Huub H. Van Rossum*, Stefan Holdenrieder, Yeo-Min Yun, Dina Patel, Marc Thelen, Junghan Song, Nick Unsworth, Katherine Partridge, Melanie Moore, Wei Cui, Lakshmi Ramanathan, Qing H. Meng, Bart E.P.B. Ballieux, Catharine Sturgeon and Hubert Vesper

# External quality assessment-based tumor marker harmonization simulation; insights in achievable harmonization for CA 15-3 and CEA

https://doi.org/10.1515/cclm-2024-0696
Received June 14, 2024; accepted August 31, 2024;
published online September 20, 2024

**Abstract**

**Objectives:** CA 15-3 and CEA are tumor markers used in routine clinical care for breast cancer and colorectal cancer, among others. Current measurement procedures (MP) for these tumor markers are considered to be insufficiently harmonized. This study investigated the achievable harmonization for CA 15-3 and CEA by using an *in silico* simulation of external quality assessment (EQA) data from multiple EQA programs using patient-pool based samples.

**Methods:** CA 15-3 and CEA data from SKML (2021), UK NEQAS (2020–2021) and KEQAS (2020–2021) were used. A harmonization protocol was defined in which MPs that were considered equivalent were used to value assign EQA samples, and recalibration was only required if the MP had a bias of >5 % with value assigned EQA. Harmonization status was assessed by determining the mean level of agreement and residual variation by CV (%).

**Results:** Only MPs from Abbott, Beckman, Roche and Siemens were available in all EQA programs. For CA 15-3, recalibration was proposed for Beckman MP only and for CEA, recalibration was proposed for Siemens MP only. When the harmonization procedures were applied, for CA 15-3 the pre-harmonization mean bias range per MP was reduced from −29.28 to 9.86 %, into −0.09–0.12 % after harmonization. For CEA, the mean bias range per MP was reduced from −23.78 to 2.00 % pre-harmonization to −3.13–1.42 % post-harmonization.

**Conclusions:** The present study suggests that a significant improvement in the harmonization status of CA 15-3 and CEA may be achieved by recalibration of a limited number of MPs.

**Keywords:** CEA; CA15-3; harmonization; tumor marker; standardization

## Introduction

Several blood-based tumor markers are available to aid in the clinical follow-up of various types of cancer [1–3]. Two of these tumor markers are cancer antigen (CA) 15–3 and carcinoembryonic antigen (CEA). The CA 15-3 tumor marker detects soluble forms of MUC-1 and is available in various sandwich immunoassay designs [4, 5]. The most widely used antibody combination for the CA15–3 immunoassay is a combination of

***Corresponding author: Huub H. Van Rossum**, Department of Laboratory Medicine, Netherlands Cancer Institute, Amsterdam, Plesmanlaan 121 1066CX, The Netherlands, E-mail: h.v.rossum@nki.nl
**Stefan Holdenrieder,** Institute of Laboratory Medicine, Munich Biomarker Research Center, Deutsches Herzzentrum, Munich, Germany
**Yeo-Min Yun,** Department of Laboratory Medicine, Konkuk University Medical Center, Seoul, Republic of South Korea
**Dina Patel,** UK NEQAS Immunology, Immunochemistry & Allergy, Northern General Hospital, Sheffield, UK
**Marc Thelen,** SKML, Nijmegen, The Netherlands; and Department of Laboratory Medicine of the Radboud University Medical Center, Nijmegen, The Netherlands. https://orcid.org/0000-0003-1771-669X
**Junghan Song,** Department of Laboratory Medicine, Seoul National University Bundang Hospital and College of Medicine, Seongnam, Republic of South Korea
**Nick Unsworth and Catharine Sturgeon,** UK NEQAS [Edinburgh], Department of Laboratory Medicine, Royal Infirmary of Edinburgh, Edinburgh, UK
**Katherine Partridge and Melanie Moore,** Therapeutic Reference Materials,Medicines and Healthcare Products Regulatory Agency (MHRA), South Mimms, Hertfordshire, UK
**Wei Cui,** Department of Laboratory Medicine, National Cancer Center/ National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
**Lakshmi Ramanathan,** Clinical Chemistry Service, Department of Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA
**Qing H. Meng,** Department of Laboratory Medicine, Division of Pathology and Laboratory Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
**Bart E.P.B. Ballieux,** Department of Clinical Chemistry, Leiden University Medical Center, Leiden, The Netherlands
**Hubert Vesper,** Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA, USA

the DF3 and 115D8 antibodies, both of which have well-characterized MUC-1 epitope binding sites [6–8]. Alternatively, the CA 27.29 test uses a competitive immunoassay design that uses a single antibody and epitope recognition site to measure soluble MUC-1. Elevated levels of CA 15-3 are found in the majority of patients with advanced breast cancer and have been shown to have prognostic value [5, 9, 10]. However, at earlier stages of breast cancer CA 15-3 lacks sensitivity and specificity for the detection of breast cancer. Elevated levels of CA15–3 concentrations have been found in other cancers, such as ovarian, pancreatic, and lung cancer, as well as in benign conditions such as hepatitis, pneumonitis, and vitamin B12 deficiency [5, 11]. Recent clinical guidelines do not specifically mention the use of CA 15-3, but do refer to tumor markers (CA 15-3, CEA and CA 27–29) as part of the clinical work-up of metastatic breast cancer disease and use as adjunctive assessments to help make decisions regarding metastatic breast cancer [12, 13]. Particularly in the absence of readily measurable disease, an increase in CA 15-3, CA 27.29 or CEA may be used to indicate treatment failure [14].

CEA is a large complex glycoprotein first identified by Gold and Freedman in 1965, where it was found in fetal colon extracts and human colon carcinomas [15, 16]. Since its discovery, the immunological heterogeneity of CEA has been studied and described, and differences between measurement procedures (MP) have been investigated [17–19]. For this tumor marker, a WHO International Reference Preparation (IRP) (73/601) is available [20], but used by a limited number of assay manufacturers. It has been shown that the current MPs are insufficiently harmonized [18, 19, 21]. Clinically, CEA has a profound role in the follow-up of both localized and advanced colon cancer and its clinical use is recommended by several clinical guidelines [22–24]. Furthermore, the clinical use of CEA is recommended or mentioned in clinical guidelines for breast cancer [2, 12] and medullary thyroid carcinoma [3], and its clinical utility has been demonstrated for among others, advanced lung cancer [25, 26], and cancers of the upper gastrointestinal tract [27]. In particular, the clinical guidelines for colorectal cancer include MP independent cut-offs. However, the relevance of the previously observed MP differences on this recommendation is unknown and generally, clinicians are not aware of this issue.

A previous study by the authors examined the current harmonization status of six widely used tumor markers. It was concluded that the harmonization status of CA 15-3 and CEA, was insufficient. One common cause of disagreement among assays is inconsistent calibration among assays, which leads to so called ‘calibration bias’. The calibration bias is a systematic bias affecting all measurement results in a consistent, predictable manner. It therefore can be minimized by applying mathematical formulas to the measurement results. Here, and as a follow-up of the first EQA based study, an *in silico* harmonization simulation approach has been used to estimate the impact of consistent calibration of assays on the inter-assay variability.

