Published in final edited form as:

Clin Chim Acta. 2013 November 15; 426: . doi:10.1016/j.cca.2013.08.012.

An Assessment of 25-Hydroxyvitamin D Measurements in Comparability Studies Conducted by the Vitamin D Metabolites Quality Assurance Program

Mary Bedner*, Katrice A. Lippa*, and Susan S.-C. Tai

National Institute of Standards and Technology Material Measurement Laboratory Chemical Sciences Division

Abstract

Background—The National Institute of Standards and Technology (NIST), in collaboration with the National Institutes of Health Office of Dietary Supplements, established the first accuracy-based program for improving the comparability of vitamin D metabolite measurements, the Vitamin D Metabolites Quality Assurance Program.

Methods—Study samples were comprised of human serum or plasma Standard Reference Materials (SRMs) with 25-hydroxyvitamin D values that were determined at NIST. Participants evaluated the materials using immunoassay (IA), liquid chromatography (LC) with mass spectrometric detection, and LC with ultraviolet absorbance detection. NIST evaluated the results for concordance within the participant community as well as trueness relative to the NIST value.

Results—For the study materials that contain mostly 25-hydroxyvitamin D₃ (25(OH)D₃), the coefficient of variation (CV) for the participant results was consistently in the range from 7% to 19%, and the median values were biased high relative to the NIST values. However, for materials that contain significant concentrations of both 25-hydroxyvitamin D₂ (25(OH)D₂) and 25(OH)D₃, the median IA results were biased lower than both the LC and the NIST values, and the CV was as high as 28%. The first interlaboratory comparison results for SRM 972a Vitamin D Metabolites in Human Serum are also reported.

Conclusions—Relatively large within-lab and between-lab variability hinders conclusive assessments of bias and accuracy.

Keywords

25-hydroxyvitamin D; interlaboratory comparison; quality assurance; immunoassay; liquid chromatography; mass spectrometry

1. Introduction

The concentration of total 25-hydroxyvitamin D (25(OH)D_{Total}) in human serum is used clinically to assess vitamin D status, where 25(OH)D_{Total} represents the sum of the individual metabolites 25-hydroxyvitamin D₂ (25(OH)D₂) and 25-hydroxyvitamin D₃

*Corresponding author contact information: vitdqap@nist.gov ; 301-974-4142 (voice); 301-977-0685 (fax), NIST, VitDQAP, 100 Bureau Drive, Gaithersburg, MD 20899-8392.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

(25(OH)D₃). There are several different analytical techniques that can be used to measure 25(OH)D_{Total} including immunoassay (IA) platform methods and liquid chromatography (LC) based assays. When assay performance has been compared on a large number of serum study samples, the individual analytical techniques often provide dissimilar results for 25(OH)D_{Total} [1–8].

Proficiency testing programs have been established to assess these assay differences and to improve the accuracy of 25(OH)D_{Total} measurements in the clinical community. The most prominent is the Vitamin D External Quality Assessment Scheme (DEQAS), which has conducted interlaboratory studies for over 20 y. Until recently, laboratory performance was judged based on the percentage bias from the All-Laboratory Trimmed Mean (ALTM) for each of five serum samples that are distributed quarterly. Since the ALTM includes results for all participant methods, it may not represent the “true” value and may change over time as the methods evolve. To address this issue, DEQAS is presently being converted to an accuracy-based program for selected vitamin D metabolites.

The College of American Pathologists (CAP) also has an established Accuracy-Based Vitamin D (ABVD) survey. Three samples of human serum are distributed twice per year, and participant results are compared to target values obtained from the Centers for Disease Control and Prevention using isotope dilution LC with tandem mass spectrometric detection (LC-MS/MS). Participants pass if their results fall within a defined percentage of the target values (information from the CAP website [9]).

