Anal Bioanal Chem (2013) 405:4437–4441
DOI 10.1007/s00216-013-6864-7

NOTE

The development and implementation of quality assurance programs to support nutritional measurements

L. C. Sander · M. Bedner · D. L. Duewer · K. A. Lippa · M. M. Phillips · K. W. Phinney · C. A. Rimmer · M. M. Schantz · K. E. Sharpless · S. S. -C. Tai · J. B. Thomas · S. A. Wise · L. J. Wood · J. M. Betz · P. M. Coates

Received: 21 November 2012 / Revised: 14 February 2013 / Accepted: 21 February 2013 / Published online: 4 April 2013

Abstract The National Institute of Standards and Technology administers quality assurance programs devoted to improving measurements of nutrients and related metabolites in foods, dietary supplements, and serum and plasma samples. These programs have been developed in collaboration with the National Institutes of Health to assist measurement communities in their efforts to achieve accurate results that are comparable among different laboratories and over time. Targeted analytes include micronutrients, botanical markers, nutritional elements, contaminants, fatty acids, and vitamin D metabolites.

Keywords Quality assurance · Accuracy · Precision · Comparability · Concordance · Nutrients · Dietary supplements · Vitamin D metabolites · Fatty acids

Published in the topical collection *Functional Foods and Dietary Supplements* with guest editors Melissa M. Phillips and Catherine A. Rimmer.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-013-6864-7) contains supplementary material, which is available to authorized users.

L. C. Sander (✉) · M. Bedner · D. L. Duewer · K. A. Lippa · M. M. Phillips · K. W. Phinney · C. A. Rimmer · M. M. Schantz · K. E. Sharpless · S. S. -C. Tai · J. B. Thomas · S. A. Wise · L. J. Wood
Chemical Sciences Division, National Institute of Standards and Technology, 100 Bureau Drive, MS 8392, Gaithersburg, MD 20899, USA
e-mail: lane.sander@nist.gov

J. M. Betz · P. M. Coates
Office of Dietary Supplements, National Institutes of Health, 6100 Executive Blvd., Rm. 3B01, Bethesda, MD 20892, USA

Introduction

The National Institute of Standards and Technology (NIST) provides a variety of resources to assist measurement communities in achieving measurement quality objectives. These resources include Standard Reference Materials (SRMs), analytical methods, and quality assurance programs. Since the mid-1980s, NIST has administered eight quality assurance (QA) programs, each of which has been designed for different measurement communities [1]. Laboratory participation in these programs is voluntary, although funding agencies may require participation as a condition of issued grants and contracts. The QA programs are not proficiency testing programs, and criteria are not established for acceptable performance. Rather, the programs can be considered performance improvement exercises with a primary goal of achieving accurate (precise and unbiased) results that are comparable across a measurement community over time. This report highlights the achievements of NIST QA programs to support measurement comparability for nutrients and related metabolites and provides examples of how programs are tailored to the needs of diverse measurement communities.

Micronutrients Measurement Quality Assurance Program

The NIST/NCI Micronutrients Measurement Quality Assurance Program (MMQAP) was established in 1984 to assess measurement quality metrics related to the determination of specific fat-soluble vitamins and carotenoids in human serum/plasma that were thought to protect against various cancers. These compounds initially included retinol, α-tocopherol, and total β-carotene, but in later years, the program

expanded to include *cis* and *trans* isomers of β-carotene, α-carotene, lutein, zeaxanthin, β-cryptoxanthin, lycopene, δ-tocopherol, γ-tocopherol, 25-hydroxyvitamin D, vitamin K, and coenzyme Q10. Some of the early participants expressed interest in the determination of vitamin C in serum, and in the mid-1990s, separate exercises were initiated to support measurements of this micronutrient. Two exercises are administered each year for both of the vitamin C and fat-soluble vitamin components. The data are evaluated, and summary statistics are generated; participants are also provided with individualized reports of their performance.

Because the MMQAP is now a mature program with data that span nearly 30 years, retrospective analysis of performance indicators can provide insight into challenges associated with individual analytes and the effectiveness of the program to produce improved measurement comparability [2–4].

Plots showing the concordance and apparent precision for six nutritional analytes, measured from 1984 to 2012, are provided in Fig. 1. Dramatically improved performance is evident during the initial 5 years of the program. In subsequent years, these performance indicators have fluctuated within narrow bounds that are analyte dependent. For example, for the past 10 years, the apparent precision for total retinol has remained less than 5 % relative standard deviation (RSD) whereas the apparent precision for total α-carotene has remained greater than 10 % RSD. Similar trends can be observed for concordance. Somewhat unexpected sources of measurement variability have been identified. These include misidentification of components and ambiguity in definitions for analytes reported as the sum of components. Expected sources of variability include calibration errors, purity corrections, chromatographic interferences, and instrumental biases.