# Materials and methods

## EQA samples

The patient-pool based EQA samples provided for the previous harmonization study, were used [21]. For CEA, 52 samples (KEQAS, SKML and UK NEQAS) and for CA 15-3 29 samples (SKML and UK NEQAS) were available. EQA data are from 2020 to 2021 for KEQAS and UK NEQAS and 2021 for SKML. Only MPs available in all EQA programs were included in the study, which included the MPs from Abbott (Alinity and Architect), Beckman (DxI/Access), Roche (Cobas) and Siemens (Atellica).

## Harmonization protocol

The *in silico* harmonization simulation protocol used for the CA 15-3 and CEA tumor markers, is shown in Table 1. The protocol consists of 5 steps.

In the first step, the systematic difference between each pair of MPs was investigated both by Passing-Bablok regression and Bland-Altman difference analysis. Also, the correlation was investigated by calculation of Spearman’s correlation coefficient. Secondly, MPs were selected to be used to value assign EQA values. Therefore, MPs that had a systematic difference of less than 5 % were considered equivalent. MP with a mean relative difference, as obtained by pre-harmonization Bland-Altman analysis within 5 %, a slope obtained by Passing-Bablok within the 0.95–1.05 range or its 95 % CI including 1.00 were considered equivalent. The two reasons to use the pragmatic 5 % criterion for both CEA and CA15–3 were that i) the lowest optimum allowable bias of 4.6 % for CA-15–3, based on biological variation was around this value [28] and ii) the obtained inaccuracy of estimating the slopes from the regression analysis, representing by the 95 % CI, was more or less of this magnitude. The selected equivalent MPs were the considered as “reference MPs” to together value assign each EQA sample by averaging the EQA values of these MP. If no pair of MPs was considered equivalent, the EQA value assignment was based on the consensus mean of all MPs. As a third step, a Passing-Bablok regression and Bland-Altman mean difference analysis are performed between the value-assigned EQA values and each individual MP. Fourth, based on the regression analysis obtained in the previous step, MPs that were not significantly biased (slope within 0.95–1.05 range), were selected to not

**Table 1:** EQA-based *in silico* harmonization simulation procedure.

| |
|---|
| **1) Determine systematic bias between each pair of MPs** |
| **1A Using Passing-Bablok regression analysis**<br>Obtain slope, intercept and 95 % CI thereof<br>**1B perform Bland-Altman difference analysis**<br>Obtain mean relative difference and 95 % CI thereof |
| **2) Consensus-based EQA value assignment** |
| **2A Select "reference MPs" to be used for consensus-based EQA value assignment**<br>Based on: including those MPs that are already equivalent<br>MPs are considered equivalent if one of the next three criteria is met:<br>i) 95 % CI of slope obtained by Passing-Bablok regression includes 1.00<br>ii) Obtained slope obtained by Passing-Bablok regression is within the 0.95 to 1.05 range<br>iii) Mean relative difference obtained by Bland-Altman analysis is within the −5.00 % to 5.00 % range<br>**2B EQA sample value assignment**<br>Calculate for each EQA sample a pre-harmonization consensus value by plain mean of values obtained by MPs selected as "reference MPs" |
| **3) Perform regression analysis between value assigned EQA values and MP results for each MP** |
| Regression is performed by passing-bablok regression analysis<br>Obtain slope, intercept and 95 % thereof |
| **4) Determine recalibration procedure** |
| **4A Determine recalibration necessity for each MP**<br>Based on: requiring recalibration only when systematic difference with value assigned EQA samples is >5 %<br>No recalibration is required if:<br>i) 95 % CI of slope obtained by Passing-Bablok regression includes 1.00<br>ii) Obtained slope obtained by Passing-Bablok regression is within the 0.95 to 1.05 range<br>iii) Pre-harmonization Mean relative difference obtained by Bland-Altman analysis is within the -5.00 % to 5.00 % range<br>**4B Determine recalibration procedures for each MP when considered necessary**<br>Recalibration procedure is based on obtained slope and/or intercept obtained by passing-bablok regression in step 3 |
| **5) Mathematically recalibrate MPs** |
| Calculate new EQA values for each recalibrated MP |

MP, measurement procedure.

require a recalibration. A secondary analysis was applied for CEA for which the results were marked as CEA*. Here, also the MPs with a mean relative difference within the −5–5 % range, were excluded to require recalibration. For all remaining MPs, a recalibration procedure based on the obtained slope and/or intercept was proposed as harmonization procedure. Fifth, calculate the recalibrated EQA values using the obtained recalibration procedure.

### Validation of the performance of the obtained TM harmonization procedures

Both the pre- and post-tumor marker (TM) harmonization status, was determined by using the consensus mean of all MP as the reference value and calculating the mean relative MP difference from the consensus mean by Bland-Altman plot difference analysis. For the post-harmonization analysis; for those MPs that did not require recalibration the original EQA values were used. Those MPs that did required recalibration, the EQA values after (*in silico*) recalibration were used.

### Statistical analysis

Data management and statistical analysis were performed using Analyse-it for Microsoft Excel software (version 6.15). The Level of Agreement/mean relative systematic differences were calculated using Bland-Altman analysis based on relative differences. Residual differences between the EQA sample target value and the harmonized measurement value observed with each EQA sample was calculated as measure of the residual scatter. Therefore, the CV (%) of this scatter was calculated by dividing the 95 % CI obtained by Bland-Altman analysis, by 4 to roughly reflect 1 SD.

## Results

### MP method comparison

The first step of the harmonization protocol was to perform MP comparisons for each combination of MPs. The regression analysis performed for each pair of MPs for CA15–3 is presented in Figure 1 and for CEA in Figure 2. The slope and intercept obtained from the Passing-Bablok regression, as well as the Spearman's correlation coefficient obtained, including the 95 % CIs, are presented in Table 2. The correlation coefficients were all ≥0.96.

### Selection of comparable MPs, EQA value assignment, and MP comparison with value assigned EQA samples

As a second step of the harmonization procedure, those MPs to value assign EQA samples, were selected. For CA15–3, based on both the regression analysis and the mean

