Federica Braga*, Erika Frusciante, Simona Ferraro and Mauro Panteghini

Trueness evaluation and verification of inter-assay agreement of serum folate measuring systems

https://doi.org/10.1515/cclm-2019-0928
Received September 2, 2019; accepted December 12, 2019

Abstract

Background: Definitive data to establish if the use of the WHO International Standard (IS) 03/178 as a common calibrator of commercial measuring systems (MSs) has improved the harmonization of serum total folate (tFOL) measurements to a clinically suitable level are lacking. Here, we report the results of an intercomparison study aimed to verify if the current inter-assay variability is acceptable for clinical application of tFOL testing.

Methods: After confirming their commutability, the IS 03/178 and National Institute for Standards and Technology SRM 3949 L1 were used for evaluating the correctness of traceability implementation by manufacturers and the MSs trueness, respectively. The inter-assay agreement was verified using 20 patient pools. The measurement uncertainty (U) of tFOL measurements on clinical samples was also estimated. An outcome-based model for defining desirable performance specifications for bias and imprecision for serum tFOL measurements was applied.

Results: The majority of evaluated MSs overestimated the WHO IS value of +5% or more with the risk to produce an unacceptably high number of false-negative results in clinical practice. The mean inter-assay CV on all pools and on those with tFOL values >3.0 µg/L (n = 15) was 12.5% and 7.1%, respectively. In neither case the goal of 3.0% was fulfilled. The residual bias resulted in an excessive U of tFOL measurement on clinical samples.

Conclusions: The implementation of traceability of tFOL MSs to the WHO IS 03/178 is currently inadequate, resulting in an inter-assay variability that does not permit the use of a common threshold for detecting folate deficiency.

***Corresponding author: Federica Braga,** Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), Università degli Studi di Milano, Milano, Italy; and UOC Patologia Clinica, ASST Fatebenfratelli Sacco, via GB Grassi 74, 20157 Milano, Italy, Phone: +390239042743, E-mail: federica.braga@unimi.it
Erika Frusciante, Simona Ferraro and Mauro Panteghini: Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), Università di Milano, Milan, Italy

Keywords: decision-making; standardization; total folate; trueness.

Introduction

Folate plays a crucial role in the biosynthesis of nucleic acids, in the biogenesis of methyl groups and in amino acid metabolism [1]. The detection and correction of folate deficiency prevents the onset of megaloblastic anemia and reduces the risk of neural tube defects in pregnancy [2]. American/Australian populations are characterized by a low prevalence of folate deficiency (ranging from 0.1 to 1%) due to wheat flour folate fortification. However, in the European region, where no mandatory fortification programs are in place, the folate status appears to be widely variable across countries, and some populations are characterized by a suboptimal folate intake [3]. In this setting, the assessment of the individual folate status remains valuable and the measurement of serum folate, reflecting recent intakes, is the test of choice for detecting vitamin deficiency.

A comparison of automated serum folate assays performed in 2003 clearly highlighted the need for standardization efforts [4]. According to the metrological traceability theory, to become equivalent, laboratory results must be traceable to higher-order references [5]. Components of a working reference measurement system (RMS) include: (a) the unequivocal definition of the measurand as the quantity subject to measurement; (b) a reference measurement procedure (RMP), which specifically measures the analyte as defined; and (c) purified (primary) and matrixed (secondary) reference materials [6].

Serum folate cannot be defined as a single measurable entity because the proportion of individual folate species may vary in any given serum sample. The main (82%–93%) circulating folate form is the 5-methyltetrahydrofolate (5-methylTHF) [7]. The concentration of additional forms (pteroylglutamic acid [PGA], 5-formylTHF, THF, unmetabolized folic acid) depends on factors such as dietary supplementation and food fortification [3]. Technically, folate assays are not immunoassays because they rely on the competition between a labeled folate standard and unlabeled folate from biological samples for

Published online January 11, 2020

a folate-binding protein (FBP), usually a β-lactoglobulin from cow's milk. The currently available assays are built on the use of highly specific FBPs extracting from the sample 5-methylTHF and other vitamers. Therefore, the measurand detected by commercial assays in serum samples can be described as 'total folate' (tFOL), i.e. the sum of 5-methylTHF, 5-formylTHF and PGA [8].

An isotope dilution-liquid chromatography-tandem mass spectrometry (ID-LC-MS/MS) RMP calibrated with pure compound folate standards is available and listed in the Joint Committee for Traceability in Laboratory Medicine database [9]. This RMP has been used to assign tFOL target values to the WHO International Standard (IS) 03/178, established in 2005 [10]. Therefore, a complete RMS for serum tFOL assays is in place and, at present, to comply with the requirements of the European Union (EU) *In Vitro* Diagnostics Directive [11], the majority of manufacturers trace their measuring systems (MSs) to it by calibrating their internal calibrators to the WHO IS 03/178 reference material [12]. However, data are still lacking for definitively establishing if the use of the IS 03/178 as the common calibrator of commercial MSs has actually improved the between-assay harmonization, allowing the intermethod bias to decrease to a level suitable for clinical application of folate measurements [13]. External quality assessment (EQA) programs could help in understanding the current status of folate measurements, but the lack of value-assignment by RMP and non-commutability of control materials used in the schemes are a limiting issue [14]. Therefore, further investigations about the disagreement, if any, among commercially available MSs and the possible application of common decisional thresholds for their clinical use are warranted. Accordingly, we planned this study for checking the trueness of major commercial MSs for serum tFOL measurements, the status of inter-assay agreement, comparing it with established analytical performance specifications (APS) and for estimating measurement uncertainty of folate testing on clinical samples.