Over the course of administering the MMQAP, statistical approaches have evolved to help analysts better understand their performance and to utilize this information in improving their measurement capabilities. In early exercises, measurement results were summarized based on calculation of the mean and standard deviation to characterize consensus values and population dispersion. This approach is sensitive to the presence of outlier data that may result from method biases, calculation errors, and other blunders. To address these limitations, a different approach based on the use of robust statistics has evolved. Assuming that the majority of the results for a given analyte in a given sample can be described as normally distributed, the median provides a

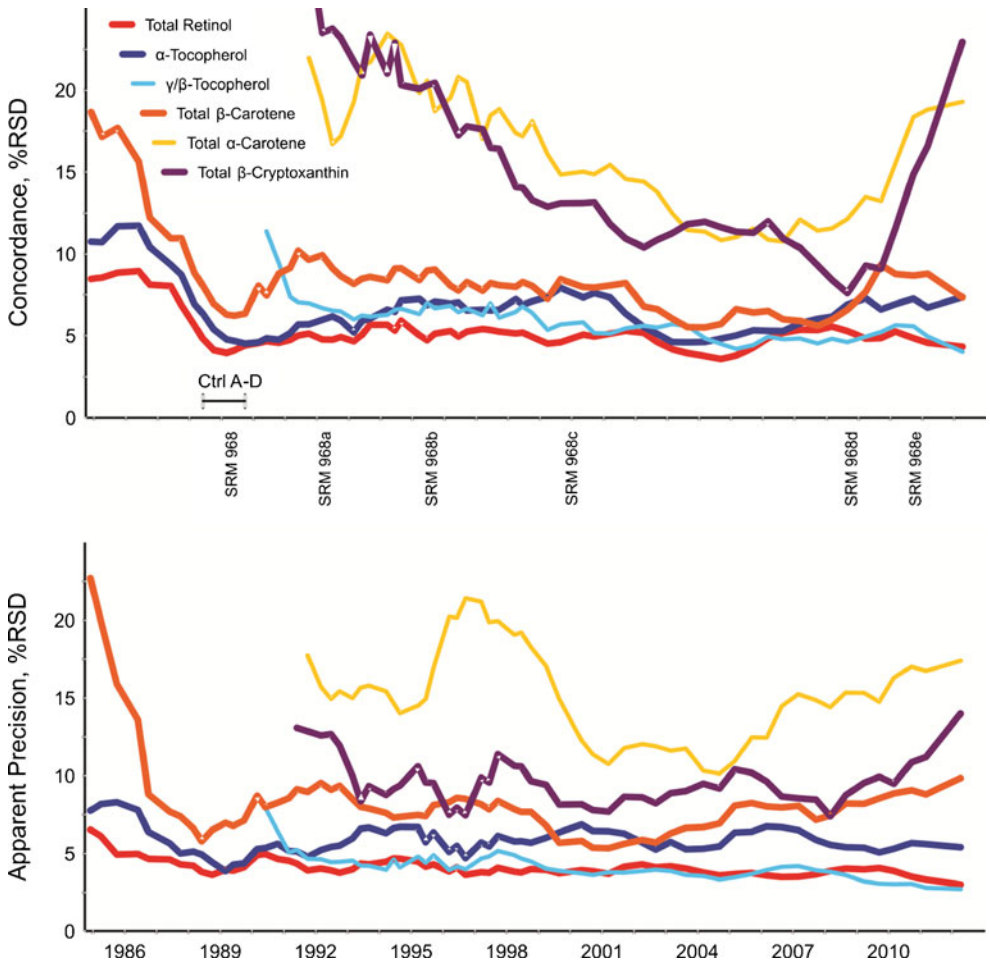

Fig. 1 Interlaboratory composite concordance and apparent precision for selected micronutrients, plotted as a function of individual MMQAP exercises spanning the years 1984 to 2012. The use of various SRMs as control materials is indicated in the *upper plot*

robust estimate of the consensus value. The standard deviation for the majority population was initially estimated from the range of the central 50 % of the results and more recently from the median of the absolute deviations from the median. Use of the median and either of these robust estimates of the standard deviation is more objective and is less sensitive to outlier data.