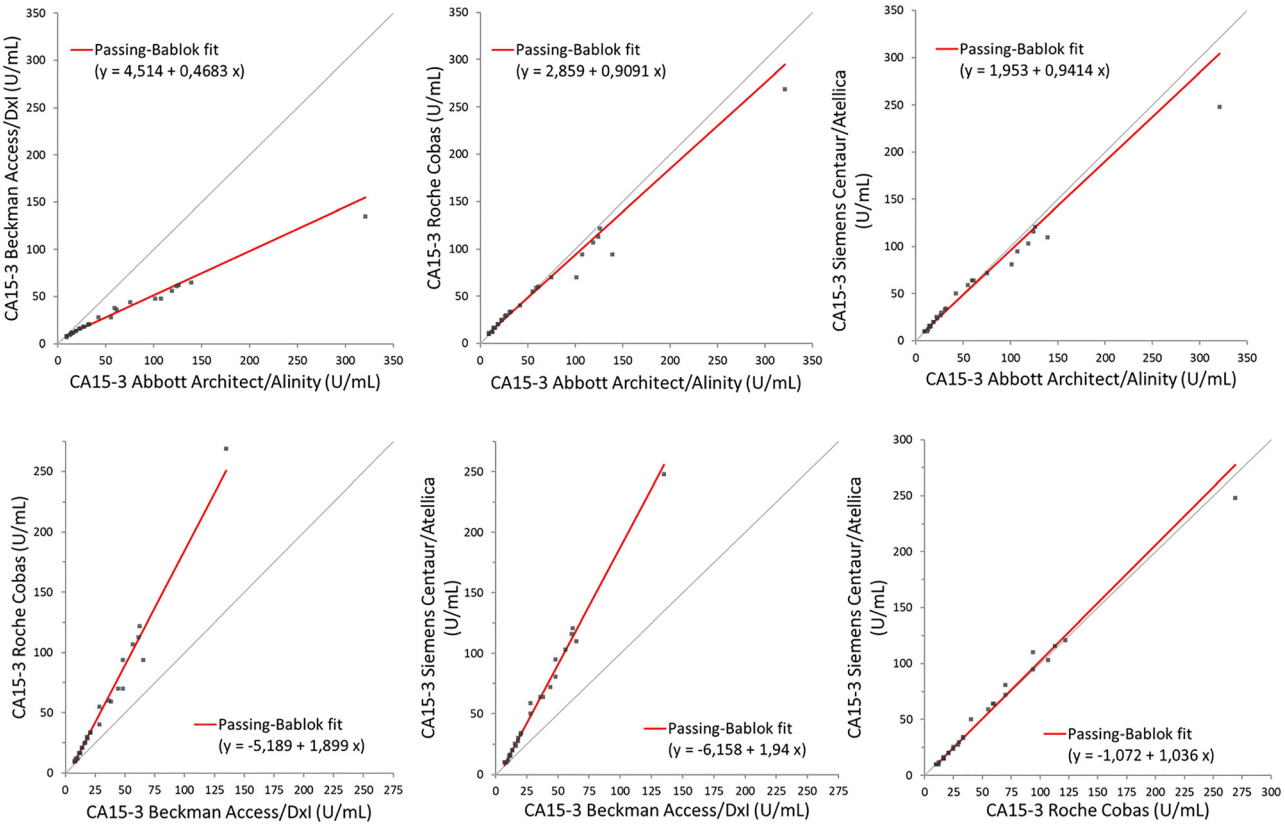


**Figure 1:** Pre-harmonization between MP method comparison for CA 15-3. Pre-harmonization method comparison study by Passing-Bablok regression analysis. Grey line represents the identifier and red line the obtained regression line.

difference, the Abbott, Roche, and Siemens MPs were considered comarable, and the consensus mean of these three MP was used to value assign the EQA samples, see Table 3. For CEA, the Abbott, Beckman and Roche MP were considered comparable and the consensus mean of these three MP was used to value assign the EQA samples. As a third step of the harmonization protocol, regression analyses as well as Bland-Altman analysis for agreement were performed between the value assigned EQA samples and each MP. The slopes, intercepts and mean relative differences obtained are shown in Table 3.

## Determining *in silico* recalibration procedures

The MPs that required recalibration based on the prespecified criteria were Beckman for CA15–3, and Beckman, Roche and Siemens for CEA. For the alternative CEA*, recalibration of only the Siemens MP was required. The mathematical recalibration procedures were based on the slope and intercept obtained by regression of the MP with the value-assigned EQA samples (step 3). The recalibration procedures obtained are shown in Table 3.

## Validation of obtained harmonization protocol

Both pre- and post-harmonization status were investigated by using the consensus mean of all MPs as a reference value and determining the level of agreement using Bland-Altman plots. The obtained mean MP difference and 95 % level of agreement (LoA) are shown in Table 3. The post-harmonization method comparison and regression analysis for both CA 15-3 and CEA is presented in Figure 3. Post-harmonization Bland-Altman plots are shown in Figure 4 for CA 15-3 and in Figure 5 for CEA in which also the pre-harmonization EQA values of those MPs that required recalibration, are presented.

For CA15–3, the estimated pre-harmonization mean MP difference ranged from –29.28 to 9.86 %, and the estimated post-harmonization mean MP difference ranged from –0.09 to 0.12 %, with the highest 95 % CI ranging from –17.41 to

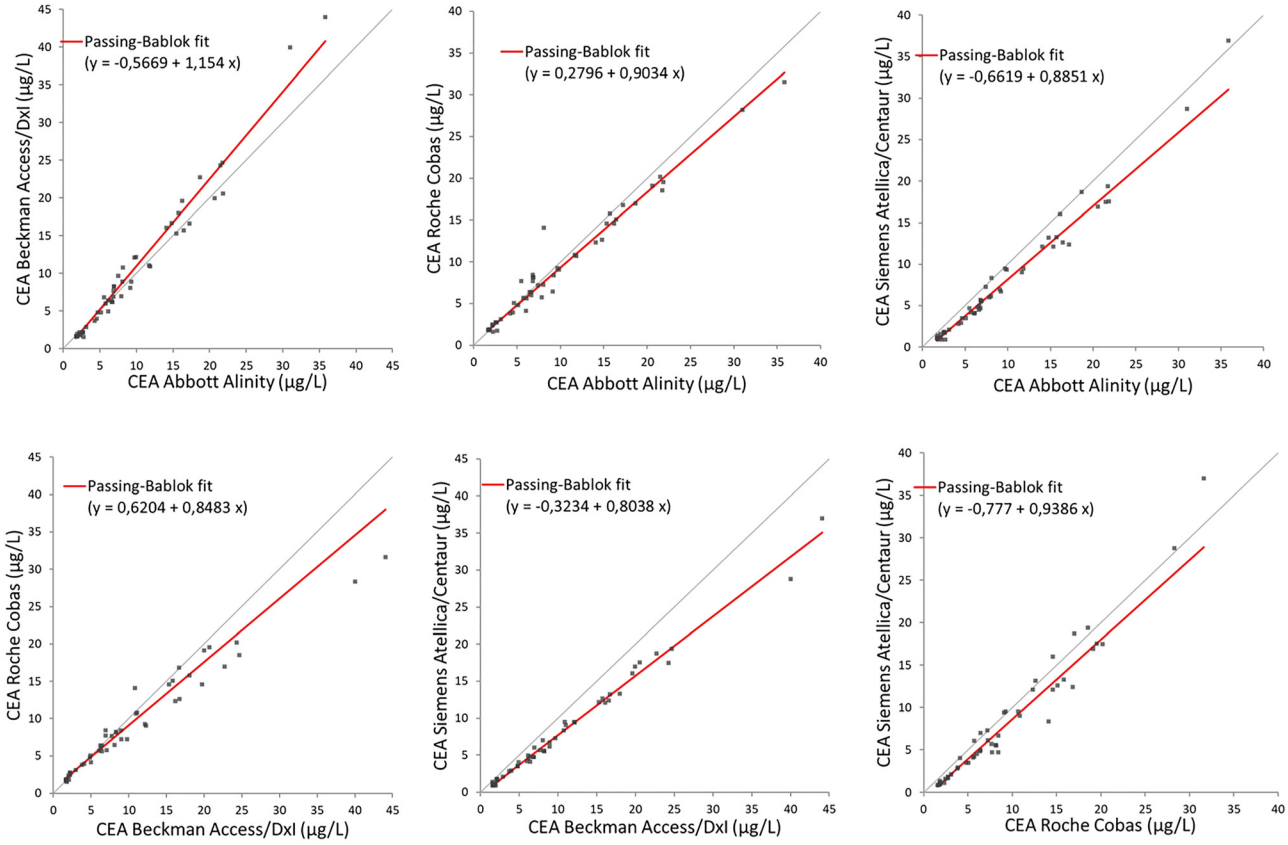


**Figure 2:** Pre-harmonization between MP method comparison for CEA. Pre-harmonization method comparison study by Passing-Bablok regression analysis. Grey line represents the identifier and red line the obtained regression line.