Materials and methods

Evaluated MSs

We included in this study four MSs: Access Dxl (Beckman Coulter), Advia Centaur (Siemens Healthcare Diagnostics), Alinity i (Abbott Diagnostics) and Cobas e801 (Roche Diagnostics), all 'Conformité Européenne'/CE marked. According to the data from EQA schemes, these manufacturers cover ~85% of folate measurements [8]. The main MS characteristics are summarized in Table 1. Regarding

Table 1: Characteristics of measuring systems for total folate measurement included in the study.

Manufacturer	Platform	Commercial name	Manufacturer's instruction for use version	Method principle	Declared limit of detection, µg/L	Declared upper measurement limit, µg/L	Calibrator name	Stated traceability
Abbott Diagnostics	Alinity i	Folate reagent kit	B8P140, November 2017	Chemiluminescent microparticle immunoassay[a]	1.40	20.0	Folate calibrators	WHO IS 03/178
Beckman Coulter	Dxl Access	Access Folate reagent pack	B03898, February 2019	Paramagnetic particle chemiluminescent immunoassay[a]	0.85	24.8	Access Folate calibrators	WHO IS 03/178
Roche Diagnostics	Cobas e801	Elecsys Folate III	07027290190, May 2019	Electrochemiluminescence binding assay	1.20	20.0	CalSet Folate	WHO IS 03/178
Siemens Healthcare Diagnostics	Advia Centaur	Folate (FOL)	10629859_EN, December 2017	Chemiluminescent assay	0.35	24.0	FOL calibrators	Internal standard manufactured using highly purified material (5-methyl tetrahydrofolate)

[a]Technically, folate assays are not immunoassays because they rely on the competition between a labeled folate standard and unlabeled folate from biological samples for a folate-binding protein.

traceability, Advia Centaur is unusual in that it uses gravimetrically prepared manufacturer's internal standards of 5-methylTHF to which MS calibrators are traceable.

Samples

As three out of four manufacturers declared that their calibrators were traceable to the WHO IS 03/178, we used this reference material for evaluating the correctness of traceability implementation by manufacturers. Based on the certificate of analysis, the IS has an assigned value of 5.33 µg/L tFOL, when reconstituted with 1.0 mL distilled water, as determined using ID-LC-MS/MS [10]. The uncertainty of the certified value is not available.

The recently released National Institute of Standards and Technology (NIST) Standard Reference Material (SRM) 3949 (frozen human serum from donors) was used as the EQA material for checking the MS trueness. A unit of SRM 3949 consists of one vial each of three materials: low (L1), medium (L3) and high (L2) folates, with corresponding tFOL values of 7.50 µg/L, 18.45 µg/L and 24.71 µg/L, respectively, as determined by ID-LC-MS/MS, and expanded uncertainty (U) (k = 2) of 0.18 µg/L, 0.22 µg/L and 0.35 µg/L, respectively. The estimated amount of tFOL concentration in nmol/L was converted to µg/L using a conventional conversion factor of 2.266 [10]. As the tFOL concentration of L2 was higher than the upper measurement limit of the majority of evaluated MSs (Table 1), this material was not used in our study. Both IS 03/178 and SRM 3949 were prepared following the instructions in the respective certificates of analysis.

Clinical samples were used for the commutability assessment of two reference materials and for the verification of the inter-assay agreement. We originally prepared 24 serum pools (~15 mL each) from residual sera of the daily routine, excluding from the collection hemolyzed, lipemic or icteric samples (interference indices estimated on an Abbott Architect c4000 platform). In particular, the hemolysis index was ≤6 (i.e. free hemoglobin concentration ≤0.06 g/L) for all pools. The recommendations of the Clinical and Laboratory Standards Institute (CLSI) C37-A guideline was observed for making pools [15]. tFOL concentrations in the pools were distributed as follows: nine <3.0 µg/L, 13 between 3.0 and 10.0 µg/L, and two >10.0 µg/L. Particularly, we kept tFOL well below the 22.0-µg/L concentration that might be expected in subjects receiving staple foods fortified with PGA [3]. Each pool was divided into aliquots and stored in polypropylene cryovials at −80 °C until use (maximum storage time was 2 months). Because the study involved anonymized leftover samples, it did not require approval by an Ethics Committee.

Experimental design

In the week before analyses, the optimal performance of MSs was verified with the support of the corresponding manufacturer. All analyses were performed in a single analytical run by the same trained technician by strictly applying the manufacturer's recommended instructions and following the analysis sequence as shown in Supplementary Figure 1. The calibration of each MS was assessed based on the acceptable ranges for control materials supplied by each manufacturer.