In 2009, the National Institute of Standards and Technology (NIST), in collaboration with the National Institutes of Health Office of Dietary Supplements (NIH-ODS) established the first accuracy-based program for improving the comparability of laboratory measurements for 25(OH)D_{Total}, the NIST/NIH Vitamin D Metabolites Quality Assurance Program (VitDQAP). In addition to the VitDQAP, NIST and NIH-ODS collaboratively support vitamin D metabolite metrology through Standard Reference Materials (SRMs) including SRM 972 Vitamin D in Human Serum, which was the first certified reference material for vitamin D metabolites and was made available for sale in 2009 [10]. SRM 972 consisted of four levels that reflect the different analytical challenges facing laboratories that measure vitamin D: “normal” and “deficient” levels of 25(OH)D₃, as well as a level with each fortified 25(OH)D₂ and fortified 3-epi-25-hydroxyvitaminD₃ (3-epi-25(OH)D₃). The demand for SRM 972 was greater than anticipated, and the supplies were depleted in a little over a year. The replacement, SRM 972a Vitamin D Metabolites in Human Serum, was released in early 2013. SRM 972a uses only human serum with native levels of 25(OH)D₂ and 25(OH)D₃ (Level 4 contains fortified 3-epi-25(OH)D₃), potentially addressing technique-specific biases that were observed with SRM 972. Also in 2009, NIST released SRM 2972, 25-Hydroxyvitamin D₂ and D₃ Calibration Solutions, which was designed to improve standardization of vitamin D assays, particularly LC-based methods.

The VitDQAP has conducted comparability studies twice a year, typically in the summer and winter. To date, there have been six completed comparability studies of the VitDQAP, identified here as Study 1 (Winter 2010) through Study 6 (Summer 2012). During the first three years of the program, the number of participants has grown from 16 to 85, representing 14 different countries and all clinically-relevant measurement sectors: government, academia, hospitals, and testing laboratories. To date, any laboratory that has an established method to measure 25(OH)D_{Total} and has requested enrollment has been able to participate in the program. VitDQAP currently does not charge a fee for participation, and participants are assigned code numbers that are known only to them and to NIST.

For each comparability study, control and study samples were distributed to participants. SRM 2972 was provided as a control material for assay calibration or verification. In addition, three to four blinded samples of human serum and/or plasma SRMs (study materials) were distributed that represented different analytical challenges. The study samples were evaluated at NIST with isotope dilution techniques using LC with mass spectrometric detection (LC-MS) [11] and/or a LC-MS/MS reference measurement procedure (RMP) [12], both of which separate 25(OH)D₃ and 3-epi-25(OH)D₃. Participants were requested to provide individual concentration values for 25(OH)D₂ and 25(OH)D₃ along with a 25(OH)D_{Total} concentration for each of the study samples. In the most recent study, Summer 2012, participants were also asked to provide values for 3-epi-25(OH)D₃.

NIST compiled the results and statistically evaluated the data for concordance within the participant community as well as trueness relative to the NIST value for each study of the VitDQAP. Unlike DEQAS and the CAP ABVD, VitDQAP is not a proficiency testing program. Therefore, participants do not pass or fail based on how well their results agree with either the consensus range or the NIST value for each material. Results from the first exercises are presented and discussed in this report, including the first interlaboratory results for the new SRM 972a Vitamin D Metabolites in Human Serum.

2. Materials and Methods

2.1. Control and study materials

Control and study samples were distributed to participants with dry ice to keep the liquid samples frozen during shipping. SRM 2972, which is comprised of separate ethanolic solutions with certified concentrations of 25(OH)D₂ and 25(OH)D₃ [13], was distributed as a control material for each comparability study of VitDQAP. Participants were asked to report single values for 25(OH)D₂ and 25(OH)D₃; results were compiled and included in the reports to participants [14–19]. The control solutions were determined to be incompatible with many participant methods because of the ethanol solvent, particularly those utilizing IA techniques. For that reason, the participant results for the controls are mentioned for completeness but are not presented here.