Other statistical indicators and graphical tools have been developed to characterize measurement performance of individual participants over time (see Electronic supplementary material, Fig. S1) [2]. In the context of a performance improvement program, these indicators provide useful measures of performance that are based on the capabilities of the community (i.e., the exercise participants) rather than the capabilities of a reference laboratory. Measurement *comparability* provides an indication of agreement with the consensus value for a given analyte and exercise and is defined by the difference between the analyst's measurement and the consensus value. Measurement *concordance* has been defined as the average comparability for a given participant across all similar-matrix samples in a given interlaboratory exercise. In a similar fashion, the *apparent precision* has been defined as the standard deviation of the participant's comparabilities for all samples in an exercise.

Dietary Supplement Laboratory Quality Assurance Program

The Dietary Supplement Health and Education Act (DSHEA, 1994) amended the Food, Drug and Cosmetic Act to establish dietary supplements as a new regulatory category distinct from food and drugs. This law imposed certain requirements on manufacturers to characterize the identity, purity, and composition of dietary supplements as well as the ingredients used in their manufacture. The Dietary Supplement Laboratory Quality Assurance Program (DSQAP) was established as a collaborative effort between NIST and NIH-Office of Dietary Supplements (ODS) to support the measurement infrastructure of the dietary supplement community [5, 6]. The primary focus of the DSQAP is measurement performance improvement, and the program offers tools that can assist participants in demonstrating compliance with current Good Manufacturing Practices established by the FDA.

Unlike many other quality assurance programs, the DSQAP functions as a survey program. Different measurement challenges are addressed in each exercise, which include a broad spectrum of sample types and analyte classes. Because each exercise targets different analytes and sample matrices, individual studies reflect short-term sources of measurement variability. Exercises include up to six individual studies comprised of measurement of phytochemicals, fatty acids, fat-soluble vitamins, water-soluble vitamins, contaminants, and nutritional elements. Control materials are distributed with the unknown samples, and when possible, SRMs (or other certified reference materials) are utilized for this purpose. The control samples can be used to verify that the analytical methods are performing as expected and to assess sources of variability across a given study.

An example of a summary report for the DSQAP is shown in Fig. 2. Analyte levels are ordered from smallest to largest and are identified with the laboratory code (Fig. 2a). A target value and target range is provided based on NIST data, and the community consensus median and consensus variability are shown for the study participants. This depiction of the data provides a way for participants to easily assess their performance relative to the other participants in the study, and relative to the target values. The

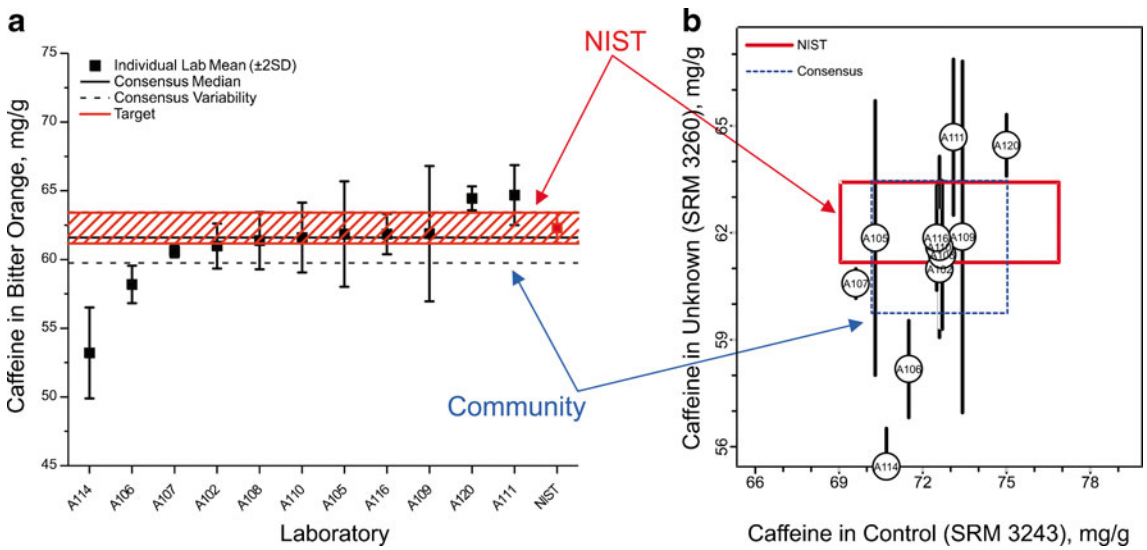

Fig. 2 Summary report for caffeine, DSQAP Exercise A. *Error bars* represent ±2 standard deviations for individual labs

Youden plot (Fig. 2b) facilitates comparisons of data reported for the unknown with data reported for the control. Boxes represent the target and consensus variability ranges for unknown and control data. The sources of measurement biases are sometimes apparent from these plots. For example, if a laboratory reports values for both control and unknown samples that are consistently low or high compared with target values, this may indicate calibration problems. However, if biases are apparent for only the control or the unknown sample, this could suggest that the differences are due to sample matrix effects. Differences between the community consensus and the target value could result from methodological differences, such as might occur for incomplete extraction recovery by participants.