APS for tFOL measurements

Previous studies evaluating the performance of tFOL assays employed biological variability-derived APS [16]. In line with the resolution of the 2014 EFLM Strategic Conference, the rationale for assigning measurands to one of the APS models recommended by the conference has been reported [17]. In particular, for serum tFOL the outcome-based model for deriving APS measurements should be used because: (a) as humans are unable to synthesize folates, serum folate is not subjected to any homeostatic control, and (b) changes that hold clinical meaning are only one-sided, i.e. vitamin deficiency is diagnosed when values are lower than established thresholds [18]. We applied this model for defining desirable APS for bias and imprecision for serum tFOL measurements yielding a tolerable misclassification rate, defined as 1.7% false-negative results in subjects expected to have overt (tFOL < 2.0 µg/L) or possible (tFOL < 4.0 µg/L) vitamin deficiency. Accordingly, at tFOL concentration around 2.0 µg/L, an imprecision <12% (as CV) and a bias <11% are allowable. At a tFOL concentration of 4.0 µg/L, APS become more stringent, with the desirable bias being <3.0% when the assay CV is <3.0%, and <2.0% when the CV is <5.0% [18].

In this study, we applied the most appropriate APS for bias on the basis of the considered tFOL concentrations and the mean assay CV obtained on serum pools. Particularly, when the analysis involved clinical samples with a tFOL concentration <3.0 µg/L (the mean intra-MS CV obtained on these samples was ≤9.5%), the desirable bias goal was <11%. Differently, for tFOL values >3.0 µg/L, the desirable bias goal was set at 3.0%, as the mean intra-MS CV at these concentrations was ≤3.9%. For assessing the commutability of the employed reference materials according to the IFCC approach, we adopted a less stringent bias goal (4.5%) by moving from desirable to minimum quality level according to the classic approach by Fraser et al. ([desirable goal/0.50] * 0.75) [19]. Finally, we derived the desirable U goal of serum tFOL measurements at the clinical sample level (mean tFOL concentration of 4.0 µg/L) by consulting the figure 1 of ref. 18 and identifying on the 1.7% misclassification isocontour line the CV corresponding to the allowable bias of 3.0%, multiplied by a coverage factor of 2, which is 5.0%.

Statistical analyses

The commutability of IS 03/178 and SRM 3949 L1 materials was assessed based on the pairwise comparison of results on 14 pools with tFOL concentrations between 3.0 and 14.0 µg/L to cluster the concentrations of the clinical samples closer to that of the reference materials and improve the statistical analysis and conclusions regarding commutability [14]. The commutability of SRM 3949 L3 was not evaluated because of the lack of correspondence between its tFOL concentration and those of collected patient pools. Two statistical approaches were used to assess commutability: (a) observing the CLSI guideline EP30-A [20], using Deming regression for each pair of MS, followed by the calculation of 95% prediction intervals; (b) observing the IFCC recommendations [21], using the difference in bias between the reference material and clinical sample pools. Conclusions regarding the commutability of the reference materials were drawn based on the positions of their values with respect to each prediction interval (CLSI approach) or on the difference in bias compared with a predefined criterion based on the outcome-based model

(see previous text) (IFCC approach). The minimum quality level for acceptable bias was selected based on the use of the reference materials in this study as EQA materials.

To evaluate the correctness of traceability implementation of MSs to the WHO IS 03/178, a recovery analysis of tFOL target value in the material was performed and the results were compared with the desirable goal for bias to avoid undue false-negative results (+3.0%).

For verification of MS agreement, a preliminary visual inspection using box and whisker plots was performed on mean values obtained for pools by each MS. Therefore, the inter-assay CV on all pools and only on those with tFOL values >3.0 µg/L were calculated and compared with the desirable goal for bias (3.0%). Considering the median value of each pool for all MSs as reference, the Passing-Bablok regression analysis and difference plots were performed for each MS. For each comparison, mean estimates of percentage difference on all pools and only on those with folate values >3.0 µg/L were also calculated.

For trueness evaluation of the MSs with SRM 3949, the estimated bias was considered statistically significant if: $(x_{ref} - \bar{x}_{method}) > 2 * u_{bias}$, where x_{ref} is the tFOL certified value of the reference material, $\bar{x}_{method}$ is the mean of duplicates obtained for this material by each evaluated MS, and u_{bias} is equal to $\sqrt{(u_{ref}^2 + SD_{mean}^2)}$. Note that u_{ref} is the U from the reference material certificate divided by 2, and SD_{mean} is the standard deviation of the mean value of SRM 3949 obtained by each MS, calculated by the equation: $SD_{mean} = SD/\sqrt{n}$, where SD is the standard deviation of the duplicate measurement of the reference material and $n = 2$.

The mean U of tFOL measurements on clinical sample pools was estimated by combining the bias between the obtained mean of the WHO IS 03/178 and its target value, and the mean imprecision of a given MS, obtained from duplicate measurements of patient pools, to the uncertainty of the respective calibrator with the concentration level nearest to the IS 03/178. In estimating U, a coverage factor of 2 (95.45% level of confidence) was applied $(U = 2 * \sqrt{[u_{cal}^2 + bias^2 + CV^2]})$ [22].

Results

Commutability assessment of IS 03/178 and SRM 3949 L1

Supplementary Figures 2–7 and Figures 8–13 report in detail the results of IS 03/178 and SRM 3949 L1 commutability assessment by the CLSI and IFCC approaches, respectively. Based on the CLSI approach, both materials were within the 95% prediction intervals for the Deming regression in all pairwise comparisons showing their commutability for use with all MSs examined. A summary representation of commutability conclusions according to the IFCC statistical approach is shown in Supplementary Figure 14. Because of the relatively high scatter in the clinical sample results, the commutability of the two reference materials was frequently indeterminate, i.e. the mean difference between the two MSs was outside, but the error bars were inside of the predetermined maximum allowable commutability-related bias. Overall, the results confirmed the good commutability of the WHO IS 03/178 with all evaluated MSs. For SRM 3949 L1, some non-commutability problems were seen with Advia Centaur in the evaluation using the more demanding IFCC approach, suggesting that this material was commutable enough to be used as the trueness control material in this study, but the interpretation of the results obtained with Advia would require some caution.