To date, the following human serum and/or plasma SRMs have been distributed as study materials in the VitDQAP: SRM 968d Fat-Soluble Vitamins, Carotenoids and Cholesterol in Human Serum Level 1 (SRM 968d L1) and Level 2 (SRM 968d L2); SRM 968e Fat-Soluble Vitamins, Carotenoids and Cholesterol in Human Serum Level 1 (SRM 968e L1), Level 2 (SRM 968e L2), and Level 3 (SRM 968e L3); SRM 972 Vitamin D in Human Serum Level 3 (SRM 972 L3); SRM 972a Vitamin D Metabolites in Human Serum Level 1 (SRM 972a L1), Level 2 (SRM 972a L2), and Level 3 (SRM 972a L3); and SRM 1950 Metabolites in Human Plasma. For all of these materials except SRM 968d (both levels) and SRM 968e (all levels), the NIST result for 25(OH)D_{Total} is the sum of the certified values for 25(OH)D₃ and 25(OH)D₂, and the 95% confidence limit (U_{95}) was approximated using the individual uncertainties reported for the two analytes. For SRM 968d and SRM 968e, the NIST value for 25(OH)D₃ was obtained using an LC-MS/MS RMP [12], and the U_{95} confidence interval includes components for both measurement variability and measurement uncertainty associated with the density of the materials.

Table 1 lists the 10 unique study materials and NIST values for each of the six completed comparability studies of VitDQAP. While SRM 2972 was distributed as a known control to the participants in each study, SRM 968d L1 was used as a blinded human serum control and was distributed in each comparability study except for Study 1.

2.2. Participant methods

Participants used all of the major techniques to evaluate the control and study samples, including chemiluminescence immunoassay (CLIA), enzyme immunoassay (EIA), radioimmunoassay (RIA), LC-MS, LC-MS/MS, LC with ultraviolet absorbance detection (LCUV), and in Study 2, LC with electrochemical detection. For graphical purposes, the LC-MS and LC-MS/MS participant results are collectively referred to as LC-MSⁿ. Information regarding kit vendors for the IA methods is not collected or reported by the VitDQAP, as it is not NIST policy to endorse or potentially discriminate against any particular company. The number of participants providing results (N) for all techniques and for each major technique (IA and LC) in the first six studies is reported in Table 1. The majority of the VitDQAP participants used LC-MS/MS methods with multiple reaction monitoring to differentiate the mass transitions for 25(OH)D₂ and 25(OH)D₃ and at least one stable isotope labeled internal standard. Complete details for each participant method may be found in the comparability study reports [14–19].

2.3. Data analysis

The results for each control and study sample were compiled and evaluated by NIST. For the individual IA methods (CLIA, EIA and RIA) and LC-UV, N was often 8, so robust summary statistics may not be representative of the community and were not determined. Therefore, the community consensus statistics were determined for all reported methods, the IA methods only, and the LC methods only. The community results included the median values for 25(OH)D_{Total}, the median absolute deviation estimate (MAD_e), a robust estimate of the standard deviation (SD), and the coefficient of variation (CV).

3. Results

LC methods can determine 25(OH)D₂ and 25(OH)D₃ through the selection of column and mobile phase conditions that separate the metabolites, but IA methods measure only 25(OH)D_{Total}. Therefore, only the results for 25(OH)D_{Total} can be compared to evaluate the laboratory performance of both IA and LC techniques. Table 1 provides a summary of the participant median and CV values for 25(OH)D_{Total} in each of the study materials for all methods, the IA methods only, and the LC methods only. A more detailed assessment of the key results for selected study samples follows.

3.1. Serum vs plasma

To determine if there is a matrix effect on the measurements for 25(OH)D_{Total}, VitDQAP has distributed both serum and plasma materials for evaluation by the participants. To assess a potential matrix effect, it is instructive to compare the results from SRM 1950, a human plasma material, to those obtained for SRM 972a L1, a human serum material, both of which have similar 25(OH)D₃ concentrations and low amounts of 25(OH)D₂.

