



# Lessons from the NIST micronutrients quality assurance program for vitamin C, 1993 to 2015: sample stability, assay reproducibility, and use of controls to improve comparability

David L. Duewer<sup>1</sup> · Jeanice Brown Thomas<sup>2</sup> · Katherine E. Sharpless<sup>3</sup> · Sam A. Margolis<sup>4</sup>

Received: 8 September 2020 / Revised: 7 October 2020 / Accepted: 20 October 2020 / Published online: 2 November 2020

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020

## Abstract

Vitamin C is a necessary micronutrient that is involved in many biological processes. In preserved human plasma and serum, vitamin C is most meaningfully analyzed as total ascorbic acid (TAA). From 1993 through 2015, the National Institute of Standards and Technology (NIST) coordinated 40 interlaboratory studies (ILS) devoted to improving the between-participant comparability of TAA measurements. The results from these ILS demonstrate that the concentration of TAA ([TAA]) is stable for at least 20 years in serum diluted 1 + 1 (volume fraction) with 10% mass concentration aqueous metaphosphoric acid (MPA) and stored at  $-80\text{ }^{\circ}\text{C}$ . The between-participant relative reproducibility precision, expressed as a coefficient of variation (CV), improved from over 16% to under 9% over the course of the studies. Normalization of test samples (i.e., *ex post-facto* recalibrating the as-submitted results) using participant-prepared serum-free calibration solutions did not improve reproducibility. Normalization to one control sample having a similar serum-based matrix as the test samples improved the CV to 7%; normalization to two such controls reduced the CV to 4%. Multicenter studies that require the highest degree of measurement comparability among the participants should consider calibrating with materials that have a serum-based matrix as similar as possible to that of the samples of interest.

**Keywords** Ascorbic acid · Interlaboratory studies · Reproducibility · Stability

## Introduction

Vitamin C, a composite of L-ascorbic acid (AA) and dehydro-L-ascorbic acid (DHAA), is a micronutrient with many biological

activities of potential clinical interest in addition to the prevention of scurvy [1, 2]. As with other measurement processes used to inform decisions, those used for vitamin C need to reliably provide metrologically valid results largely independent of when and where the measurements are made [3–5].

In 1984, what is now the National Institute of Standards and Technology (NIST) started investigating how to help improve the within- and between-laboratory comparability of vitamin C measurements in human plasma and serum. This work was conducted as part of a National Cancer Institute (NCI)-sponsored program charged with developing analytical methods, reference materials, and interlaboratory studies (ILS) for micronutrients hypothesized to have cancer chemopreventive activity [6]. Informed by a series of exploratory ILS, a chromatographic reference method for AA and DHAA was developed, and various approaches to providing stable reference materials were explored. [7–9]. While vitamin C test samples were distributed in several early ILS [10] conducted under the auspices of what became known as the Micronutrients Measurement Quality Assurance Program (MMQAP), only in 1993 was it established that the vitamin

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00216-020-03021-9>.

✉ David L. Duewer  
david.duewer@NIST.gov

<sup>1</sup> Chemical Sciences Division, Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899-8390, USA

<sup>2</sup> Chemical Sciences Division, Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899-8392, USA

<sup>3</sup> Special Programs Office, National Institute of Standards and Technology, Gaithersburg, MD 20899-4701, USA

<sup>4</sup> National Institute of Standards and Technology, Gaithersburg, MD 20899-8390, USA

C-related measurand best suited for ILS study was the sum of AA and DHAA (total L-ascorbic acid, TAA) [11, 12]. Also in 1993, NIST established that the TAA concentration ([TAA]) of serum diluted 1 + 1 (volume fraction) with 10% mass concentration aqueous metaphosphoric acid (MPA) was: 1) stable for at least 5 years when stored at  $-80^{\circ}\text{C}$  and 2) could be successfully determined by all of the measurement methods then available. The relative reproducibility precision of the results provided by these methods was  $\approx 15\%$  [13].

The NCI funding for MMQAP expired in 1997. NIST elected to continue development of a certified reference material for vitamin C in human serum, using participation fees to partially defray direct ILS expenses. Standard Reference Material® (SRM®) 970 Ascorbic Acid in Frozen Human Serum was released for sale in 2000 [14]. SRM 970 provided two 1 + 1 mixtures of human serum and 10% MPA: the high-normal Level I and the low-normal Level II. The vitamin C ILS continued until 2015. The stability of the SRM 970 solutions was monitored and approaches to reducing between-participant measurement variability were tested. The MMQAP vitamin C program was ended due to declining participation, exhaustion of the supply of SRM 970, and the availability of international proficiency tests for vitamin C initiated in 2010 by the Royal College of Pathologists of

Australia Quality Assurance Programs (RCPAQAP) [5] and in 2014 by the German INSTAND e.v. External Quality Assurance Study (EQAS) program [15].

Forty MMQAP vitamin C ILS were conducted from 1993 through 2015 that exclusively used MPA-stabilized serum test samples. These samples contained no other stabilizers. Most of the 23 test materials studied were distributed multiple times over this 22-year period. A number of lessons learned from the results from these studies may be relevant to current measurement practice, particularly as preservation with MPA continues to be a widely-used approach to stabilizing the vitamin C content of plasma and serum samples [16–18].