Correctness of traceability implementation to the WHO IS 03/178

The MSs were evaluated for bias relative to the WHO IS 03/178 (assigned value, 5.33 µg/L). The mean values (±SD) obtained for the IS 03/178 by Access Dxl, Advia Centaur, Alinity and Cobas e801 were 5.81 (±0.14) µg/L, 5.77 (±0.13) µg/L, 5.60 (±0.42) µg/L and 4.89 (±0.36) µg/L, respectively, with a mean percentage recovery of target value of 109.0%, 108.3%, 105.1% and 91.7%, respectively. The evaluation of a single-level reference material does not permit to establish if the poor alignment detected in our study for all evaluated MSs is transferable to tFOL concentrations around the threshold for diagnosing overt vitamin deficiency (i.e. 2.0 µg/L), where positive bias until 11% is allowable. What is certain is that a bias of +5% or more at tFOL values around the threshold for possible deficiency (i.e. 4.0 µg/L) may produce an unacceptably high number of false-negative results (2.5% or more), falsely reassuring laboratory users about the vitamin status of tested individuals and potentially delaying needful supplementation [18]. On the other hand, a negative bias of ~0.50 µg/L, as detected for Roche Cobas, may increase the number of falsely low tFOL results by 5% (Andrew W. Lyon and Simona Ferraro, data on file, unpublished), thus impacting the test cost-effectiveness.

Verification of agreement among tFOL MSs

Four low-concentration pools were undetectable by Alinity (i.e. <1.40 µg/L). Therefore, results from these samples were removed when the inter-assay agreement was verified. Figure 1 shows box and whisker plots of tFOL mean concentrations obtained by different MSs on 20 patient pools. The Access Dxl showed the lowest within-method analytical variation (mean CV on duplicates, 1.94%), whereas the CV was larger for Advia Centaur (3.20%). Alinity and Cobas showed the highest CV, with values of 5.17% and 5.98%, respectively. The mean inter-assay CV on

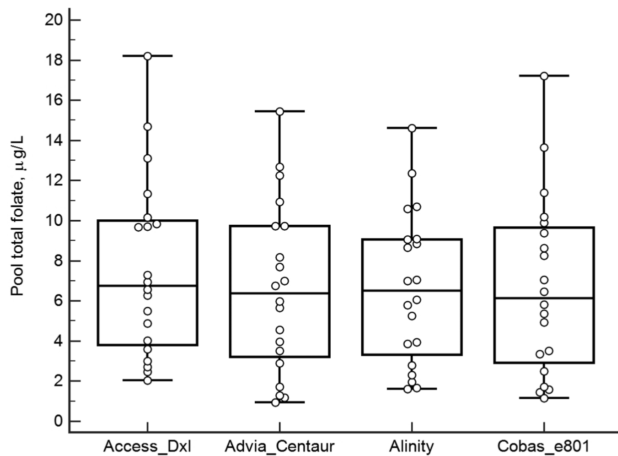

Figure 1: Box and whisker plots of total folate mean concentrations obtained by the evaluated measuring systems in 20 patient pools. The boxes show the median, 25th and 75th percentile values, and whiskers represent the 10th and 90th percentile values.

all pools and on those with tFOL values >3.0 µg/L (n = 15) was 12.5% and 7.1%, respectively. In neither case the goal of 3.0% was fulfilled. Results for each MS were then assessed against median tFOL concentrations measured by all MSs using the Passing-Bablok regression analysis. Table 2 displays the regression parameters together with the mean estimates of percentage difference over all pools and over only those with tFOL values >3.0 µg/L. Absolute and percentage difference plots are shown in Figure 2. It should be noted that this experiment was not a trueness check of different MSs (as tFOL values assigned to pools by RMP were not available), but results were, however, helpful in understanding the status of MS agreement.

The Access Dxl gave results markedly higher than MS medians (mean, +18.2%), this difference becoming more evident at tFOL concentrations <3.0 µg/L. Advia Centaur and Cobas e801 showed lower tFOL values and, once again, the situation clearly worsened when the analysis focused on the tFOL concentrations <3.0 µg/L. The Alinity appeared to be more aligned, even if results >9.0 µg/L denoted a lack in measurement linearity. All MSs had intercept significantly different from 0, related to the bias nonlinearity.

Evaluation of MS trueness by using SRM 3949 as the EQA material

Figure 3 reports tFOL results (with corresponding uncertainty) obtained with the evaluated MSs on SRM 3949 L1 and L3. Access Dxl and Advia Centaur showed a marked and relatively constant overestimation of both materials (25.6% and 17.6% for L1 and L3, and 25.5% and 21.5% for L1 and L3, respectively). Alinity overestimated L1 (+6.67%), while L3 was underestimated (−8.67%). For Cobas e801, only results for L1 were displayed (mean bias, +7.87%) as L3 gave results, although overestimated, out of assay linearity (>20.0 µg/L). Overall, all MSs showed a significant positive bias when compared to ID-LC-MS/MS-assigned values (see Table in the online Data Supplement) that may produce a great impact on clinical interpretation of tFOL results [18].