The single reported values for 25(OH)D_{Total} in SRM 1950 (from Study 5) and SRM 972a L1 (Study 6) are plotted in Fig. 1A and B, respectively, and are distinguished by different symbols for each of the individual methods (CLIA, EIA, RIA, LC-MSⁿ, and LC-UV). From the single reported values for all datasets for a given technique (IA or LC), the consensus median and the consensus variability ($2 \times \text{MAD}_e$) were determined and are also plotted. For the IA data for both SRM 1950 and SRM 972a L1, the non-Gaussian data distribution contributes to dispersion of the central 50% of the data, resulting in a relatively large MAD_e. Hence, the consensus variability based on MAD_e is an overestimation of the 95% confidence limits about the median, and a meaningful assessment of the consensus range and the outlying results is hindered for the IA results. For both matrix materials SRM 1950 and SRM 972a L1, the consensus median values for both the LC and IA results are biased

higher than the NIST value (Fig. 1, Table 1). When the results for the individual methods are compared, no significant differences were observed for these plasma and serum SRMs except for the LC-UV results, which exhibit significant dispersion for SRM 972a L1 (Fig. 1B). Even though some CLIA manufacturers claim that their methods are not appropriate for plasma matrices [4], the CLIA methods provided similar results to the other IA methods for both of these study materials. Also, it is well-documented that method performance for both LC and IA is quite comparable for materials that contain predominantly 25(OH)D₃, which is consistent with the results for SRM 1950 and SRM 972a L1 [20].

3.2. Endogenous versus fortified 25(OH)D₂

VitDQAP has also investigated laboratory performance on human serum in which 25(OH)D₂ is a major component of 25(OH)D_{Total}. In SRM 972 L3, the 25(OH)D₂ comprises ≈ 60% of 25(OH)D_{Total} and was fortified using an ethanolic spike. Conversely, SRM 972a L3 contains only endogenous 25(OH)D₂, which comprises ≈ 40% of 25(OH)D_{Total}. The participant results for 25(OH)D_{Total} in SRM 972 L3 (Study 3) and SRM 972a L3 (Study 6) are plotted in Fig. 1C and D, respectively.

The results for 25(OH)D_{Total} in SRM 972 L3 (Fig. 1C) are markedly different than those observed for SRM 1950 and SRM 972a L1 (Fig. 1A and B). The IA and LC techniques yield bimodal results for SRM 972 L3, and the IA median value is ≈ 35% lower than the both the LC median and the NIST values (Fig. 1C, Table 1). In addition, the consensus ranges for the IA and LC results exhibit only partial overlap. The NIST value for SRM 972 L3 falls within the consensus range for the LC methods but not for the IA methods; the NIST value was obtained with a combination of LC-MS and LC-MS/MS methods, which explains the better agreement with the participant LC results.

When samples spiked with 25(OH)D₂ or 25(OH)D₃ in ethanol were provided as study samples in the DEQAS program, the IA methods that did not utilize an extraction step also under-recovered the augmented metabolites, and the ethanol was demonstrated not to be the source of the bias in the measurements [21]. The augmented 25(OH)D₂ is not incorporated into the matrix in the same manner as the endogenous metabolites, which leads to an underrepresentation of the 25(OH)D_{Total} with some of the IA methods for fortified materials including SRM 972 L3.