## Methods and materials

### MMQAP vitamin C Interlaboratory comparability studies

Table 1 lists the number of participants, test samples, and control samples distributed in the MMQAP vitamin C ILS from 1993 thru 2015. These studies focused on measurements of [TAA] in MPA-stabilized human serum. Table 1 also

**Table 1** Number of Participants, Test Samples, Control Samples, and Calibration Solutions per Study

Period 1: 1993 to 2000						Period 2: 2001 to 2008						Period 3: 2008 to 2015					
RR <sup>a</sup>	Date <sup>b</sup>	P <sup>c</sup>	T <sup>d</sup>	N <sup>e</sup>	C <sup>f</sup>	RR <sup>a</sup>	Date <sup>b</sup>	P <sup>c</sup>	T <sup>d</sup>	N <sup>e</sup>	C <sup>f</sup>	RR <sup>a</sup>	Date <sup>b</sup>	P <sup>c</sup>	T <sup>d</sup>	N <sup>e</sup>	C <sup>f</sup>
4	9/1993	14	3			14	4/2001	13	4		3	29	8/2008	13	4	1	4
5	5/1994	15	2			15	9/2001	15	3		4	30	3/2009	14	4	1	4
6	9/1994	11	2			16	4/2002	16	3		4	31	9/2009	13	4	1	4
7	5/1995	17	2	1		17	9/2002	13	3		4	32	2/2010	13	4	1	4
8	12/1995	15	2	1		18	2/2003	13	3		4	33	9/2010	12	4	1	4
9	6/1996	17	2			19	9/2003	13	4		4	34	2/2011	12	4	1	4
10	1/1997	20	2		1	20	3/2004	11	4	2	4	35	8/2011	10	4	1	4
11	10/1998	17	2		1	21	9/2004	10	4	2	4	36	4/2012	10	4	1	4
12	3/1999	14	2		1	22	2/2005	11	4	2	4	37	8/2012	9	4	2	4
13	5/2000	18	4		1	23	9/2005	10	4	2	4	38	3/2013	9	4	2	
						24	3/2006	12	4	2	4	39	8/2013	9	4	2	
						25	8/2006	11	4	2	4	40	4/2014	10	4	2	
						26	2/2007	11	4	1	4	41	9/2014	9	4	2	
						27	8/2007	11	4	1	4	42	4/2015	7	4	2	
						28	2/2008	13	4	1	4	43	8/2015	7	4	2	

a) “Round Robin” study identifier

b) Target date (month/year) for participants to provide results to NIST

c) Number of participants in the study that provided results

d) Number of test samples distributed in the study

e) Number of NIST-provided matrix-based control samples in the study

f) Number of ascorbic acid calibration solutions participants were requested to prepare using NIST-provided solid ascorbic acid

displays the number of aqueous ascorbic acid calibrants that participants were requested to prepare.

Calibrants were solutions of high-purity AA in serum-free 5% mass concentration aqueous MPA, thus matching the MPA concentration in the test and control samples. NIST supplied the solid AA along with the test and control samples. Participants were asked to prepare the solutions according to a protocol provided by NIST, calculate their expected [TAA], measure the [TAA] of the solutions as if they were test samples, and then compare calculated and experimental results. In the 1997 to 2000 ILS, participants were asked to prepare a single solution having nominal [TAA] of 50  $\mu\text{mol/L}$ ; later studies also requested solutions of nominal concentration {25, 12.5, and 0}  $\mu\text{mol/L}$ . To confirm their preparation, participants were asked to determine the absorbance maximum of the 50  $\mu\text{mol/L}$  solution and calculate the mass attenuation coefficient ( $E^{1\%}$  extinction coefficient). The results provided by the analysis of these solutions were intended to help participants evaluate the performance of their measurement systems. Participants were instructed not to analyze the control or test samples until they were satisfied that their measurement processes were performing properly.

Control samples were serum-based materials that had been previously distributed as test samples. Participants were informed of the target [TAA] of these materials and were asked to confirm that their measurement processes yielded results that were within about 10% of the target before analyzing the test samples.

The protocols for each study are documented in the 7880 series of NIST Internal/Interagency Reports (NISTIRs); these documents and associated data are publicly available free of charge. Table S1 in the Electronic Supplementary Material (ESM) provides links to these reports.

## Test and control samples

During the 1993 to 2015 period, 23 MPA-stabilized sera were distributed as test samples; 6 of these materials were also distributed as control samples. These sample materials were prepared at various times over the course of the MMQAP vitamin C program, using ten different serum pools. To enable use with colorimetric methods, none of these samples contained dithiothreitol, a strongly colored antioxidant that had been included in some test materials prepared prior to 1993 [19]. To provide sample materials that were similar to the participants' "real" samples, all samples were delivered as (frozen) liquids (stored at  $-80\text{ }^{\circ}\text{C}$  and shipped on dry ice) rather than as lyophilized (and relatively temperature insensitive) solids. ESM Table S2 describes these materials.

The test and control samples delivered to study participants were coded with the study name and index; the sample materials themselves were assigned arbitrary names at the time of production. ESM Table S3 provides the association between the sample names and the identifiers used in the ILS.

## Methods of analysis

Participants in every ILS were asked to report their analysis methods. The methods used included a variety of commercial autoanalyzers ( $\approx 20\%$  of all methods), colorimetry ( $\approx 13\%$ ), liquid chromatography (LC) with electrochemical (EC) detection (32%), LC with absorbance (Abs) detection ( $\approx 26\%$ ), LC with fluorescence (Fluor) detection ( $\approx 8\%$ ), and LC with tandem mass spectrometric (MS/MS) detection ( $\approx 2\%$ ). ESM Table S4 lists the proportion of methods used during three seven-year periods: 1993 to 2001, 2001 to the first 2008 ILS, and the second 2008 study thru 2015. While the proportions did not change dramatically over time, the use of autoanalyzer, LC-Abs, and LC-MS/MS methods increased while the use of colorimetric and LC-EC methods decreased.

## Reporting units

In the vitamin C studies before 1993, confusion was observed concerning the units used to report AA, DHAA, and TAA concentrations. Some participants reported results in terms of what was in the test materials (concentration in the final 1 + 1 solution) and others in terms of the implied [TAA] of the undiluted serum (i.e., twice the [TAA] in solution). Confusion was also apparent between reporting in mass concentration ( $\mu\text{g/mL}$ ) or molar concentration ( $\mu\text{mol/L}$ ). Since the majority of participants preferred  $\mu\text{mol/L}$ , from 1993 on results were recorded and analyzed in terms of micromoles of TAA per liter solution.