MU of tFOL MSs on clinical samples

Table 3 reports estimates of mean U on patient pools for each of the evaluated MSs. Because the U of IS 03/178 was not available, it should be noted that the reported uncertainties of manufacturers' calibrators traced to this material are not combined to that of the employed reference and, therefore, they are in principle underestimated. Nevertheless, all evaluated MSs were unable to fulfil the APS for U (5.0%) on clinical samples. Results indicated the residual bias as the main source of uncertainty in all MSs.

Discussion

To comply with the requirements of the EU Directive, the majority of manufacturers of tFOL MSs currently calibrate

Table 2: Passing-Bablok regression analysis and mean percentage difference of measured total folate values for patient pools vs. median total folate concentrations for the evaluated measuring systems.

Measuring system	Regression parameters		Mean difference, % (CI) [all pools, n = 20]	Mean difference, % (CI) [pools with values >3.0 µg/L, n = 15]
	Slope (95% CI)	Intercept, µg/L (95% CI)		
Alinity i	0.918 (0.873–0.978)	0.36 (0.11–0.64)	0.8 (−2.5–4.1)	−1.7 (−4.2–0.8)
Dxl Access	1.036 (0.950–1.087)	0.60 (0.23–0.93)	18.2 (9.9–26.5)	9.0 (6.1–11.9)
Cobas e801	1.028 (0.988–1.081)	−0.33 (−0.59 to −0.10)	−5.2 (−8.4 to −2.1)	−2.9 (−5.4 to −0.3)
Advia Centaur	1.027 (0.969–1.072)	−0.34 (−0.58 to −0.04)	−6.8 (−11.5 to −2.2)	−2.5 (−4.5 to −0.4)

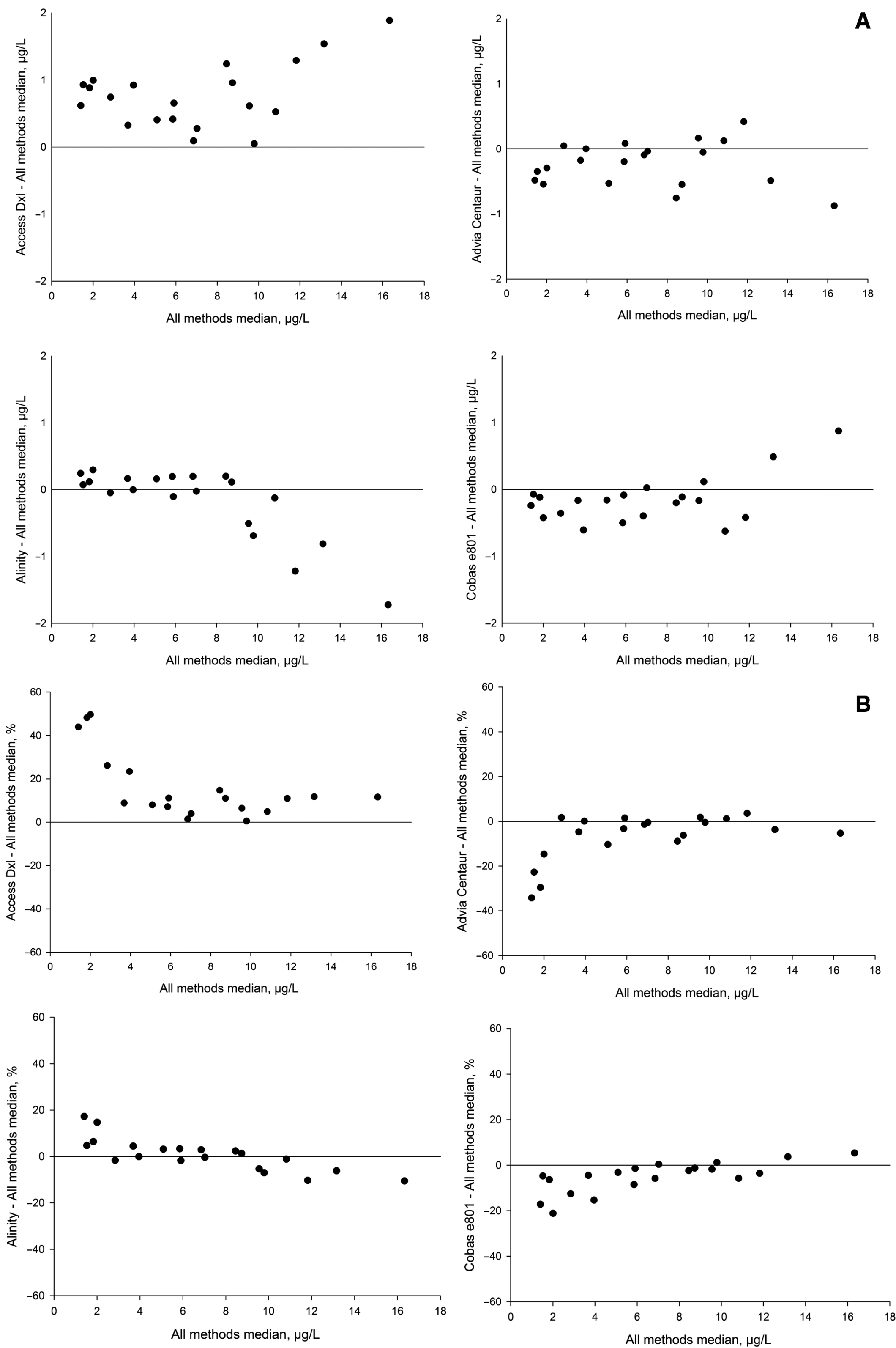

Figure 2: Absolute (panel A) and percentage (panel B) difference plots of each measuring system (MS) vs. all MSs median total folate (tFOL) values for 20 pools.
Access gave results markedly higher than MS medians, this difference becoming more evident at tFOL concentrations <3.0 µg/L. Advia Centaur and Cobas showed lower tFOL values, with the situation worsening at tFOL concentrations <3.0 µg/L. Alinity appeared to be more aligned, even if results >9.0 µg/L denoted a lack in measurement linearity.