The results for 25(OH)D_{Total} in SRM 972a L3 (Fig. 1D) are similar to those obtained for SRM 972 L3, but the difference between the IA and LC results is not as marked. The IA median value is lower than the NIST value by ≈ 14%, but the NIST value falls within the consensus range for the IA results. Even though the IA methods perform better on SRM 972a L3, which contains only endogenous 25(OH)D₂, the 25(OH)D_{Total} is still underrepresented when compared to both the LC value and the NIST value. The most likely culprit is non-equivalent response of the IA methods to the 25(OH)D₃ and 25(OH)D₂ that contribute to the 25(OH)D_{Total} (see review of IA methods in [22]). Conversely, the LC median result is higher than the NIST result, but this is typical observation for all materials evaluated in the VitDQAP and does not suggest a technique bias from the 25(OH)D₂ metabolite. However, to obtain accurate values for 25(OH)D_{Total}, LC methods require independent measurements of both 25(OH)D₂ and 25(OH)D₃. In SRM 972a L3, there were three outlying LC results that underestimated the 25(OH)D_{Total} either because they did not measure or did not detect the 25(OH)D₂.

3.3. 3-Epi-25(OH)D₃ in study materials

Using the NIST methods, the 3-epi-25(OH)D₃ metabolite was detected in most of the serum/plasma study materials regardless of the 25(OH)D₃ concentration. However, the 3-

epi-25(OH)D₃ concentration was below the NIST limit of quantitation (LOQ; ≈ 0.5 ng/ml or 1.25 nmol/l) for SRM 968e L1, which has a low concentration of 25(OH)D₃. For most of the SRM study materials that have 25(OH)D₃ concentrations ≥ 10 ng/ml (25 nmol/l), the 3-epi-25(OH)D₃ was above the NIST LOQ, but no value was assigned. The exceptions include the SRMs that were specifically designed for vitamin D metabolite measurements (SRM 972 and SRM 972a), for which NIST has provided certified or reference concentration values for 3-epi-25(OH)D₃ (Table 1). When SRM 972 L3, SRM 972a L1, and SRM 972a L3 were distributed as study materials in the VitDQAP, only 3 of the participants (all using LC-MS/MS methods) reported values for this metabolite.

3.4. Results for the serum control, SRM 968d L1

To monitor participant performance over time, a blinded human serum control, SRM 968d L1, was distributed during each round of VitDQAP except for Study 1. Fig. 2 presents a box-and-whisker plot for the results for SRM 968d L1 in each of the studies. The participant results for SRM 968d L1 exhibit consistent precision for both major techniques, and the majority of the IA and LC results are higher than the NIST value for this material. In addition, the IA median result is consistently higher than the LC median value (middle line in each of the boxes). It appears that the participant methods and calibrants (or kits for IA methods) have remained relatively unchanged over the timeframe established by the five most recent VitDQAP studies (≈ 2 y).

3.5. Correlation of 25(OH)D_{Total} results with a clinical range

The serum 25(OH)D_{Total} concentration levels associated with vitamin D deficiency and adequacy are still widely debated in the clinical community, and several different reference ranges have been established. One example of current guidance regarding 25(OH)D_{Total} concentrations and human health is from the Institute of Medicine (IOM) [23], which indicates that serum concentrations below 12 ng/ml (30 nmol/l) are indicative of deficiency, between 12 ng/ml (30 nmol/l) and 20 ng/ml (50 nmol/l) are associated with inadequacy, and between 20 ng/ml (50 nmol/l) and 50 ng/ml (125 nmol/l) are considered adequate. The ability of a laboratory to accurately measure 25(OH)D_{Total} is of particular concern for low levels that might be associated with deficiency and inadequacy, regardless of which reference range is used.

To illustrate how laboratory results correlate with 25(OH)D_{Total} adequacy or deficiency, Fig. 3 plots the average values ($\pm 2 \times$ SD) for duplicate samples of SRM 972a L2 (from Study 5) in conjunction with the clinically-relevant 25(OH)D_{Total} concentration levels as defined by the IOM. For SRM 972a L2, the within-lab precision was approximated with a relative SD, which ranged from 0.3% to 37% and was $\approx 5\%$ for $\approx 40\%$ of the participating labs. In addition, the CV was 11% for all methods, which indicates a lack of measurement comparability and correlates with the equal dispersion of results among the inadequate and adequate 25(OH)D_{Total} concentration ranges. Such dispersion hinders the designation of vitamin D status for serum samples with 25(OH)D_{Total} concentrations near clinically-relevant levels like SRM 968d L1. Furthermore, since lab results with potentially high variabilities have been used to define reference ranges for 25(OH)D_{Total}, it is unclear if the ranges are truly representative of the 25(OH)D_{Total} levels in the general population or if they are biased.