To facilitate comparison of the MMQAP vitamin C results with serum vitamin C levels as reported in the literature, the recorded results are here multiplied by a factor of 2 as a transformation into micromoles per liter serum.

## Interlaboratory study robust summary statistics

Interlaboratory study data often contain "outliers" – results that are not concordant with the majority of the data. Summarizing the performance of the majority is typically accomplished by identifying and excluding the discordant results before summarizing as the mean and standard deviation or by using robust estimators that are relatively insensitive to the presence of discordant values [20]. Explicit identification can create tension between study coordinators and participants; use of robust statistics is both simpler and less susceptible to misunderstandings.

The MMQAP studies used the median as the robust estimator of "average" values. While not the most sophisticated statistic, the median is robust, reasonably efficient, familiar, and easily explained [21]. The MMQAP reports summarized the spread of the concordant results using the adjusted median absolute deviation from the median ( $\text{MAD}_E$ ) [22]. However, this estimator tends to underestimate the dispersion of truly

normally distributed data (poor statistical efficiency), particularly when the suspect values cannot be assumed to be symmetrically distributed about the median [23]. In this report we replace the  $MAD_E$  estimates with those from the robust, efficient, and symmetry-insensitive  $Q_n$  estimator [23, 24].

## Results and discussion

### MPA-stabilized Total ascorbic acid stored at $-80\text{ }^\circ\text{C}$

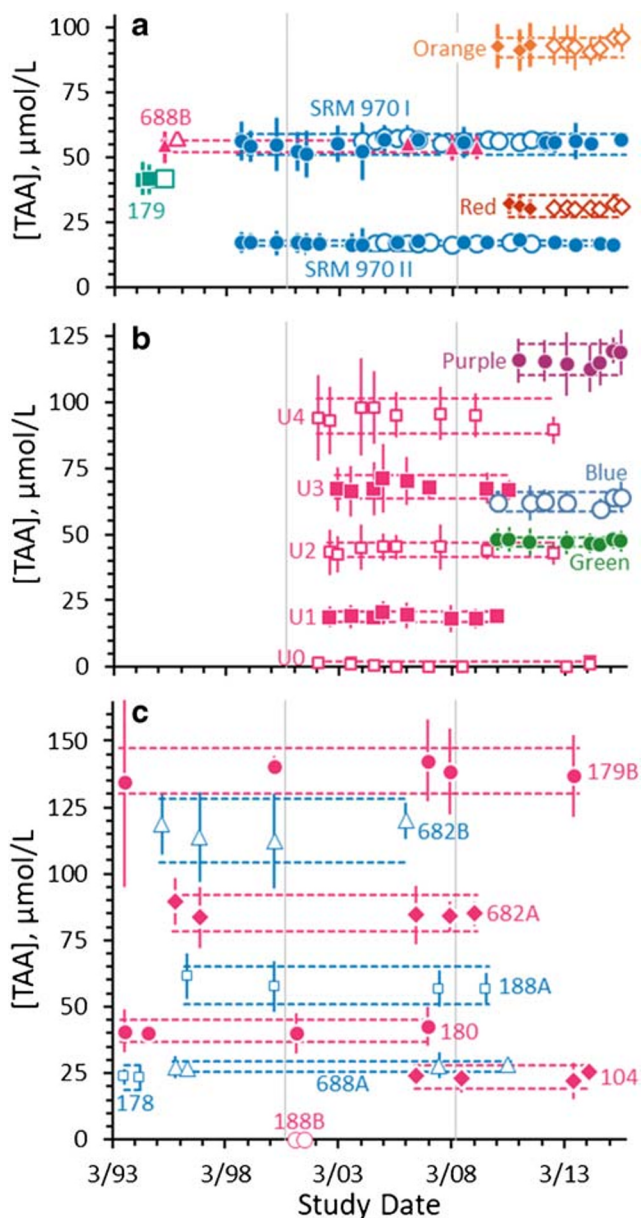
Figure 1 displays in a dot-and-bar format the consensus [TAA] value and the between-participant variability for every sample in each of the ILS. The consensus values (the “dots”) are the medians of the reported results for the given ILS; the variabilities (the “bars”) are robust standard deviations. Fig. 1 also summarizes the combined results for all distributions of each test material with 95% level of confidence intervals estimated from: the number of times the material was distributed as a test sample, the mean of the median results from each of those distributions ( $x_i$ ), and the standard deviation of those medians.

Figure 1A displays results for all of the sample materials distributed as both test and control materials. No appreciable difference was observed in the medians between the two uses. Fig. 1B displays results for the “U0” to “U4” series that was produced at NIST and the “Red” to “Purple” series that was produced commercially using NIST-supplied materials and protocol. No performance differences were attributable to the producer of the materials. Fig. 1C displays results for the legacy samples produced at NIST before 1996. Little or no evidence of [TAA] decline in any of the samples was observed.

The “104” sample was produced in 1989 and evaluated in a small ILS in 1990. The NIST and ILS mean values and their standard deviations from that time were  $(23.0 \pm 0.7)\text{ }\mu\text{mol/L}$  and  $(22.1 \pm 0.7)\text{ }\mu\text{mol/L}$ , respectively [9]. Since relatively few samples of this material were produced, it was not distributed in the MMQAP studies until the number of participants per ILS had declined sufficiently to ensure that samples damaged in shipment could be replaced. The mean and its standard uncertainty for the four distributions of the test sample is  $(23.7 \pm 0.7)\text{ }\mu\text{mol/L}$ , suggesting that the [TAA] content in this early MPA-reserved serum did not degrade over 24 years.