17.22 %. For CEA the estimated pre-harmonization mean MP difference ranged from –23.78 to 2.00 %. Using the procedure that required recalibration of Beckman, Roche and Siemens MP a post-harmonization mean MP difference was observed that ranged from –8.44 to 6.19 %. Using the alternative CEA* harmonization procedure, which required recalibration of only the Siemens MP, a post-harmonization mean MP difference was observed that ranged from –3.13 to 1.42 %. In addition, the coefficient of variation (CV) as a measure or post-harmonization residual scatter and harmonization inaccuracy of individual samples are also presented in Table 3.

## Discussion

A study was conducted to investigate the achievable harmonization by MP recalibration for CA 15-3 and CEA. The procedure used patient-pool based EQA samples, a prespecified step-by-step harmonization protocol and was performed *in silico*. The results obtained suggest that recalibration of selected MPs can significantly improve the overall harmonization status of both CA 15-3 and CEA.

Adequate harmonization of MPs for the same measurand is essential to enable the use of common clinical decision limits and to allow interchangeability of test results. Often, commutable IS and reference measurement procedures are available for measurands to achieve and maintain adequate MP harmonization. However, for most tumor markers, these are not available. For the tumor markers CEA and CA 15-3, a previous study has demonstrated that the harmonization status was outside the total allowable bias criterion based on biological variation [21]. As a follow-up, this study aimed to investigate the potential achievable harmonization that recalibration of the MPs would allow. Therefore, a protocol was performed, based on the available EQA data obtained from programs that used residual patient materials, to obtain elevated tumor marker concentrations. It should be noted that the commutability, a critical and essential requirement for this purpose, has not been demonstrated and was assumed [21]. In addition, a practical requirement was added to minimize the number of MP recalibrations required to allow the highest adoption by IVD companies, clinical laboratories, and end users. Furthermore, this would allow direct traceability to clinical studies that have clinically

**Table 2:** Pre-harmonization MP comparison. For both CA15-3 and CEA the results of the method comparison for each combination of MP are presented.

| | CA15-3 | | CEA | |
|---|---|---|---|---|
| | **Estimate** | **95 % CI** | **Estimate** | **95 % CI** |
| **Abbot vs. Beckman** | | | | |
| Slope | 0.47 | 0.43 to 0.55 | 1.15 | 1.02 to 1.24 |
| Intercept | 4.51 | 3.06 to 5.77 | −0.57 | −1.02 to 0.24 |
| Correlation | 0.99 | 0.99 to 1.00 | 0.98 | 0.97 to 0.99 |
| Bland-Altman mean difference, % | −43.12 | −50.96 to 35.29 | 1.01 | −3.40 to 5.42 |
| **Abbot vs. Roche** | | | | |
| Slope | 0.91 | 0.85 to 0.96 | 0.90 | 0.88 to 0.93 |
| Intercept | 2.86 | 1.76 to 4.78 | 0.28 | 0.10 to 0.46 |
| Correlation | 0.99 | 0.98 to 1.00 | 0.97 | 0.95 to 0.98 |
| Bland-Altman mean difference, % | 0.29 | −5.05 to 5.63 | −3.64 | −8.33 to 1.04 |
| **Abbot vs. Siemens** | | | | |
| Slope | 0.94 | 0.84 to 1.07 | 0.89 | 0.84 to 0.96 |
| Intercept | 1.95 | −0.19 to 4.65 | −0.66 | −1.14 to 0.46 |
| Correlation | 0.99 | 0.99 to 1.00 | 0.98 | 0.97 to 0.99 |
| Bland-Altman mean difference, % | 0.45 | −4.09 to 5.00 | −28.42 | −34.02 to 22.82 |
| **Beckman vs. Roche** | | | | |
| Slope | 1.90 | 1.65 to 2.04 | 0.85 | 0.73 to 0.93 |
| Intercept | −5.19 | −7.31 to 2.25 | 0.62 | 0.33 to 1.03 |
| Correlation | 0.99 | 0.99 to 1.00 | 0.98 | 0.96 to 0.99 |
| Bland-Altman mean difference, % | 43.69 | 38.75 to 48.63 | −4.62 | −9.36 to 0.11 |
| **Beckman vs. Siemens** | | | | |
| Slope | 1.94 | 1.81 to 2.06 | 0.80 | 0.77 to 0.84 |
| Intercept | −6.16 | −8.07 to 4.48 | −0.32 | −0.51 to 0.11 |
| Correlation | 0.99 | 0.99 to 1.00 | 0.99 | 0.98 to 0.99 |
| Bland-Altman mean difference, % | 43.79 | 37.83 to 49.74 | −29.67 | −33.10 to 26.23 |
| **Roche vs. Siemens** | | | | |
| Slope | 1.04 | 1.00 to 1.12 | 0.94 | 0.88 to 1.07 |
| Intercept | −1.07 | −2.99 to 0.00 | −0.78 | −1.26 to 0.62 |
| Correlation | 0.99 | 0.99 to 1.00 | 0.98 | 0.96 to 0.99 |
| Bland-Altman mean difference, % | 0.18 | −2.96 to 3.28 | −24.92 | −31.09 to −18.75 |

validated these tumor markers in the past. In particular, steps 1, 2 and 4 of the harmonization protocol were critical to meeting this practical preference. Step 1, the comparison between individual MPs, and step 2, the selection of those MPs considered equivalent, were used to identify MPs with significantly different characteristics from the "the majority" of MPs. The EQA value assignment was then based, where possible, on these already comparable MPs. Particularly in the case of a single or limited number of MPs with significantly different tumor marker results, this approach would avoid unnecessary recalibrations of all MPs to the overall consensus mean. Also, the criterion used in step 4 to determine MPs requiring recalibration, not exceeding a 5 % bias in either the slope of the obtained regression line, or the mean difference, limits the number of unnecessary recalibrations. Although the 5 % criterion used was more pragmatic, it could alternatively be related to tumor marker-specific performance criteria and related to the exact accuracy of estimating any recalibration procedure. Others, particularly in the context of the thyroid stimulating hormone (TSH) harmonization initiative, have provided a statistical basis for harmonization protocols without reference measurement procedures [29]. Challenges with the proposed calculation procedure include, in particular, the outlier exclusion procedure and the complexity of the analysis. The statistical approach presented in this study, using non-parametric regression analysis, is rather insensitive to outliers and the preferred method for skewed distributions. This approach was chosen since particularly for tumor markers, which are known for their heterogeneous clinical occurrence, in combination with differences in MP assay design, outliers were expected. The statistical approach used here therefore did not require outlier removal or transformation to more normally distributed populations. As part of the TSH harmonization, the procedure for defining the consensus means for value assignment was presented [30], but here the criteria for MP selection were different, namely to reflect MPs that are symmetrically located from the