4. Discussion

In the first six comparability studies of the VitDQAP, the inter-laboratory CV for the serum and plasma materials that contain predominantly 25(OH)D₃ was consistently in the range from $\approx 7\%$ to $\approx 19\%$, which is similar to the CV reported for DEQAS distributions since

2009 [24;25]. Papers describing results from DEQAS and a recent interlaboratory comparison exercise for 25(OH)D_{Total} measurements have attributed improvements in both the accuracy and comparability of the different vitamin D assays to the use of SRM 972 [1;26]. The VitDQAP was established after the release of SRM 972 and SRM 2972, which could partially explain the consistency of the participant results over the first six comparability studies.

Although inter-laboratory precision was unchanged during the first 6 comparability studies, accuracy remains an issue in VitDQAP for both IA and LC techniques. For most SRM study samples evaluated in the VitDQAP, the labs performed consistently when evaluated as individual methods (i.e., CLIA, EIA, RIA, and LC-MSⁿ) and as a technique (IA vs LC), with the exception of LC-UV. However, the median results were almost always biased higher than NIST, regardless of technique. What are these biases? For the methods reported by many of the LC participants in the VitDQAP, the 3-epi-25(OH)D₃ coelutes with 25(OH)D₃ and is detected by the same ions in LC-MS and LC-MS/MS and absorbance wavelength in LC-UV, leading to a positive bias in the 25(OH)D_{Total} results. Initially, it was thought that the 3-epimer was only a concern in infant serum, but many studies have demonstrated that the 3-epi-25(OH)D₃ is present in both infant and adult serum [10;12;27–34]. In a recent investigation of 214 patient samples by Lensmeyer et al., the 3-epi-25(OH)D₃ was detected in 99% of the samples with a mean relative amount of epimer to 25(OH)D₃ of 4.8% [34]. Therefore, the epimer is a consistent source of bias for LC methods that do not chromatographically resolve it from 25(OH)D₃.

The 3-epi-25(OH)D₃ is not detected by and is not a source of bias for any of the major IA methods, according to the results from a recent DEQAS exercise [26]. Even though the IA methods do not detect 3-epi-25(OH)D₃, the participant IA method results were biased higher than both the LC method results and the NIST values for many of the VitDQAP study samples that contain predominantly 25(OH)D₃ including the serum control SRM 968d L1 (Fig. 3). A potential bias for IA methods that has not been explored in VitDQAP is interference from the 24,25-dihydroxyvitamin D metabolites, which could yield high results for 25(OH)D_{Total} [20]. However, for serum materials that contained appreciable amounts of 25(OH)D₂ such as SRM 972 L3 and SRM 972a L3, the IA results were biased low, which was likely due to non-equivalent response to 25(OH)D₃ and 25(OH)D₂. In a recent study comparing the results for 25(OH)D_{Total} in plasma collected from 203 healthy, sick, and pregnant individuals, 25(OH)D₂ was detectable (using LC-MS/MS, which can distinguish between 25(OH)D₃ and 25(OH)D₂) in only 11, or 5.4%, of the samples [4]. The results from this study suggest that 25(OH)D₂ is not as prevalent a bias in many clinically-different patient populations as the 3-epi-25(OH)D₃ metabolite.

Another source of bias in the measurements that can affect both IA and LC methods is the purity of the reference standards used to prepare the calibrants or kits. At NIST, calibrants are prepared gravimetrically with reference standards that were assessed for purity using LC-UV, LC-MS, thermogravimetric analysis and/or Karl Fisher titration (to determine water content, which is often a major impurity). If the impurities in the calibration standards are not accounted for, the quantitative results will be biased high regardless of which major technique is used.