### Changes in performance over time

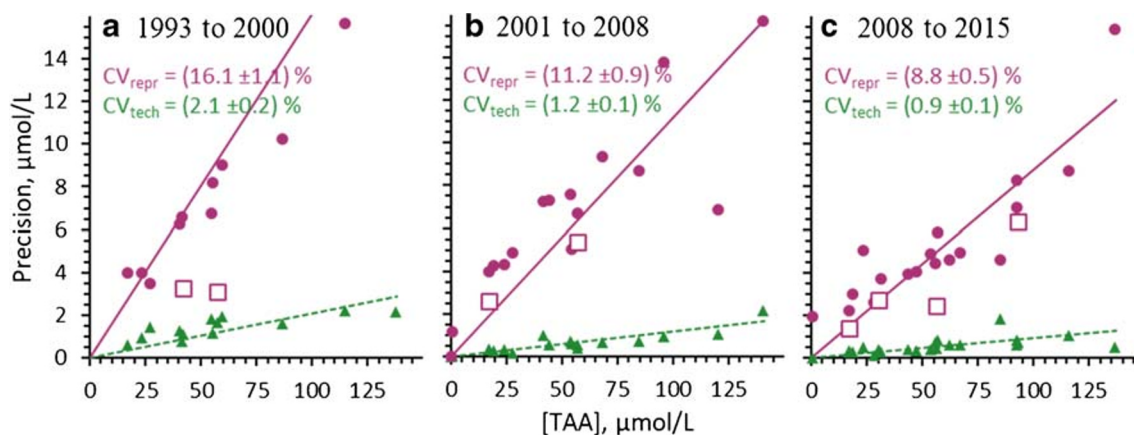
Figure 2 displays estimates of within-participant (technical) and between-participant (reproducibility) precision as functions of [TAA] for the three seven-year periods from 1993 to 2015.



**Fig. 1 Summary Results for Test Materials Over Time.** Each symbol represents the median participant total ascorbic acid concentration, [TAA], for one test material in one study; bars represent robust  $Q_n$  standard deviation estimates. Dashed horizontal lines bound approximate 95% level of confidence intervals on the medians for all studies that used the material. The vertical lines mark the seven-year intervals. A) Materials used as test materials (solid symbols) and as controls (open symbols). B) Test materials prepared in 2001 at NIST (squares) and commercially in 2009 (circles). C) Test materials prepared at NIST from 1989 to 1995

### Technical precision

In all of the MMQAP vitamin C ILS, participants were asked to report results for two replicate analyses of each calibrant, control sample, and test sample. This design was intended to enable assessing within-participant variability while keeping the analytical burden on the participants within reasonable



**Fig. 2 Test Material Within- and Between-Participant Precision Estimates Over Time.** Each panel reports within- and between-participant precision estimates for total ascorbic acid concentration, [TAA], over one of three seven-year periods: A) 1993 to 2000 (RR4 to RR13), B) 2001 to 2008 (RR14 to RR28), and C) 2008 to 2015 (RR29 to RR43). Each symbol represents results for one test material, combining summary results from all studies that used the material during the period. Solid triangles represent technical precision estimated from the two

results per sample reported for each study. The dotted line represents the proportionality of technical precision to [TAA]; the slope of this line defines the coefficient of variation,  $CV_{\text{tech}}$ . Solid circles represent reproducibility precision estimated by the  $Q_n$  robust standard deviations of participant results for test samples. The solid line represents the proportionality of reproducibility precision to [TAA]; the slope of this line defines the coefficient of variation,  $CV_{\text{repr}}$ . Open squares represent  $Q_n$  reproducibility precision estimates for control samples

limits. Participants were free to interpret this request in light of their own policies. Some participants likely reported results for replicate injections of one sample preparation (instrumental precision conditions) while others reported results from independent preparations (repeatability precision conditions). In consequence, this mixture of within-participant variability is summarized as “technical” precision. For each material  $i$  distributed as a control or test sample in distribution  $j$  and reported in replicate by participant  $k$ ,  $s_{ijk}$  is the standard deviation of those replicates,  $s_{ij}$  is the median of the  $s_{ijk}$ , and  $s_i$  is the pooled standard deviation (square root of the mean  $s_{ij}^2$ ) for all the distributions within a given time period. These results are then combined by regression using the model  $s_i = \beta_{\text{tech}} x_i$ . The technical precision for the period can be expressed as a percent relative standard deviation (coefficient of variation, CV):  $CV_{\text{tech}} = 100\beta_{\text{tech}}$ .

The composite  $CV_{\text{tech}}$  estimates improve over time, declining from 2.1% to 0.9%. These values are compatible with the 2% to 3% repeatability estimates reported for several methods [25–27]. However, whether the improvement reflects improved control of within-laboratory measurements or a greater proportion of participants reporting instrumental rather than repeatability precision cannot be determined. In any case, within-participant variability does not contribute much to the between-participant variability.

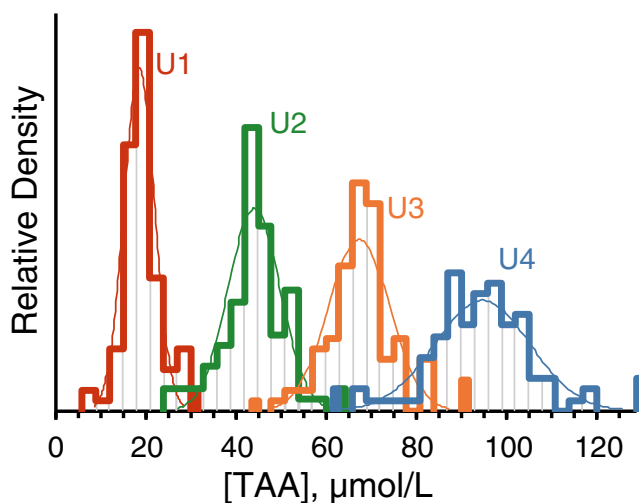
## Reproducibility

The relative ILS reproducibility precision can, in principle, be estimated similarly: calculate the mean result for each material  $i$  in distribution  $j$  reported by each of the  $k$  participants, the

standard deviation of those means ( $\sigma_{ij}$ ), the pooled standard deviation for all the distributions within the period ( $\sigma_i$ ), and combine these values using the regression model  $\sigma_i = \beta_{\text{repr}} x_i$ . The reproducibility for the period can be expressed as  $CV_{\text{repr}} = 100\beta_{\text{repr}}$ .