**Table 3:** Obtained CA15-3 and CEA harmonization procedures.

| | CA15-3 | CEA | CEA[a] |
|---|---|---|---|
| **Current (pre-) harmonization status** | | | |
| LoA with pre-harmonization consensus mean (95 %CI) | | | |
| Abbott | 9.75 % (−12.50 to 32.00 %) | 0.96 % (−17.78 to 19.70 %) | Same as CEA |
| Beckman | −29.28 % (−47.81 to −10.76 %) | 2.00 % (−17.22 to 21.22 %) | Same as CEA |
| Roche | 9.67 % (−1.85 to 21.19 %) | −2.59 % (−22.17 to 16.99 %) | Same as CEA |
| Siemens | 9.86 % (−1.86 to 21.58 %) | −23.78 % (−47.29 % to −0.28 %) | Same as CEA |
| **Step 1: Determine systematic bias between each pair of MPs** | See Table 2 and Figure 1 | See Table 2 and Figure 2 | Same as CEA |
| **Step 2: MPs selected for consensus based EQA value assignment** | Abbott, Roche, Siemens | Abbott, Beckman, Roche | Same as CEA |
| **Step 3: Regression analysis of value assigned EQA samples and each MP** | | | |
| Slope (95 % CI) | | | |
| Abbott | 1.040 (−0.9869 to 1.112) | 0.9882 (0.9553 to 1.018) | Same as CEA |
| Beckman | 0.5261 (0.4733 to 0.5456) | 1.109 (1.031 to 1.168) | Same as CEA |
| Roche | 0.9557 (0.9367 to 0.9797) | 0.9142 (0.8557 to 0.9461) | Same as CEA |
| Siemens | 0.9821 (0.9545 to 1.043) | 0.8812 (0.8353 to 0.9266) | Same as CEA |
| Intercept (95 % CI) | | | |
| Abbott | −1.810 (−3.342 to −0.6502) | 0.09339 (−0.04141 to 0.2510) | Same as CEA |
| Beckman | 3.070 (2.663 to 4.234) | −0.4862 (−0.7630 to −0.2116) | Same as CEA |
| Roche | 1.439 (0.7378 to 2.175) | 0.2959 (0.1678 to 0.5145) | Same as CEA |
| Siemens | 0.4310 (−0.6749 to 1.477) | −0.6709 (−0.8717 to −0.4496) | Same as CEA |
| Mean relative difference (95 % CI), % | | | |
| Abbott | −0.46 (−3.53 to 2.61) | 0.40 (−2.19 to 3.00) | Same as CEA |
| Beckman | −43.79 (−49.70 to −37.88) | 1.40 (−1.22 to 4.03) | Same as CEA |
| Roche | −0.19 (−2.77 to 2.39) | −3.24 (−6.06 to −0.43) | Same as CEA |
| Siemens | −0.01 (−1.93 to 1.90) | −28.19 (−32.73 to 23.66) | Same as CEA |
| **Step 4: Defined recalibration procedure** | | | |
| Abbott | N.A. | N.A. | N.A. |
| Beckman | CA15-3 recal=[CA15-3 - 3.07]/0.526 | CEA recal=CEA/1.11 | N.A. |
| Roche | N.A. | CEA recal=CEA/0.914 | N.A. |
| Siemens | N.A. | CEA recal=[CEA + 0.671]/0.8812 | CEA recal=[CEA + 0.671]/0.8812 |
| **Post-harmonization protocol harmonization status** | | | |
| LoA with post-harmonization consensus mean (95 %CI) | | | |
| Abbott | −0.09 % (−17.41 to 17.22 %) | 0.59 % (−18.66 to 19.83 %) | 0.35 % (−18.85 to 19.55 %) |
| Beckman | −0.08 % (−11.96 to 11.80 %) | −8.44 % (−26.07 to 9.19 %) | 1.36 % (−17.48 to 20.20 %) |
| Roche | 0.05 % (−13.31 to 13.42 %) | 6.19 % (−15.80 to 28.18 %) | −3.13 % (−23.87 to 17.60 %) |
| Siemens | 0.12 % (−9.82 to 10.05 %) | 1.66 % (−12.67 to 15.99 %) | 1.42 % (−12.78 to 15.63 %) |
| Post-harmonization leftover CV, % | | | |
| Abbott | 8.7 % | 9.6 % | 9.6 % |
| Beckman | 5.9 % | 8.8 % | 9.4 % |
| Roche | 6.9 % | 11.0 % | 10.4 % |
| Siemens | 5.0 % | 7.2 % | 7.1 % |

Step 1: Consensus value was calculated using pre-harmonization MP values of all MPs (Abbott, Beckman, Roche and Siemens). Step 2: Consensus value was based on MP selected at Step 2. Step 3: Consensus value was calculated using post-harmonization MP values of all MPs (Abbott, Beckman, Roche and Siemens).

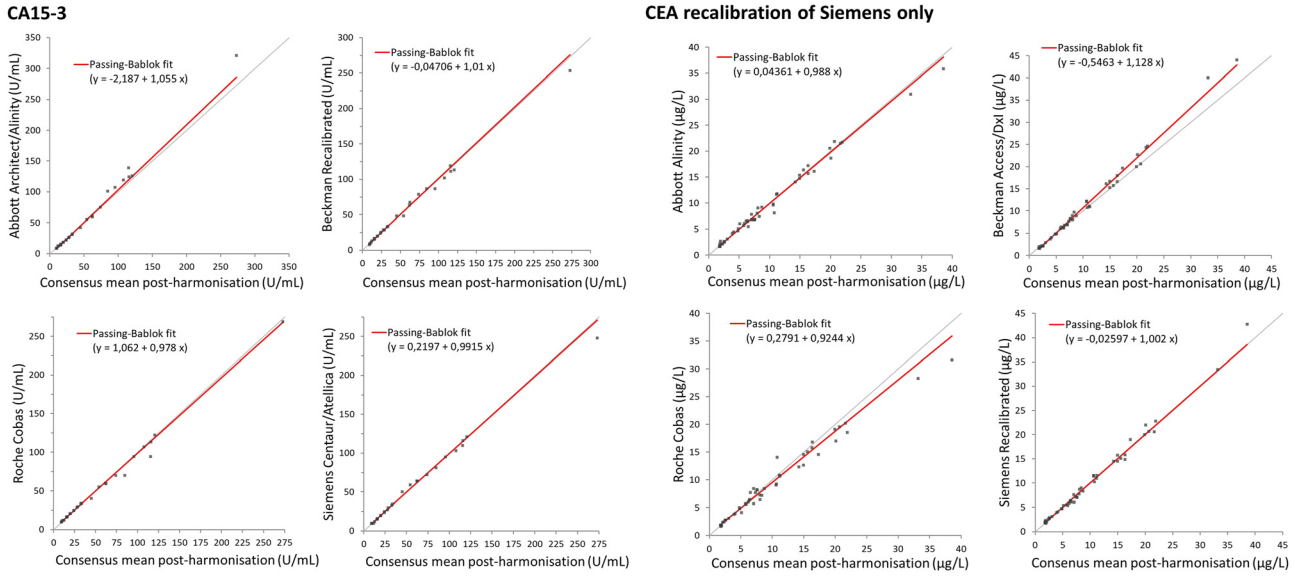


**Figure 3:** Post-harmonization MP comparison with consensus mean. Post harmonization regression analysis of consensus mean of all MP with individual MPs. Gray line represents the identifier and red line the obtained Passing-Bablok regression line.