Lessons learned from past exercises provide useful guides for measurement improvement. When specific challenges are identified, the root cause is often relevant to other applications and analytical methods. For example, extraction efficiency has been demonstrated to vary with the type of sample matrix under study. Dried botanicals offer the biggest challenge, followed by finished products, extracts, and solutions. Experience has shown that sample extraction is one of the likely sources of measurement bias for natural-matrix materials. Other common sources of bias include calibration errors, lack of method specificity, handling losses, analyte degradation, and blunders (e.g., calculation errors, mislabeled samples, transposed numbers, incorrect units). The unpredictability of blunders makes them difficult to eliminate; however, the routine use of control materials can assist in their detection.

Vitamin D Metabolites Quality Assurance Program

Vitamin D is essential to bone metabolism and calcium homeostasis, and deficiencies have been linked to increased risks in the development of cancer, diabetes, and other diseases. Vitamin D is obtained from the diet (vitamin D_2 and vitamin D_3) and through exposure to sunlight on the skin (vitamin D_3). Both forms are converted in the liver to respective 25-hydroxyvitamin D (25(OH)D) metabolites. These metabolites are further converted to 1,25-dihydroxyvitamin D ($1,25(OH)_2D$), the active form of vitamin D. Because 25(OH)D has a relatively long half-life of 2 to 3 weeks, the concentration of 25(OH)D in serum is the preferred indicator of vitamin D status [7]. Clinical methods used in the assessment of vitamin D status have been subject to a high level of scrutiny as a consequence of recent high-profile litigation and adverse reports in the public media. Evidence-based guidelines for vitamin D status depend on the use of precise and unbiased assays for vitamin D metabolite(s) that are comparable among different laboratories over time.

The NIST/NIH Vitamin D Metabolites Quality Assurance Program (VitDQAP) was established in 2009 with the principle objective of improving the accuracy of measurements of 25(OH)D in serum and plasma samples. As with the other QA programs administered by NIST, participants use methods of their choice to measure 25(OH)D in the samples. Participants who use commercial assay kits based on immunoassays report 25(OH)D as a total value, whereas participants who use chromatographic methods report $25(OH)D_2$ and $25(OH)D_3$ separately. The principle objective of the VitDQAP is harmonizing measurements made by the vitamin D measurement community through the reduction of method biases.

Representative results from a recent VitDQAP exercise are illustrated in Fig. 3. To assess potential method-specific biases, levels of 25(OH)D have been plotted individually for each method (levels for chromatographic methods represent the sum of $25(OH)D_2$ and $25(OH)D_3$). Immunoassays generally provided slightly higher results in this study compared with chromatographic-based techniques, and the all-lab median was higher than the reference laboratory value. Future VitDQAP exercises will utilize samples that are high in $25(OH)D_2$ and 3-epi-$25(OH)D_3$ as challenge samples to assess possible measurement biases that may result from atypical patient samples.

A key component in achieving measurement comparability is the use of reference materials as measurement controls and for method calibration. SRM 972 and its replacement SRM 972a Vitamin D Metabolites in Human Serum were developed for use as control materials for 25(OH)D assays