As described above, the MMQAP used the  $MAD_E$  to estimate  $\sigma_{ij}$ . This estimator not only is less statistically efficient than the  $Q_n$ , it assumes that the non-concordant values are symmetrically distributed about the median. The histograms for the U1 to U4 series of test samples shown in Fig. 3 are representative of the general tendency for the as-reported vitamin C probability distributions to favor smaller values over larger (left-skew). We therefore here estimate the  $\sigma_{ij}$  using the  $Q_n$  estimator [23, 24]. For the 168 data sets with at least 5 results and  $x_i > 0$ , the  $MAD_E$  estimates are on average 18% smaller than  $Q_n$  while the standard deviation is on average nearly twice  $Q_n$ .

The  $CV_{\text{repr}}$  improved over time. We hope that this improvement reflects experience gained through participation in the MMQAP but recognize that it may just reflect changes in the number and nature of the ILS participants as well as differences in the analytical methods used. Evidence from vitamin C studies coordinated by other organizations is difficult to compare since as-reported data are not available to non-participants and the data analysis methods are not described in detail. However, the INSTAND e.v. EQAS program makes vitamin C summary information available for 44 single-distribution lyophilized samples in their Oct-2014 to June-2020 programs [15]. The  $CV_{\text{repr}}$  estimated from their “Overview” information is a roughly constant 11% over this six-year period (ESM Fig. S1 reports our analysis of these data.) The Royal College of Pathologists of Australia Quality Assurance



**Fig. 3 Representative Probability Distributions for Test Materials.** The thick-line histograms are empirical unit-area probability distributions for the reported total ascorbic acid concentration, [TAA], of four test sample materials {designated U1, U2, U3, and U4} prepared at NIST in 2001. The light vertical lines mark the histogram bin boundaries. The light smooth curves represent normal distributions for the samples parameterized with the ILS estimates

Program (RCPAQAP) does not make summary information available, but a recent publication states that  $CV_{\text{repr}}$  for their 2010 initial vitamin C program was 13% and decreased to 7% in 2019 [5].

Except for the two 1995 ILS, the  $CV_{\text{repr}}$  for materials used as control samples is about the same as when used as test samples. This suggests that participants used the same measurement and reporting processes for the test and control materials regardless of having knowledge of the target [TAA] in the controls.

#### Preparation and evaluation of calibration solutions

A secondary goal of the MMQAP vitamin C program was assessment and enhancement of the participants' ability to

prepare metrologically traceable calibration solutions. ESM Fig. S2 shows that the probability density functions for the ratio between the measured and calculated [TAA] and the calculated  $E^{1\%}$  narrowed over time. The final, fully-developed protocols provided to participants for preparing and evaluating the calibrants are contained in Appendix E of [28].

#### Normalization

Figure 4 displays the impact of three normalization methods on the technical and reproducibility precision of the test and control sample results from the 2008 to 2015 period. These methods are all linear transformations:

$$x'_{ijk} = \frac{x_{ijk} - \alpha}{\beta}, \quad s'_{ijk} = \frac{s_{ijk}}{\beta}$$

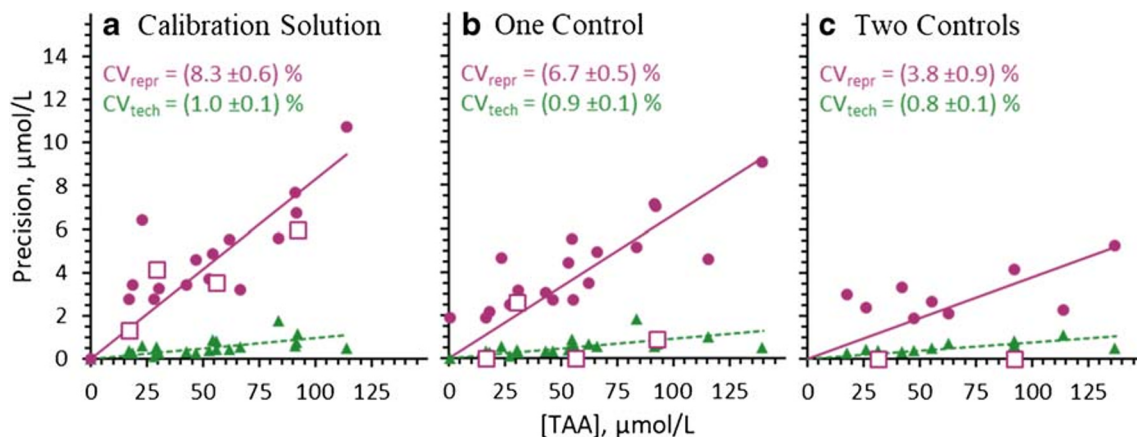
where  $i$  indexes the material,  $j$  the ILS,  $k$  the participant,  $\alpha$  is an offset parameter and  $\beta$  sets the scale.

Figure 4A reports the effect of normalizing to the calibration solutions prepared by each participant: Here,  $i$  is the nominal 50  $\mu\text{mol/L}$  solution, the offset  $\alpha$  parameter is set equal to zero, and the scaling  $\beta$  parameter is defined as the ratio between the measured and calculated values:

$$\beta = \frac{x_{ijk}(\text{measured})}{x_{ijk}(\text{calculated})}$$

This method only marginally improves the estimated  $CV_{\text{repr}}$  for this period, from  $(8.8 \pm 0.5)\%$  to  $(8.3 \pm 0.6)\%$ . Using regression to estimate both offset and scale parameters from the four participant-prepared calibration solutions is even less effective, leaving the  $CV_{\text{repr}}$  essentially unchanged at  $(8.7 \pm 0.6)\%$ .