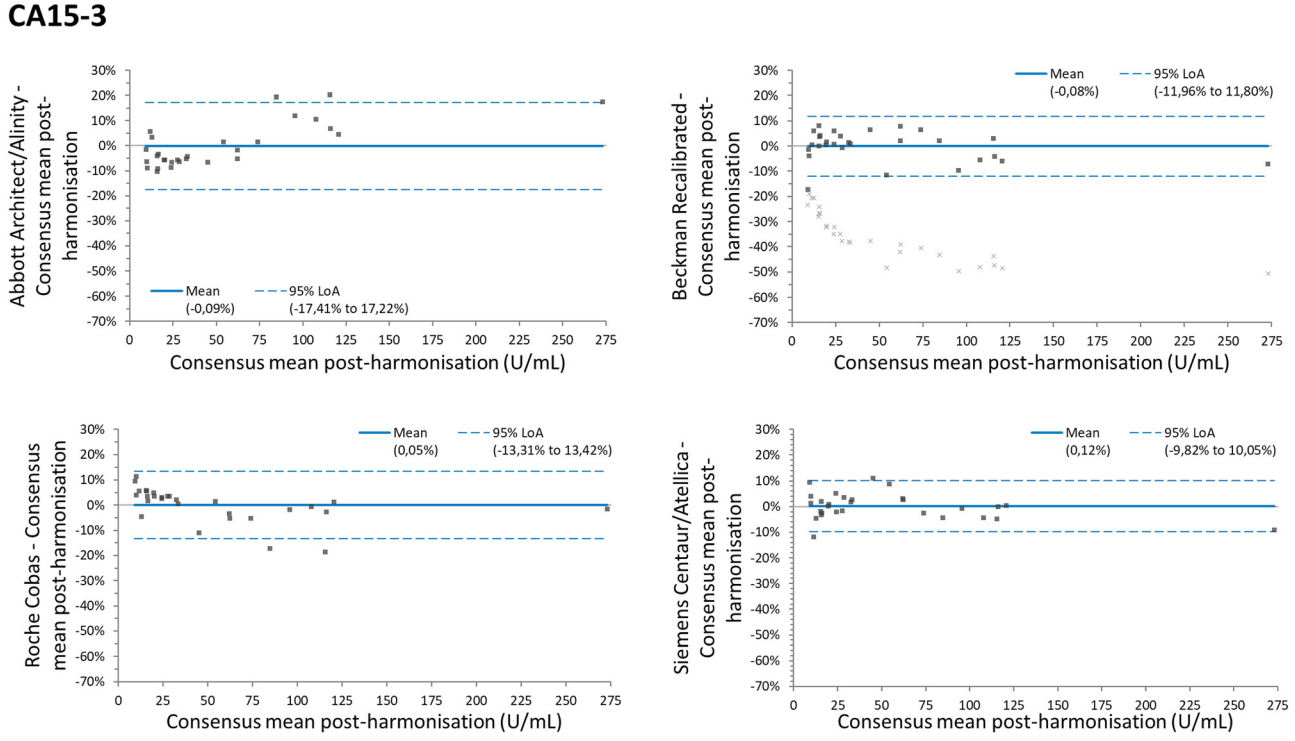


**Figure 4:** Post-harmonization Bland-Altman plots for CA 15-3. Post-harmonization relative differences between consensus mean of all MP and individual MP are presented. Dark gray squares represent post-harmonization EQA values. Light gray crosses represent pre-harmonization EQA values of recalibrated MPs.

consensus mean of all MPs and having low scatter. The selection criteria used in our study were based only on selecting MPs that, on average, were already considered to provide similar tumor marker values and thus to be equivalent. Overall, when the protocol was applied the mean bias per MP range was reduced from –29.28–9.86 % to –0.09–0.12 % for CA 15-3 and from –23.78 – 2.00 % to –3.13–1.42 % for CEA. The post-harmonization mean differences with the overall consensus mean were all within the optimal allowable bias criterion of 4.6 % for CA 15-3 and 7.5 % for CEA,

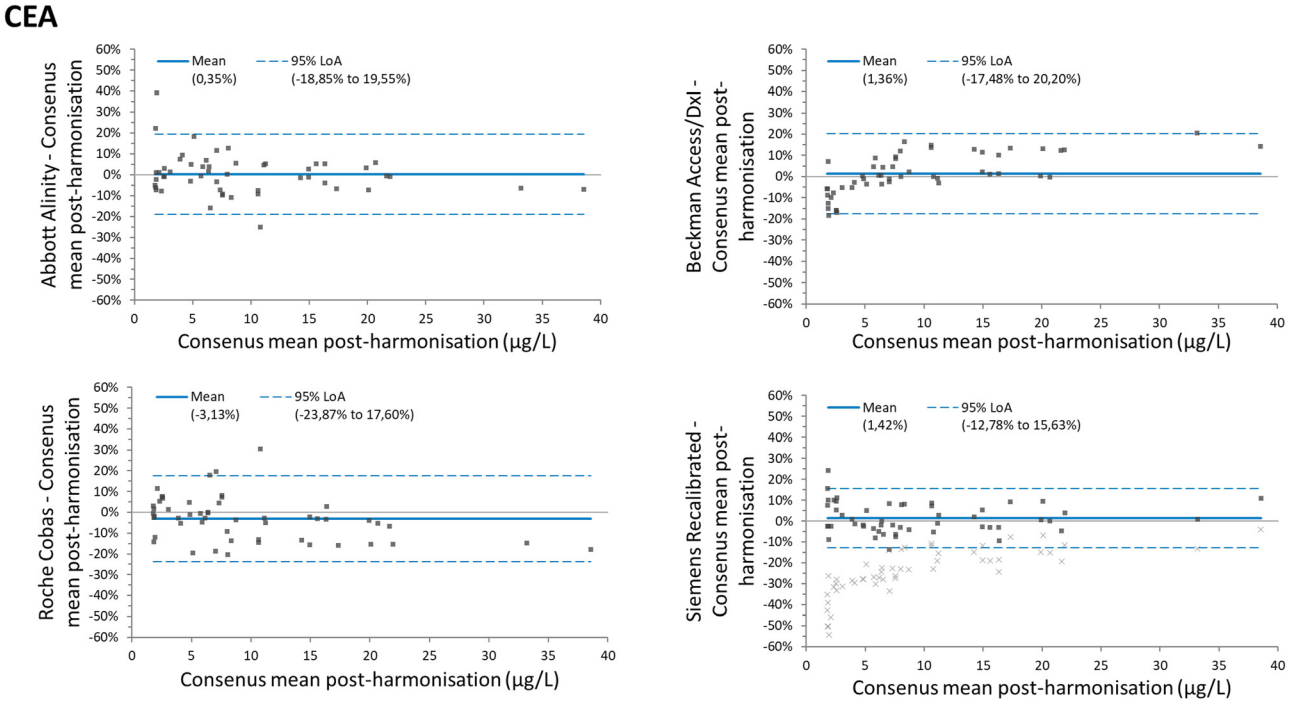


**Figure 5:** Post-harmonization Bland-Altman plots for CEA. Post-harmonization relative differences between consensus mean of all MP and individual MP are presented. Dark gray squares represent post-harmonization EQA values. Light gray crosses represent pre-harmonization EQA values of recalibrated MPs.

based on biological variation [28, 31]. The residual scatter of the individual samples was determined by calculating the coefficient of variation, which ranged from 5.0 to 8.7 % for CA 15-3 and 7.1–10.4 % for CEA. This metric most likely reflects a combination of MP imprecision, number of laboratory results per MP included, potentially a lack of commutability and true clinical heterogeneity of the TM. Since the EQA values are generally based on a large number of individual laboratory observations ranging from 6-100 participants per MP, this may largely reflect true clinical tumor marker heterogeneity, although for the Abbott Alinity MPs only two participants were included in the 2021 SKML EQA program.