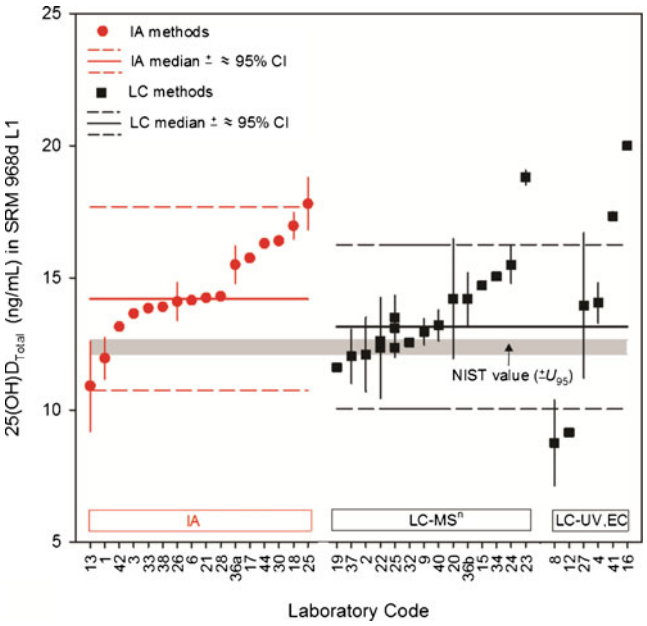

Fig. 3 Results for $25(OH)D_{Total}$ in SRM 968d Level 1 from the VitDQAP Summer 2010 exercise. Immunoassay methods include radioimmunoassay, enzyme immunoassay, and chemiluminescent immunoassay. *Error bars* represent ±1 standard deviation for individual laboratory mean results. The "*≈95 % CI*" represents the approximate 95 % confidence interval

[7]. SRM 972 (now depleted) consisted of four serum pools with differing levels of 25(OH)D_2, 25(OH)D_3, and 3-epi-25(OH)D_3. Levels of some metabolites were adjusted through dilution and spiking procedures. Based on comments from users of this SRM, the replacement material was formulated using pools of prescreened sera to achieve the desired metabolite levels (only the 3-epi-25(OH)D_3 metabolite was fortified in one pool by spiking).

Interlaboratory Analytical Comparison Study of Total Fatty Acid Concentrations in Human Serum

Risk factors for the development of cardiovascular disease have long been associated with consumption of dietary fat; however, in recent years, emphasis has moved away from measures of total fat to characterization of individual fatty acids. Sources of dietary fat are characterized by levels of saturated fatty acids, mono- and polyunsaturated fatty acids, and *trans* fatty acids. Serum lipids are likewise characterized by their constituent fatty acids, and the resulting fatty acid profiles may provide indications of risks for the development of atherosclerosis and other cardiovascular diseases [8].

The most common methods for the determination of fatty acids employ a derivatization step to form the methyl esters (FAMEs) followed by analysis using gas chromatography with flame ionization detection or GC/MS. Liquid chromatography is also used for fatty acid analysis. As with GC-based techniques, derivatization is required; however, this step is included for purposes of detection rather than separation. A variety of derivatizing reagents are available to label the fatty acids with fluorophores or chromophores.

A quality assurance program for improving the accuracy of measurements of fatty acids in human serum has recently been established by NIST, NIH-ODS, and CDC. Samples distributed during the summer of 2012 included three levels of a candidate SRM. These samples are human serum materials that are representative of (1) donors who have not taken fish or flaxseed oil supplements for 1 month prior to collection, (2) donors who have taken flaxseed oil supplements for a minimum of 1 month prior to collection, and (3) donors who have taken fish oil supplements for a minimum of 1 month prior to collection. In addition to these unknowns, SRM 1950 Metabolites in Human Plasma was distributed to participants for use as a control. Twenty-four fatty acids were identified for determination in the initial exercise (see Table S1, Electronic supplementary material).

Conclusions

Quality assurance programs have been developed to assist measurement communities achieve measurement quality objectives. The goals of the programs differ slightly depending on the needs of the participants, but improved measurement concordance and, where possible, accuracy remain common objectives.

References

1. Quality Assurance Programs (2012). http://www.nist.gov/mml/csd/qaps.cfm, NIST. Accessed 13 March 2013
2. Duewer DL, Kline M, Sharpless KE, Brown Thomas JM, Gary KT (1999) Anal Chem 71:1870–1878
3. Duewer DL, Kline MC, Sharpless KE, Thomas JB (2000) Anal Chem 72:4163–4170
4. Duewer DL, Kline M, Sharpless KE, Brown Thomas JM, Stacewicz-Sapuntzakis M, Sowell AL (2000) Anal Chem 72:3611–3619
5. Phillips MM, Rimmer CA, Wood LJ, Lippa KA, Sharpless KE, Duewer DL, Sander LC, Betz JM (2011) J AOAC Int 94:803–814
6. Betz JM, Fisher KD, Saldanha LG, Coates PM (2007) Anal Bioanal Chem 389:19–25
7. Phinney KW, Bedner M, Tai SSC, Vamathevan VV, Sander LC, Sharpless KE, Wise SA (2012) Anal Chem 84:956–962
8. Bhupathiraju SN, Tucker KL (2011) Clin Chim Acta 412:1493–1514