Figure 4B reports the effect of normalizing to the reference value of a control material, setting the offset to zero and defining the scale parameter as



**Fig. 4 Test Material Within- and Between-Participant Precision Estimates Following Normalization.** Each panel reports within- and between-participant precision estimates for total ascorbic acid concentration, [TAA], for three approaches to normalizing between-participant

variability: A) scaled to the participant-prepared calibration solutions, B) scaled to the ratio of the control sample (or to the average ratio of two control samples), and C) interpolated between results for two control materials. Format as in Fig. 2

$$\beta = \frac{x_{1jk}(\text{measured})}{x_1(\text{reference})} \text{ or } \left( \frac{x_{1jk}(\text{measured})}{x_1(\text{reference})} + \frac{x_{2jk}(\text{measured})}{x_2(\text{reference})} \right) / 2$$

depending on whether one or two controls were distributed. This reduces the  $CV_{\text{repr}}$  to  $(6.7 \pm 0.5) \%$ .

Figure 4C reports the effect of linearly interpolating between two controls where the parameters are defined

$$\beta = \frac{x_{2jk}(\text{measured}) - x_{1jk}(\text{measured})}{x_2(\text{reference}) - x_1(\text{reference})}, \quad \alpha = x_{1jk}(\text{measured}) - \beta x_1(\text{reference}).$$

This reduces the  $CV_{\text{repr}}$  to  $(3.8 \pm 0.9) \%$ , although the available data are about as well described as having a [TAA]-independent standard deviation of  $(2.8 \pm 0.9) \mu\text{mol/L}$ .

The failure of calibrant normalization to improve reproducibility demonstrates that the between-participant variability is not driven by differences among the participants' routine calibration materials or strategies. The improvement provided by normalizing to one control sample identifies that whatever does drive the between-participant differences, its effect is systematic across the MPA-preserved samples. The considerable improvement provided by normalizing to two controls suggests that concentration-independent as well as concentration-dependent differences exist.

ESM Fig. S3 displays the changes in the probability density functions induced by normalization. The predominant effect is to reduce the width of the distributions, but without much reduction in the left-skew asymmetry.

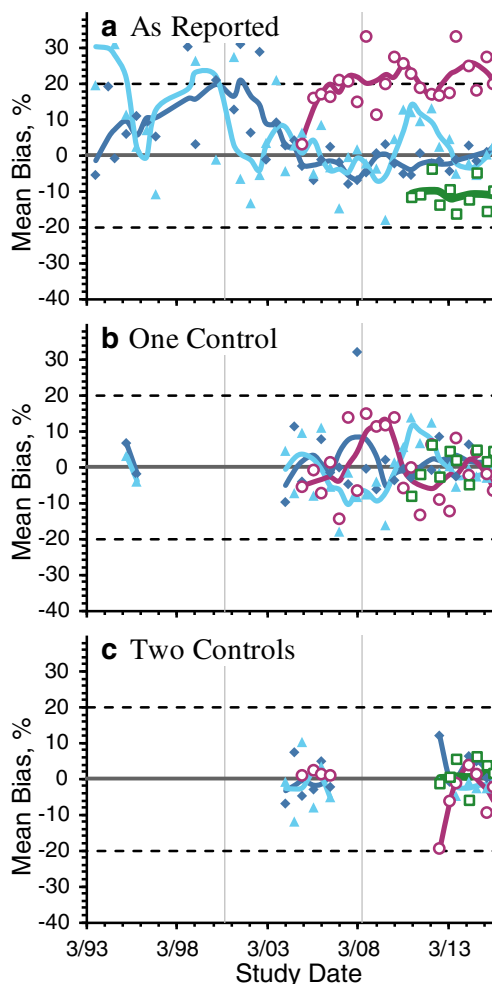
Figure 5 displays the effect of normalization on the performance of four exemplar participants, two of whom participated in all 40 MMQAP Vitamin C ILS. For each ILS, a relative mean bias for each participant can be estimated as:

$$B_{jk} = 100 \left( \frac{\sum_i^{n_j} x_{ijk} - x_i}{\sigma_i} \right) / n_j$$

where  $n_j$  is the number of test samples distributed in the  $j$ th ILS. Assuming that  $\sigma_i$  is equal to  $0.1 x_i$  (i.e.,  $CV_{\text{repr}}$  is 10% for all test samples in all ILS) facilitates visualization, although it underestimates the variability during 1993 to 2001 and slightly overestimates it for 2008 to 2015.

Figure 5A displays the mean bias for the as-reported results. The  $B_{jk}$  for the two long-term participants stabilized around zero bias in the early 2000s, a few years after the number of test samples distributed per ILS was increased. We believe that the two-year long increase in bias starting in 2011 for one of these participants reflects changes in analytical methodology. However, the  $B_{jk}$  bias for a participant who joined the program in 2005 is fairly consistently +20% and that for a participant who joined in 2011 is consistently -10%.

While control samples became a routine component of the program only in 2004 (see Table 1), Fig. 5B reveals that, when control data is available, scaling the sample results to one control greatly reduces the average bias. Fig. 5C demonstrates that projecting onto the line defined by two controls typically reduces the variability of the bias estimates. Note, however, that normalization can amplify bias if the participants' control sample results aren't representative of the test sample results (e.g., because of matrix miss-match, analysis at different times or with different reagents, or measurement error).



**Fig. 5 Participant Performance Over Time and Normalization Approaches.** Each panel reports mean relative test sample measurement bias for four exemplar laboratories that participated in the final MMQAP Vitamin C interlaboratory studies. Each symbol denotes the mean bias in one ILS for one participant, calculated from results: A) as reported, B) scaled to the ratio of the control sample (or to the average ratio of two control samples), and C) projected onto the line defined by two control materials. The curves are quadratic-smoothed representations of the performance of these four participants over time. Dashed horizontal lines bound an approximate 95% level of confidence interval on the relative bias, assuming a constant coefficient of variation of 10% for all test materials in all studies. The vertical lines mark the seven-year intervals. The open circles and squares identify the participants whose mean biases are much reduced by normalization

## Conclusion

NIST's Micronutrient Measurement Quality Assurance Program (MMQAP) for vitamin C established that the total ascorbic acid concentration ([TAA]) of human serum in a 1 + 1 volumetric combination with an aqueous solution containing a mass fraction of 10% metaphosphoric acid (MPA) is stable for at least 20 years when stored at  $-80^{\circ}\text{C}$ .