Several limitations of the study need to be addressed. First, as previously discussed, the commutability of the EQA samples has not been demonstrated [21]. Secondly, the sample size, might compromise an accurate estimation of the obtained slope and intercept of the regression line with the value assigned EQA samples. Due to the residual scattering around the regression line, an increased number of samples would increase the reliability of the obtained mathematical recalibration. Furthermore, an external validation by a new set of samples (either individual patient samples or commutable EQA) should be used to confirm the presented results and the harmonization achievable by recalibration of CEA and CA 15-3 MPs. Also, the present investigation included only MPs available in the SKML, UK NEQAS and KEQAS EQA programs, while several other MPs for CEA and CA 15-3 are available worldwide and used in clinical care. Finally, the results obtained in this proof of principle study do not mean that the MPs identified and selected for *in silico* recalibration are “wrong” as no “correct” answer is possible due to the absence of a clear definition of the measurand, the absence of a reference MP and accepted reference materials. Despite these limitations, overall the presented methodology may also provide a relevant approach to investigate the possibility of harmonizing a larger number of (oncological) biomarkers.

## CA 15-3 harmonization

The results of the CA 15-3 method comparison study are somewhat comparable to a method comparison study performed by Slev et al. using 7 automated immunoassay platforms [32]. They also found that the CA 15-3 results of the Beckman MP were about half those of the Centaur MP (now Siemens) and that the Centaur with the Abbott and Roche MPs had a slope obtained by Passing-Bablok regression not significantly different from 1. They also observed that another MP not included in this study, the VITROS ECi, had results about 50 % higher than the Centaur MP and that the IMMULITE MP results were also comparable to the Centaur

MP with a slope not significantly different from 1 [32]. The correlation coefficients from the Slev et al. study, ranged from 0.90 to 0.96 which is slightly lower than that observed in this study. Since the method comparison by Slev et al. was presumably (not otherwise specified) based on singular measurements, it is probably also likely to be affected by the MP imprecisions [32]. For CA 15-3 there is no clear definition of the measurand and (therefore) no IS and reference measurement procedure available [33]. As the results obtained in our study are highly comparable to the study performed by Slev et al. [32], this strengthens the assumption that the EQA materials used, truly reflect the averaged individual patient sample behavior and can be used for harmonization procedures. For true CA 15-3 harmonization, an ISO 21151-compliant protocol must be used to demonstrate true commutability of the EQA materials used [34, 35]. Furthermore, an approach in which MPs that are considered equivalent e.g. MPs from Abbott, Roche, Siemens and perhaps others, are used to value assign EQA, and only those MPs that are not considered equivalent are recalibrated, seems to be able to provide a significant increase in the harmonization of CA 15-3 MPs. In addition, the EQA samples may also be useful for future monitoring of the CA 15-3 harmonization and adaptation to an initiated harmonization by other or new CA 15-3 MPs. The availability of other, high-volume commutable standards or reference materials, would however provide a more practical and sustainable harmonization and should therefore preferably be included in any harmonization study. Some residual scatter and harmonization uncertainty is expected for individual patient samples as shown by the post-harmonization CV in Table 3.

## CEA harmonization

For CEA the presented MP comparisons presented in Table 1 appear to be comparable to method comparison studies performed by others, in particular that Siemens/Centaur MP was found to produce the lowest CEA values [18, 19, 36] (see Supplementary Table 1). A reference preparation is available for CEA; IRP WHO 73/601, but of the MPs included in this study only Roche uses the IRP for CEA calibration. In addition, there is no internationally accepted reference MP for CEA [36]. A recent study investigating the commutability of the IRP diluted in different dilution matrices for 5 MPs (Abbott, Beckman, Diasorin, Maccura and Roche) found that it was commutable for 7 out of 10 MP combinations. Recalibration of the MPs incorporating a bias correction for incommutability resulted in commutability of all pairs of MPs. When used for recalibration, the mean bias range was reduced from −7.31 to 10.09 %, to −6.49 to 3.06 % and particularly affected the Beckman MP bias [36]. Alternatively, a set of serum pools as candidate reference materials were also tested for commutability and found to be commutable for all pairs of MPs [36]. Unfortunately, this important study by Zhang R. et al. did not include more MPs, especially the Siemens MP, which had the largest bias in our and other studies. However, they showed that CEA harmonization can be improved by using an IRP. Currently, the development of a new WHO reference preparation for CEA has been initiated and a similar study as performed by Zhang R. et al. using this new standard and including more, if not all, clinically available MPs seems warranted. The results presented in our study also indicate the potential for improvement of the CEA harmonization status, although some residual scatter will remain, most likely reflecting CEA MP heterogeneity, as presented by CV in Table 3.

In conclusion, the presented in silico EQA-based harmonization set-up using a pragmatic, not overall consensus based value assignment procedure, might allow for the investigation of achievable harmonization of other biomarkers. The presented data indicate that by recalibrating individual MPs for CA 15-3 and CEA, a significant improvement in global harmonization of these tumor markers can be achieved, potentially meeting the optimal allowable bias criterion based on biological variation.

**Research ethics:** Not applicable.
**Informed consent:** Not applicable.
**Author contributions:** The author have accepted responsibility fort he entire content of this manuscript and approved ist submission.
**Competing interests:** H. van Rossum support from Roche diagnostics, Abbott, Beckman, Siemens, Sysmex, Huvaros; employment or leadership roles: Huvaros, Director Huvaros BV., Chief Scientific Officer SelfSareSure Blood Collections BV, Honoraria: Huvaros BV; expert testimony: Abbott, Siemens; patents: PCT/NL2016/050315 Method for setting measuring equipment, computer program and measuring equipment, NL2019/ 079641 Blood collection device and method for the self-collection of blood by a user. S. Holdenrieder, employment or leadership roles: Board ISOBM, WHO-IARC Steering Committee, Associate Editor for several oncological and lab journals, SFZ BioCoDE, CEBIO; consultant or advisory roles: Instand, Roche, Thermo, Merck. L. Ramanathan, support from ADLM (formerly AACC) for travel to attend 2022Annual Scientific Meeting,support from Abbott Diagnostics, 2021–2023 for validating infectious disease markers on the i2000, support from Nova Biomedical, 2023, to study ionized Mg in critically ill patients.
**Research funding:** H. van Rossum: Roche, Health Holland. S Holdenrieder: EU, Roche, Volition, Sysmex. L. Ramanathan: Abbott, Nova Biomedical. The funding organizations played

no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

**Data availability:** The raw data is not available.

**Disclaimer:** The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry. Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Public Health Service, and the US Department of Health and Human Services.