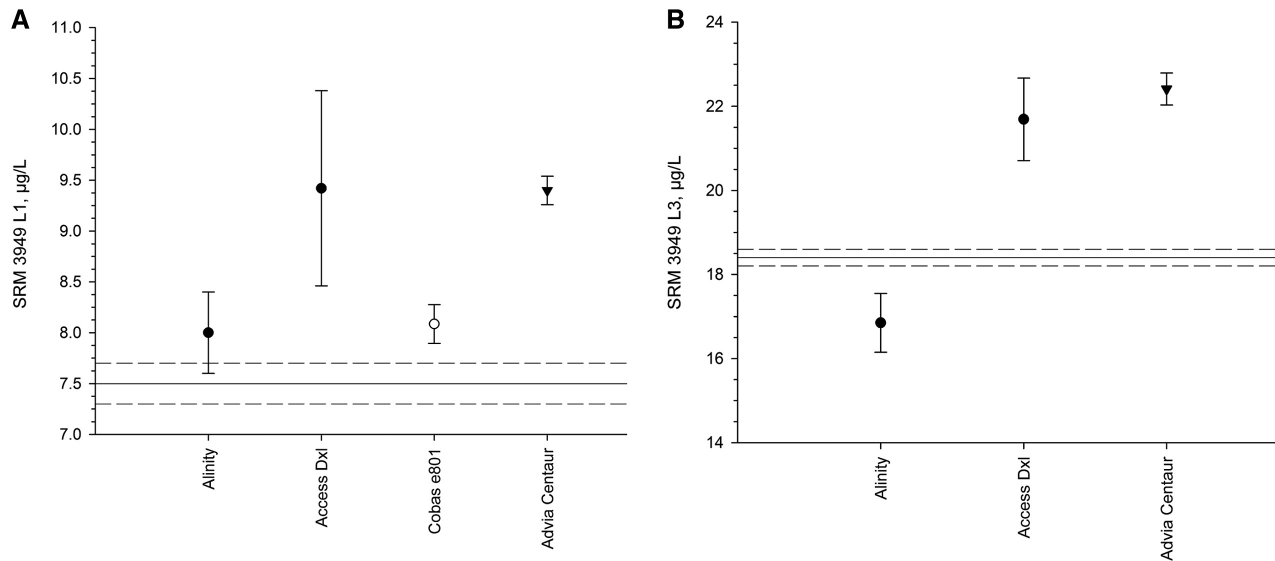

Figure 3: Total folate results by the evaluated measuring systems vs. target values for the National Institute of Standards and Technology Standard SRM 3949 L1 (panel A) and L3 (panel B).
The solid line is the target value certified for the reference material, and the dashed lines indicate its uncertainty (k = 2). The error bars for each measuring system represent the uncertainty (k = 2) for the method mean values. Results for SRM 3949 L3 by Cobas e801 are lacking because of assay linearity (>20.0 µg/L). Note that SRM L3 was not assessed for commutability in this study.

Table 3: Mean relative combined measurement uncertainty (expanded by a coverage factor of 2) on patient pools for each evaluated total folate measuring system.

Measuring system	Calibrator type	Calibrator standard uncertainty, %[a]	Bias vs. WHO IS 03/178 target value (5.33 µg/L), %[b]	Method CV, %[c]	Mean combined uncertainty on patient pools, %[d]
Alinity i	Folate calibrators	0.33–0.23[e] (5 levels, from 1.5 to 20.0 µg/L)	5.07	3.31	12.12[f]
Dxl Access	Access Folate calibrators	2.02–1.69[e] (5 levels, from 1.24 to 24.8 µg/L)	9.01	1.60	18.61[g]
Cobas e801	CalSet Folate	6.31–1.45 (2 levels, 1.75 and 15.8 µg/L)	−8.26	3.58	18.24[h]
Advia Centaur	FOL calibrators	5.25–2.75 (2 levels, 2.7 and 16.5 µg/L)	8.26	2.05	17.89[i]

[a]Except for Advia Centaur, not combined with the uncertainty of the corresponding higher-order reference material. [b]Data from this study. [c]The mean imprecision of measuring systems was obtained from duplicate measurements of six clinical sample pools with serum total folate concentrations between 3.0 and 7.0 µg/L. [d]Expanded by multiplying the standard uncertainty by a coverage factor of 2 (95.45% level of confidence). For suitable clinical application of serum total folate measurements, the expanded measurement uncertainty at the patient sample level should remain within ±5.0%. [e]It depends on the concentration level. [f]Estimated using the standard uncertainty of calibrator level D, i.e. 0.30% at a total folate concentration of 5.0 µg/L. [g]Estimated using the standard uncertainty of calibrator S3, i.e. 1.69% at a total folate concentration of 6.2 µg/L. [h]Estimated using the standard uncertainty of calibrator 2, i.e. 1.45% at a total folate concentration of 15.8 µg/L. [i]Estimated using the standard uncertainty of calibrator High, i.e. 2.75% at a total folate concentration of 16.5 µg/L.

their internal calibrators to the WHO IS 03/178 [12]. We conducted this study to establish if the use of this material as a higher-order reference has actually improved the between-assay harmonization, allowing the intermethod variability to decrease to a level suitable for clinical application of tFOL measurements.