Calibration with participant-prepared and evaluated AA aqueous calibration solutions did not improve measurement comparability whereas calibration with one or two MPA-stabilized human serum control materials reduced  $\text{CV}_{\text{repr}}$  from 9% to  $\approx 7\%$  and  $\approx 4\%$ , respectively. This indicates that the various [TAA] measurement methods used by the MMQAP participants are differently (but self-consistently) sensitive to components of the serum-based samples that are not present in the aqueous calibrants.

The typical within-participant measurement coefficient of variation ( $\text{CV}_{\text{tech}}$ ) improved from  $\approx 2\%$  to less than 1% over the 40 interlaboratory studies from 1993 through 2015. The between-participant relative reproducibility precision ( $\text{CV}_{\text{repr}}$ ) improved from  $\approx 16\%$  to  $\approx 9\%$  over this period. We believe this improvement supports the contention that reference materials such as SRM 970 and regular participation in quality assurance and/or proficiency test programs like the MMQAP, RCPAQAP, and INSTAND help improve among-laboratory measurement comparability.

While the availability of certified reference materials and participation in interlaboratory comparisons may contribute to controlling and potentially improving the reproducibility of [TAA] measurements, normalization to serum-based control materials having a matrix that is well matched to that of the samples of interest can ensure it. Multicenter studies of [TAA] in preserved plasma or serum that require the best achievable measurement comparability should consider calibration with matrix-matched control materials.

**Acknowledgments** We thank all of the analysts who participated in the MMQAP studies and workshops. Their participation, enthusiastic support, and critical feedback were critical to the MMQAP's success and to establishing NIST's new Health Assessment Measurements Quality Assurance Program (HAMQAP).

**Availability of data and materials** All MMQAP data, study design, and final reports are available in the NISTIR 7880 series.

**Funding** All MMQAP activities through FY 1997 were supported in part by the Division of Cancer Prevention and Control, National Cancer Institute, National Institutes of Health. The MMQAP Vitamin C studies from 1998 to 2015 were supported in part by participant fees.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest nor competing interests.

The anonymous donor human plasma pools used to prepare the MMQAP Vitamin C test and control materials were obtained from commercial sources after appropriate human subjects research determinations by NIST.

**Consent to participate** Not Applicable.

**Consent for publication** Not Applicable.

**Disclaimer** Not Applicable

## References

1. Granger M, Eck P. Dietary vitamin C in human health. *Adv Food Nutr Res.* 2018;83:281–310. <https://doi.org/10.1016/bs.afnr.2017.11.006>.
2. Shenoy N, Creagan E, Witzig T, Levine M. Ascorbic acid in Cancer treatment: let the Phoenix Fly. *Cancer Cell.* 2018;34(5):700–6. <https://doi.org/10.1016/j.ccell.2018.07.014>.
3. Bates CJ. Plasma vitamin C assays: a European experience. EC FLAIR concerted action no. 10: micronutrient measurement, absorption and status. *Int J Vitam Nutr Res.* 1994;64(4):283–7.
4. Armbruster D, Miller RR. The joint Committee for Traceability in laboratory medicine (JCTLM): a global approach to promote the standardisation of clinical laboratory test results. *Clin Biochem Rev.* 2007;28(3):105–13.
5. Collie JTB, Greaves RF, Jones OAH, Eastwood G, Bellomo R. Vitamin C measurement in critical illness: challenges, methodologies and quality improvements. *Clin Chem Lab Med.* 2020;58(4):460–70. <https://doi.org/10.1515/cclm-2019-0912>.
6. Duewer DL, Thomas JB, Kline MC, MacCrehan WA, Schaffer R, Sharpless KE, et al. NIST/NCI micronutrients measurement quality assurance program: measurement repeatabilities and reproducibilities for fat soluble vitamin related compounds in human sera. *Anal Chem.* 1997;69(7):1406–13. <https://doi.org/10.1021/ac9607727>.
7. Margolis SA, Davis TP. Stabilization of ascorbic acid in human plasma, and its liquid-chromatographic measurement. *Clin Chem.* 1988;34(11):2217–23. <https://doi.org/10.1093/clinchem/34.11.2217>.
8. Margolis SA, Paule RC, Ziegler RG. Ascorbic and dehydroascorbic acids measured in plasma preserved with dithiothreitol or metaphosphoric acid. *Clin Chem.* 1990;36(10):1750–5. <https://doi.org/10.1093/clinchem/36.10.1750>.
9. Margolis SA, Ziegler RG, Helzlsouer KJ. Ascorbic and dehydroascorbic acid measurement in human serum and plasma. *Am J Clin Nutr.* 1991;54(6 Suppl):1315S–8S. <https://doi.org/10.1093/ajcn/54.6.1315s>.
10. Craft NE, Duewer DL, Kline MC, Paule RC, Ruegg FC, Schaffer R, Schonberger E, Thomas JB (2018) NIST Micronutrients Measurement Quality Assurance Program Spring and Fall 1986 Comparability Studies: Results for Round Robins VII and VIII Fat-Soluble Vitamins and Carotenoids in Human Serum. NIST



