *Crit Rev Clin Lab Sci.* 60 (6) (2023) 466–482.
- [4] E.W. Randell, S. Yenice, Delta Checks in the clinical laboratory, *Crit Rev Clin Lab Sci.* 56 (2) (2019) 75–97.
- [5] Y. Ozarda, K. Sikaris, T. Streichert, J. Macri, Distinguishing reference intervals and clinical decision limits - A review by the IFCC Committee on Reference Intervals and Decision Limits, *Crit Rev Clin Lab Sci.* 55 (6) (2018) 420–431.
- [6] M.W.G. Chapter, Harmonization of results among laboratories, in: W. Clarke, M. A. Marzinke (Eds.), *Contemporary Practice in Clinical Chemistry* (fourth Edition), 17, Academic Press, 2020, pp. 289–299.
- [7] W.G. Miller, N. Greenberg, Harmonization and Standardization: Where Are We Now? *J Appl Lab Med.* 6 (2) (2021) 510–521.
- [8] H.H. van Rossum, S. Holdenrieder, B. Ballieux, T.C. Badrick, Y.M. Yun, C. Zhang, et al., Investigating the Current Harmonization Status of Tumor Markers Using Global External Quality Assessment Programs: A Feasibility Study, *Clin Chem.* 70 (4) (2024) 669–679.
- [9] W.L. Yang, Z. Lu, R.C. Bast Jr., The role of biomarkers in the management of epithelial ovarian cancer, *Expert Rev Mol Diagn.* 17 (6) (2017) 577–591.
- [10] G. Yan, H. Ju, Z. Liang, T. Zhang, Technical and clinical comparison of two fully automated methods for the immunoassay of CA 125 in serum, *J Immunol Methods.* 225 (1–2) (1999) 1–8.
- [11] M. Martin, P. Blockx, Comparison of one first and three second generation methods for the determination of CA 125, *Anticancer Res.* 17 (4b) (1997) 3171–3175.
- [12] M. Plebani, Harmonization in laboratory medicine: Requests, samples, measurements and reports, *Crit Rev Clin Lab Sci.* 53 (3) (2016) 184–196.
- [13] A. Dorgalaleh, E.J. Favaloro, M. Bahraini, F. Rad, Standardization of Prothrombin Time/International Normalized Ratio (PT/INR), *Int J Lab Hematol.* 43 (1) (2021) 21–28.
- [14] H.C. Çubukçu, F. Vanstapel, M. Thelen, L.M. van Schroyen, F.A. Bernabeu-Andreu, P. Meško Brguljan, et al., APS calculator: a data-driven tool for setting outcome-based analytical performance specifications for measurement uncertainty using specific clinical requirements and population data, *Clin Chem Lab Med.* 62 (4) (2024) 597–607.
- [15] E.M. Davelaar, G.J. van Kamp, R.A. Verstraeten, P. Kenemans, Comparison of seven immunoassays for the quantification of CA 125 antigen in serum, *Clin Chem.* 44 (7) (1998) 1417–1422.
- [16] M. Panteghini, An improved implementation of metrological traceability concepts is needed to benefit from standardization of laboratory results, *Clin Chem Lab Med* (2024).
- [17] M. Panteghini, M. Krintus, Establishing, evaluating and monitoring analytical quality in the traceability era, *Crit Rev Clin Lab Sci.* 1–34 (2025).
- [18] Team TPD. *pandas-dev/pandas: Pandas.* latest ed: Zenodo; 2020.
- [19] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, et al., Array programming with NumPy, *Nature* 585 (7825) (2020) 357–362.
- [20] Plotly Technologies Inc. Collaborative data science Publisher: Plotly Technologies Inc. Montréal, QC, 2015 [cited 2023 June 21, 2023]. Available from: <https://plot.ly>.
- [21] Çubukçu HC. *VerifyMyLab 2025* [Available from: <https://verifymylab.streamlit.app/>].
- [22] Aarsand A, Fernandez-Calle P, Webster C, Coskun A, Gonzales-Lao E, Diaz-Garzon J, et al. The EFLM Biological Variation Database [updated April 16, 2020. Available from: <https://biologicalvariation.eu/>].
- [23] G. Van Rossum, F.L. Drake, Python 3 Reference Manual, CreateSpace, Scotts Valley, CA, 2009, p. 2009.
- [24] Streamlit • A faster way to build and share data apps [cited 2023 April 12, 2023]. Available from: <https://streamlit.io/>.
- [25] M. Plebani, Harmonization in laboratory medicine: more than clinical chemistry? *Clin Chem Lab Med.* 56 (10) (2018) 1579–1586.
- [26] H.W. Vesper, G.L. Myers, W.G. Miller, Current practices and challenges in the standardization and harmonization of clinical laboratory tests, *Am J Clin Nutr.* 104 Suppl 3(Suppl 3) (2016) 907s–s912.
- [27] F. Braga, M. Panteghini, Commutability of reference and control materials: an essential factor for assuring the quality of measurements in Laboratory Medicine, *Clin Chem Lab Med.* 57 (7) (2019) 967–973.
- [28] W.G. Miller, J. Budd, N. Greenberg, C. Weykamp, H. Althaus, H. Schimmel, et al., IFCC Working Group Recommendations for Correction of Bias Caused by Noncommutability of a Certified Reference Material Used in the Calibration Hierarchy of an End-User Measurement Procedure, *Clin Chem.* 66 (6) (2020) 769–778.
- [29] A. Clerico, A. Ripoli, G.C. Zucchelli, M. Plebani, Harmonization protocols for thyroid stimulating hormone (TSH) immunoassays: different approaches based on the consensus mean value, *Clin Chem Lab Med.* 53 (3) (2015) 377–382.
- [30] L.M. Thienpont, K. Van Uytfganghe, L.A.C. De Grande, D. Reynders, B. Das, J. D. Faix, et al., Harmonization of Serum Thyroid-Stimulating Hormone Measurements Paves the Way for the Adoption of a More Uniform Reference Interval, *Clin Chem.* 63 (7) (2017) 1248–1260.
- [31] F. Braga, S. Pasqualetti, E. Frusciant, F. Borrillo, M. Chibireva, M. Panteghini, Harmonization Status of Serum Ferritin Measurements and Implications for Use as Marker of Iron-Related Disorders, *Clin Chem.* 68 (9) (2022) 1202–1210.
- [32] S.K. Mongia, M.L. Rawlins, W.E. Owen, W.L. Roberts, Performance characteristics of seven automated CA 125 assays, *Am J Clin Pathol.* 125 (6) (2006) 921–927.
- [33] Clsi, Measurement Procedure Comparison and Bias Estimation Using Patient Samples, 3rd ed. CLSI guideline EP09c., Clinical and Laboratory Standards Institute, Wayne, PA, 2018.
- [34] C.L.S.I. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third Edition. CLSI Document EP28-A3c, Clinical and Laboratory Standards Institute, Wayne, PA, 2008.
- [35] S.K. Van Houcke, S. Van Aelst, K. Van Uytfganghe, L.M. Thienpont, Harmonization of immunoassays to the all-procedure trimmed mean - proof of concept by use of data from the insulin standardization project, *Clin Chem Lab Med.* 51 (5) (2013) e103–e105.