According to the requirements for reference materials used as common calibrators [14], the IS 03/178 showed good commutability for all evaluated MSs. This was in agreement with results from a previous study [10], which demonstrated, although indirectly, the commutability of this material by applying an approach more recently advocated by the IFCC [23]. Access Dxl, Advia Centaur and Alinity showed a positive bias vs. the IS 03/178 target value that was greater than the desirable goal to avoid excessive false-negative results. By consulting the instructions for use of those MSs, we discovered that Beckman and Abbott considered as validation criterion for calibrator traceability to IS 03/178 an internal specification of ±10%. The overestimation obtained in this study seems, therefore,

to fulfill the manufacturer's internal specifications. However, it is clear from our data that the ±10% criterion leads to an excessive bias on clinical samples preventing the achievement of the combined MU goal for tFOL measurements. As previously demonstrated for other analytes, our results show that manufacturers should conform their internal protocols of trueness transfer from the certified reference material to commercial calibrators to the clinical value of the test [24, 25].

From this study, an insufficient agreement among tFOL MSs emerges, and the situation clearly worsened when the analysis focused on the tFOL concentrations <3.0 μg/L, indicating that a common threshold for vitamin deficiency detection should not be used and an MS-dependent cut-off is still required. The results obtained by using SRM 3949 for evaluating the MS trueness confirmed the existence of a highly significant bias. It should be, however, noted that SRM 3949 was developed after a human subjects research determination by NIST, supported by the NIH Office of Dietary Supplements [26]. So, it could be possible that some supplemented individuals were involved as blood sample source. As commercial MSs have a tendency to overrecover PGA, this suggests an increased bias in tFOL measurements for samples containing increased concentrations of PGA, typical of individuals receiving fortified foods [27]. As clinically requested, commercial MSs have been developed for the measurement of sera from individuals with suspected vitamin deficiency, not receiving staple foods fortified with PGA; therefore, the samples used in SRM 3949 preparation might not provide the optimum condition under which to look at the ability of the FBP-based MSs to quantify tFOL. Indeed, the NIST-reported intended use of SRM 3949 is in validating methods for determining folate vitamers in human serum, not for assessing accuracy and validating calibration of field methods used in medical laboratories [26].

Our study has some limitations. Firstly, our study did not include folate results on pools obtained by ID-LC-MS/MS RMP. This method is currently available only at the Centers for Disease Control and Prevention [26] and it was impossible for us to include it in our experiments. Definitive data about the relationship between current MSs selectivity and folate forms detected in the ID-LC-MS/MS are lacking and, as previously shown for other measurands [28], difference in the method selectivity between competitive binding folate assays and ID-LC-MS/MS may make it difficult to control the traceability chain well enough and do a real trueness evaluation by using available reference materials. Secondly, the use of pools instead of native samples may not be optimal in the commutability assessments. However, having strictly followed the CLSI C37-A recommendations for their preparation, we are confident that our pools may reasonably behave as individual patient samples [29]. Thirdly, using the IFCC statistical approach for evaluating commutability of reference materials, an indeterminate conclusion was frequently observed. As previously discussed [25], this fact may suggest that the experimental design employed in our study is not optimal for the IFCC approach. In fact, the original IFCC model recommends triplicate measurements of the clinical samples (instead of duplicates), more clinical samples (at least 30 instead of 14) and measuring the reference materials in different positions of the analytical run [21]. Apart from these considerations, even if correctly defined, a commutability criterion of 4.5% is probably not realistic for the performance of present MSs.

In conclusion, although the quality of IS 03/178 appears to be good, the implementation of traceability to it is far from perfect. We expect manufacturers producing biased MSs to invest more effort in implementing internal protocols for transferring trueness to their calibrators by applying outcome-based APS as those employed in this study.

Acknowledgments: We thank Antonella Antononi, Laboratorio Analisi Cliniche, Centro Diagnostico Italiano, Milano and Maria Teresa Sandri, Unità Operativa di Analisi Cliniche, Istituto Humanitas, Rozzano, who enthusiastically joined the study. We thank all manufacturers for their technical support. Abbott Diagnostics is acknowledged for the generous gift of folate reagents.
Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: This study was partially supported by the Research Support Plan Fund 2018 of the University of Milan.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Roberts NB, Taylor A, Sodi R. Vitamins and trace elements. In: Rifai N, Horvath AR, Wittwer CT, editors. Tietz Textbook of Clinical Chemistry snd Molecular Diagnostics. 6th ed. St. Louis, MO, USA: Elsevier; 2018:639–718.