- Internal/Interagency Report (NISTIR) 7880–38, Gaithersburg, MD. <https://doi.org/10.6028/NIST.IR.7880-38>
11. Duewer DL, Kline MC, Margolis SA, May WE, Sharpless KE, Thomas JB (2014) NIST Micronutrients Measurement Quality Assurance Program Winter, Spring, and Fall 1993 Comparability Studies: Results for Round Robins XXVII, XXVIII, and XXIX Fat-Soluble Vitamins and Carotenoids in Human Serum and Round Robin 4 Ascorbic Acid in Human Serum. NIST Internal/Interagency Report (NISTIR) 7880–31, Gaithersburg, MD. <https://doi.org/10.6028/NIST.IR.7880-31>
  12. Margolis SA, Schapira RM. Liquid chromatographic measurement of L-ascorbic acid and D-ascorbic acid in biological samples. *J Chromatogr B Biomed Sci Appl.* 1997;690(1–2):25–33. [https://doi.org/10.1016/s0378-4347\(96\)00401-x](https://doi.org/10.1016/s0378-4347(96)00401-x).
  13. Margolis SA, Duewer DL. Measurement of ascorbic acid in human plasma and serum: stability, intralaboratory repeatability, and interlaboratory reproducibility. *Clin Chem.* 1996;42(8 Pt 1):1257–62. <https://doi.org/10.1093/clinchem/42.8.1257>.
  14. Margolis SA, Vangel M, Duewer DL. Certification of standard reference material 970, ascorbic acid in serum, and analysis of associated interlaboratory bias in the measurement process. *Clin Chem.* 2003;49(3):463–9. <https://doi.org/10.1373/49.3.463>.
  15. INSTAND e.v., EQAS Online, [https://www.instand-ev.de/no\\_cache/en/eqas-online/service-for-eqa-tests/](https://www.instand-ev.de/no_cache/en/eqas-online/service-for-eqa-tests/) (Accessed 8-Sep-2020).
  16. Rossi B, Tittone F, Palleschi S. Setup and validation of a convenient sampling procedure to promptly and effectively stabilize vitamin C in blood and plasma specimens stored at routine temperatures. *Anal Bioanal Chem.* 2016;408(17):4723–31. <https://doi.org/10.1007/s00216-016-9560-6>.
  17. Clark ZD, Frank EL. Development and implementation of an HPLC-ECD method for analysis of vitamin C in plasma using single column and automatic alternating dual column regeneration. *Pract Lab Med.* 2016;6:25–37. <https://doi.org/10.1016/j.plabm.2016.09.001>.
  18. Robitaille L, Hoffer LJ. A simple method for plasma total vitamin C analysis suitable for routine clinical laboratory use. *Nutr J.* 2016;15:40. <https://doi.org/10.1186/s12937-016-0158-9>.
  19. Margolis SA (1995) Ascorbic and Dehydroascorbic Acids Analysis. In: *Methods of Analysis of Cancer Chemopreventive Agents in Human Serum.* Thomas JB, Sharpless KE, eds. NIST Special Publication (NIST SP) 874, Gaithersburg, MD. <https://doi.org/10.6028/NIST.SP.874>
  20. Rocke DM. Robust statistical analysis of Interlaboratory studies. *Biometrika.* 1983;70(2):421–31. <https://doi.org/10.1093/biomet/70.2.421>.
  21. Andrews DF, Bickel PJ, Hampel FR, Huber PJ, Rogers WH, Tukey JW (1972) *Robust Estimates of Location.* Princeton University Press
  22. Hampel FR. The Influence Curve and its Role in Robust Estimation. *J. Am Stat Assoc.* 1974;69:383–93. <https://doi.org/10.1080/01621459.1974.10482962>.
  23. Rousseeuw PJ, Croux C. Alternatives to median absolute deviation. *J Am Stat Assoc.* 1993;88:1273–83. <https://doi.org/10.1080/01621459.1993.10476408>.
  24. Wilrich P-T. Robust estimates of the theoretical standard deviation to be used in interlaboratory precision experiments. *Accred Qual Assur.* 2007;12:231–40. <https://doi.org/10.1007/s00769-006-0240-7>.
  25. Esteve MJ, Farré R, Frigola A, Garcia-Cantabella JM. Determination of ascorbic and dehydroascorbic acids in blood plasma and serum by liquid chromatography. *J Chromatogr B Biomed Sci Appl.* 1997;688(2):345–9. [https://doi.org/10.1016/s0378-4347\(96\)00248-4](https://doi.org/10.1016/s0378-4347(96)00248-4).
  26. Ihara H, Hashizume N, Matsubayashi T, Futaki K, Yoshida M, Sagawa N, et al. Stability of fat-soluble and water-soluble vitamins in artificially prepared, vitamin-enriched, lyophilized serum. *J Clin Lab Anal.* 2004;18(4):240–6. <https://doi.org/10.1002/jcla.20030>.
  27. Kand'ár R, Záková P. Determination of ascorbic acid in human plasma with a view to stability using HPLC with UV detection. *J Sep Sci.* 2008;31(20):3503–8. <https://doi.org/10.1002/jssc.200800303>.
  28. Duewer DL, Thomas JB (2013) NIST Micronutrients Measurement Quality Assurance Program Summer 2012 Comparability Studies. Results for Round Robin LXXII Fat-Soluble Vitamins and Carotenoids in Human Serum and Round Robin 37 Ascorbic Acid in Human Serum. NIST Internal Report (NISTIR) 7880–1. NIST, Gaithersburg, MD. <https://doi.org/10.6028/NIST.IR.7880-1>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**David L. Duewer** is a research chemist/data analyst with 11 years' experience in the agricultural chemical industry before joining NIST in 1991. He is an advocate for collaborative R&D at the interfaces of chemistry, biology, and measurement science, with particular interest in the design and evaluation of interlaboratory studies.



**Jeanice Brown Thomas** recently retired from NIST after working almost 44 years as a research chemist. During her career, she was involved in organic analytical research with emphasis on environmental, clinical and food analysis. She was the Program Manager for the MMQP for over 20 years.



**Katherine E. Sharpless** has been a research chemist at NIST for 30 years. She participated in the fat-soluble vitamin arm of the MMQAP, assisted with characterization of associated clinical SRMs, and coordinated NIST's development of food and dietary supplement SRMs. She is currently NIST's Open Access Officer, developing policies and procedures for making results of federally funded research available to the public.



**Sam A. Margolis** is a retired NIST research analytical biochemist. He developed the methodology for analyzing, stabilizing, and preparing the standards for the collaborative studies described in this manuscript. He managed the distribution and initial analysis of these studies from their inception through 2002.