5. Conclusions

NIST and NIH-ODS collaboratively established the VitDQAP to improve the comparability of 25(OH)D measurements in the clinical community. Major goals of the program are to reduce the inter-laboratory variability of the participant 25(OH)D results, to achieve better agreement between the participant consensus median value and the NIST value, and to

better understand the sources of bias between the results. In the VitDQAP, the within-lab variability for many of the participants was 5%, and until all labs improve to < 5%, it is very difficult to assess individual biases. Also, large within-lab variability translates to relatively high inter-laboratory imprecision (CV ranging from $\approx 7\%$ to $\approx 28\%$ for the study materials), hindering a definitive assessment of technique bias (IA and LC) and blurring conclusions about the accuracy and clinical validity of the 25(OH)D_{Total} measurements.

As a community, a lack of significant improvements has been observed during the first six comparability studies of the VitDQAP, indicating that the community has reached a performance plateau with the currently-used methods. It is recommended that individual laboratories use SRM 972a and SRM 2972, which have certified concentration values for the vitamin D metabolites, to assess and to reduce or eliminate their own method biases.

Acknowledgments

Partial funding for this work was provided by the NIH-ODS. The authors acknowledge Joseph M. Betz and Mary Frances Picciano of NIH-ODS for recognizing the need for measurement comparability and for establishing the VitDQAP in collaboration with NIST. This manuscript is dedicated to the memory of Mary Frances Picciano.

Abbreviations

3-epi-25(OH)D₃	3-epi-25-hydroxyvitamin D ₃
25(OH)D₂	25-hydroxyvitamin D ₂
25(OH)D₃	25-hydroxyvitamin D ₃
U₉₅	95% confidence limit
ABVD	Accuracy-Based Vitamin D
ALTM	all-laboratory trimmed mean
CLIA	chemiluminescence immunoassay
EIA	enzyme immunoassay
IOM	Institute of Medicine
IA	immunoassay
MADe	median absolute deviation estimate
NIST	National Institute of Standards and Technology
NIH-ODS	National Institutes of Health Office of Dietary Supplements
VitDQAP	NIST/NIH Vitamin D Metabolites Quality Assurance Program
RMP	reference measurement procedure
SRM	Standard Reference Material
25(OH)D_{Total}	total 25-hydroxyvitamin D
DEQAS	Vitamin D External Quality Assessment Scheme

References

1. Binkley N, Krueger DC, Morgan S, Wiebe D. Current status of clinical 25-hydroxyvitamin D measurement: An assessment of between-laboratory agreement. *Clin Chim Acta*. 2010; 411:1976–1982. [PubMed: 20713030]