2. WHO. Guideline: Optimal serum and red blood cell folate concentrations in women of reproductive age for prevention of neural tube defects. Geneva: World Health Organization; 2015.
3. Ferraro S, Panzeri A, Panteghini M. Tackling serum folate test in European countries within the health technology assessment paradigm: request appropriateness, assays and health outcomes. Clin Chem Lab Med 2017;55:1262–75.
4. Owen WE, Roberts WL. Comparison of five automated serum and whole blood folate assays. Am J Clin Pathol 2003;120:121–6.
5. Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. Clin Biochem 2009;42:236–40.
6. Panteghini M. Traceability, reference systems and result comparability. Clin Biochem Rev 2007;28:97–104.
7. Hannisdal R, Ueland PM, Svardal A. Liquid chromatography–tandem mass spectrometry analysis of folate and folate catabolites in human serum. Clin Chem 2009;55:1147–54.
8. Blackmore S, Pfeiffer CM, Lee A, Fazili Z, Hamilton MS. Isotope dilution-LC-MS/MS reference method assessment of serum folate assay accuracy and proficiency testing consensus mean. Clin Chem 2011;57:986–94.
9. Pfeiffer CM, Fazili Z, McCoy L, Zhang M, Gunter EW. Determination of folate vitamers in human serum by stable isotope dilution tandem mass spectrophotometry and comparison with radioassay and microbiologic assay. Clin Chem 2004;50:423–32.
10. Thorpe SJ, Heath A, Blackmore S, Lee A, Hamilton M, O’Broin S, et al. International standard for serum vitamin B12 and serum folate: international collaborative study to evaluate a batch of lyophilized serum for B12 and folate content. Clin Chem Lab Med 2007;45:380–6.
11. Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. Official Journal of the European Communities 1998 Dec 7. L331.
12. Ferraro S, Panteghini M. Folate and vitamin B12 assays after recalibration to the WHO International Standard 03/178: making the interpretation as simple as possible, but not simpler. Clin Chem Lab Med 2019;57:1112–4.
13. Ferraro S, Panteghini M. Defining the plasma folate concentration for optimal neural tube defects prevention cannot ignore the impact of the employed methodology. Am J Clin Nutr 2019 [ahead of print].
14. Braga F, Panteghini M. Commutability of reference and control materials: an essential factor for assuring the quality of measurements in Laboratory Medicine. Clin Chem Lab Med 2019;57:967–73.
15. CLSI. Preparation and validation of commutable frozen human serum pools as secondary reference materials for cholesterol measurement procedures; approved guideline. Document C37-A. Wayne (PA): CLSI; 1999.
16. Kristensen GB, Rustad P, Berg JP, Aakre KM. Analytical bias exceeding desirable quality goal in 4 out of 5 common immunoassays: results of a native single serum sample external quality assessment program for cobalamin, folate, ferritin, thyroid-stimulating hormone, and free T4 analyses. Clin Chem 2016;62:1255–63.
17. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. Clin Chem Lab Med 2017;55:189–94.
18. Ferraro S, Lyon AW, Braga F, Panteghini M. Definition of analytical quality specifications for serum total folate measurements using a simulation outcome-based model. Clin Chem Lab Med 2019;58:e66–8.
19. Fraser CG, Hyltoft Petersen P, Libeer JC, Ricos C. Proposal for setting generally applicable quality goals solely based on biology. Ann Clin Biochem 1997;34:8–12.
20. CLSI. Characterization and qualification of commutable reference materials for laboratory medicine; approved guideline. Document EP30-A (formerly C53-A). Wayne (PA): CLSI; 2010.
21. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC working group recommendations for assessing commutability part 2: using the difference in bias between a reference material and clinical samples. Clin Chem 2018;64:455–64.
22. ISO/TS 20914:2019. Medical laboratories – Practical guidance for the estimation of measurement uncertainty, 1st ed. Geneva, Switzerland: ISO, 2019.
23. Budd JR, Weykamp C, Rej R, MacKenzie F, Ceriotti F, Greenberg N, et al. IFCC working group recommendations for assessing commutability part 3: using the calibration effectiveness of a reference material. Clin Chem 2018;64:465–74.
24. Pasqualetti S, Infusino I, Carnevale A, Szőke D, Panteghini M. The calibrator value assignment protocol of the Abbott enzymatic creatinine assay is inadequate for ensuring suitable quality of serum measurements. Clin Chim Acta 2015;450:125–6.
25. Braga F, Infusino I, Frusciante E, Ceriotti F, Panteghini M. Trueness evaluation and verification of interassay agreement of 11 serum IgA measuring systems: implications for medical decisions. Clin Chem 2019;65:473–83.
26. National Institute of Standards & Technology. Certificate of Analysis. Standard Reference Material 3949, Folate Vitamers in Frozen Human Serum. Gaithersburg, MD. Certificate Issue Date: 22 October 2018.
27. Blackmore S, Pfeiffer C, Hamilton MS, Lee A. Recoveries of folate species from serum pools sent to participants in the UK NEQAS Haematinics Scheme in February and March 2004. Clin Chim Acta 2005;355(Suppl):S459.
28. Boulo S, Hanisch K, Bidlingmaier M, Arsene C-G, Panteghini M, Auclair G, et al. Gaps in the traceability chain of human growth hormone measurements. Clin Chem 2013;59:1074–82.
29. Danilenko U, Vesper HW, Myers GL, Clapshaw PA, Camara JE, Miller WG. An updated protocol based on CLSI document C37 for preparation of off-the-clot serum from individual units for use alone or to prepare commutable pooled serum reference materials. Clin Chem Lab Med 2020;58:368–74.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2019-0928).