## References


1. Gion M, Trevisiol C, Rutjes AW, Rainato G, Fabricio AS. Circulating tumor markers: a guide to their appropriate clinical use Comparative summary of recommendations from clinical practice guidelines (PART 1). Int J Biol Markers 2016;31:e332–67.
2. Gion M, Trevisiol C, Rutjes AWS, Rainato G, Fabricio ASC. Circulating tumor markers: a guide to their appropriate clinical use Comparative summary of recommendations from clinical practice guidelines (PART 2). Int J Biol Markers 2017;31:e1–52.
3. Gion M, Trevisiol C, Rutjes AWS, Rainato G, Fabricio ASC. Circulating tumor markers: a guide to their appropriate clinical use Comparative summary of recommendations from clinical practice guidelines (PART 3). Int J Biol Markers 2017;32:e147–81.
4. Duffy MJ. CA 15-3 and related mucins as circulating markers in breast cancer. Ann Clin Biochem 1999;36:579–86.
5. Duffy MJ, Evoy D, McDermott EW. CA 15-3: uses and limitation as a biomarker for breast cancer. Clin Chim Acta 2010;411:1869–74.
6. Hilkens J, Buijs F, Hilgers J, Hageman P, Calafat J, Sonnenberg A, et al. Monoclonal antibodies against human milk-fat globule membranes detecting differentiation antigens of the mammary gland and its tumors. Int J Cancer 1984;34:197–206.
7. Kufe D, Inghirami G, Abe M, Hayes D, Justi-Wheeler H, Schlom J. Differential reactivity of a novel monoclonal antibody (DF3) with human malignant versus benign breast tumors. Hybridoma 1984;3:223–32.
8. Price MR, Rye PD, Petrakou E, Murray A, Brady K, Imai S, et al. Summary report on the ISOBM TD-4 workshop: analysis of 56 monoclonal antibodies against the MUC1 mucin. San Diego, Calif., November 17-23, 1996. Tumour Biol 1998;19:1–20.
9. Gaughran G, Aggarwal N, Shadbolt B, Stuart-Harris R. The utility of the tumor markers CA15.3, CEA, CA-125 and CA19.9 in metastatic breast cancer. Breast Cancer Manag 2020;9:BMT50.
10. Yerushalmi R, Tyldesley S, Kennecke H, Speers C, Woods R, Knight B, et al. Tumor markers in metastatic breast cancer subtypes: frequency of elevation and correlation with outcome. Ann Oncol 2012;23: 338–45.
11. Trapé J, Filella X, Alsina-Donadeu M, Juan-Pereira L, Bosch-Ferrer Á, Rigo-Bonnin R. Increased plasma concentrations of tumour markers in the absence of neoplasia. Clin Chem Lab Med 2011;49:1605–20.
12. Gennari A, André F, Barrios CH, Cortés J, de Azambuja E, DeMichele A, et al. ESMO Clinical Practice Guideline for the diagnosis, staging and treatment of patients with metastatic breast cancer. Ann Oncol 2021; 32:1475–95.
13. Henry NL, Somerfield MR, Dayao Z, Elias A, Kalinsky K, McShane LM, et al. Biomarkers for systemic therapy in metastatic breast cancer: ASCO guideline update. J Clin Oncol 2022;40:3205–21.
14. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 2007;25:5287–312.
15. Gold P, Freedman SO. Demonstration of tumor-specific antigens in human colonic carcinomata by immunological tolerance and absorption techniques. J Exp Med 1965;121:439–62.
16. Gold P, Freedman SO. Specific carcinoembryonic antigens of the human digestive system. J Exp Med 1965;122:467–81.
17. Vrba R, Alpert E, Isselbacher KJ. Immunological heterogeneity of serum carcinoembryonic antigen (CEA). Immunochemistry 1976;13:87–9.
18. Park J, Lee S, Kim Y, Choi A, Lee H, Lim J, et al. Comparison of four automated carcinoembryonic antigen immunoassays: ADVIA Centaur XP, ARCHITECT I2000sr, elecsys E170, and unicel Dxi800. Ann Lab Med 2018;38:355–61.
19. Zhang K, Huo H, Lin G, Yue Y, Wang Q, Li J. A long way to go for the harmonization of four immunoassays for carcinoembryonic antigen. Clin Chim Acta 2016;454:15–9.
20. Laurence DJ, Turberville C, Anderson SG, Neville AM. First British standard for carcinoembryonic antigen (CEA). Br J Cancer 1975;32:295–9.
21. van Rossum HH, Holdenrieder S, Ballieux BEPB, Badrick T, Yun Y-M, Zhang C, et al. Investigating the current harmonisation status of tumor markers using global external quality assessment programs: a feasibility study. Clin Chem 2024;70:669–79.
22. Cervantes A, Adam R, Roselló S, Arnold D, Normanno N, Taïeb J, et al. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol 2023;34:10–32.
23. Argilés G, Tabernero J, Labianca R, Hochhauser D, Salazar R, Iveson T, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2020; 31:1291–305.
24. Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. J Clin Oncol 2006;24:5313–27.
25. Muller M, Hoogendoorn R, Moritz RJG, Van der Noort V, Lanfermeijer M, Korse CM, et al. Validation of a clinical blood-based decision aid to guide immunotherapy treatment in patients with non-small cell lung cancer. Tumour Biol 2021;43:115–27.
26. van den Heuvel M, Holdenrieder S, Schuurbiers M, Cigoianu D, Trulson I, van Rossum H, et al. Serum tumor markers for response prediction and monitoring of advanced lung cancer: a review focusing on immunotherapy and targeted therapies. Tumour Biol 2023;46: S233–68. https://doi.org/10.3233/TUB-220039.
27. Jing JX, Wang Y, Xu XQ, Sun T, Tian BG, Du LL, et al. Tumor markers for diagnosis, monitoring of recurrence and prognosis in patients with upper gastrointestinal tract cancer. Asian Pac J Cancer Prev 2014;15: 10267–72.
28. The EFLM biological variation database. [cited 2022. Available from: https://biologicalvariation.eu.
29. Stöckl D, Van Uytfanghe K, Van Aelst S, Thienpont LM. A statistical basis for harmonization of thyroid stimulating hormone immunoassays using a robust factor analysis model. Clin Chem Lab Med 2014;52:965–72.
30. Thienpont LM, Van Uytfanghe K, De Grande LAC, Reynders D, Das B, Faix JD, et al. Harmonization of serum thyroid-stimulating hormone measurements paves the way for the adoption of a more uniform reference interval. Clin Chem 2017;63:1248–60.

31. Coşkun A, Aarsand AK, Sandberg S, Guerra E, Locatelli M, Díaz-Garzón J, et al. Within- and between-subject biological variation data for tumor markers based on the European Biological Variation Study. Clin Chem Lab Med 2022;60:543–52.
32. Slev PR, Rawlins ML, Roberts WL. Performance characteristics of seven automated CA 15-3 assays. Am J Clin Pathol 2006;125:752–7.
33. Van Dalen A. Analytical requirements and standardization of CA 15-3. Scand J Clin Lab Invest Suppl 1995;221:102–4.
34. ISO 17511. *In vitro* diagnostic medical devices - requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples. Geneva, Switzerland: International Organization for Standardization; 2020.
35. ISO 21151. *In vitro* diagnostic medical devices - requirements for international harmonistion protocols establishing metrological traceability of values assigned to calibrators and human samples. Geneva, Switzerland: International Organization for Standardization; 2020.
36. Zhang R, Xu Z, Zhao R, Fu W, Song Y, Wang Q, et al. Accurate method for value assignment of carcinoembryonic antigen reference materials. J Clin Lab Anal 2023;37:e24936.

**Supplementary Material:** This article contains supplementary material (https://doi.org/10.1515/cclm-2024-0696).