2. Farrell CJL, Martin S, McWhinney B, Straub I, Williams P, Herrmann M. State-of-the-art vitamin D assays: A comparison of automated immunoassays with liquid chromatography-tandem mass spectrometry methods. *Clin Chem*. 2012; 58:531–542. [PubMed: 22230812]
3. Binkley N, Krueger D, Cowgill CS, et al. Assay variation confounds the diagnosis of hypovitaminosis D: A call for standardization. *J Clin Endocrinol Metab*. 2004; 89:3152–3157. [PubMed: 15240586]
4. Heijboer AC, Blankenstein MA, Kema IP, Buijs MM. Accuracy of 6 routine 25-hydroxyvitamin D assays: Influence of vitamin D binding protein concentration. *Clin Chem*. 2012; 58:543–548. [PubMed: 22247500]
5. Snellman G, Melhus H, Gedeberg R, et al. Determining vitamin D status: A comparison between commercially available assays. *Plos One*. 2010; 5
6. Lai JKC, Lucas RM, Banks E, Ponsonby AL. Ausimmune Investigator Group. Variability in vitamin D assays impairs clinical assessment of vitamin D status. *Intern Med*. 2012; 42:43–50.
7. Barake M, Daher RT, Salti I, et al. 25-Hydroxyvitamin D assay variations and impact on clinical decision making. *J Clin Endocrinol Metab*. 2012; 97:835–843. [PubMed: 22238386]
8. Hypponen E, Turner S, Cumberland P, Power C, Gibb I. Serum 25-hydroxyvitamin D measurement in a large population survey with statistical harmonization of assay variation to an international standard. *J Clin Endocrinol*. 2007; 92:4615–4622.
9. College of American Pathologists. 2013 Surveys and Anatomic Pathology Education Programs. 2013 www.cap.org.
10. Phinney KW, Bedner M, Tai SSC, et al. Development and certification of a Standard Reference Material for vitamin D metabolites in human serum. *Anal Chem*. 2012; 84:956–962. [PubMed: 22141317]
11. Bedner M, Phinney KW. Development and comparison of three liquid chromatography-atmospheric pressure chemical ionization/mass spectrometry methods for determining vitamin D metabolites in human serum. *J Chromatogr A*. 2012; 1240:132–139. [PubMed: 22533908]
12. Tai SSC, Bedner M, Phinney KW. Development of a candidate reference measurement procedure for the determination of 25-hydroxyvitamin D-3 and 25-hydroxyvitamin D-2 in human serum using isotope-dilution liquid chromatography-tandem mass spectrometry. *Anal Chem*. 2010; 82:1942–1948. [PubMed: 20136128]
13. NIST SRM website. 2013. <http://www.nist.gov/srm>
14. Bedner M, Lippa KA. NIST/NIH Vitamin D Metabolites Quality Assurance Program report of participant results: Winter 2010 comparability study (exercise 1). NISTIR 7890. National Institute of Standards and Technology. 2012 DOI <http://dx.doi.org/10.6028/NIST.IR.7890>.
15. Lippa KA, Bedner M, Tai SSC. NIST/NIH Vitamin D Metabolites Quality Assurance Program report of participant results: Summer 2010 comparability study (exercise 2). NISTIR 7891. National Institute of Standards and Technology. 2012 DOI <http://dx.doi.org/10.6028/NIST.IR.7891>.
16. Bedner M, Lippa KA. NIST/NIH Vitamin D Metabolites Quality Assurance Program report of participant results: Winter 2011 comparability study (exercise 3). NISTIR 7892. National Institute of Standards and Technology. 2012 DOI <http://dx.doi.org/10.6028/NIST.IR.7892>.
17. Lippa KA, Bedner M, Tai SSC. NIST/NIH Vitamin D Metabolites Quality Assurance Program report of participant results: Summer 2011 comparability study (exercise 4). NISTIR 7893. National Institute of Standards and Technology. 2012 DOI <http://dx.doi.org/10.6028/NIST.IR.7893>.
18. Bedner M, Lippa KA, Tai SSC. NIST/NIH Vitamin D Metabolites Quality Assurance Program report of participant results: Winter 2012 comparability study (exercise 5). NISTIR 7894. National Institute of Standards and Technology. 2013 DOI <http://dx.doi.org/10.6028/NIST.IR.7894>.
19. Lippa KA, Bedner M. NIST/NIH Vitamin D Metabolites Quality Assurance Program report of participant results: Summer 2012 comparability study (exercise 6). NISTIR 7895. National Institute of Standards and Technology. 2013 DOI <http://dx.doi.org/10.6028/NIST.IR.7895>.
20. Carter GD, Carter R, Jones J, Berry J. How accurate are assays for 25-hydroxyvitamin D? Data from the international vitamin D external quality assessment scheme. *Clin Chem*. 2004; 50:2195–2197. [PubMed: 15375018]

21. Carter GD, Jones JC, Berry JL. The anomalous behaviour of exogenous 25-hydroxyvitamin D in competitive binding assays. *J Steroid Biochem Mol Biol.* 2007; 103:480–482. [PubMed: 17197167]
22. Wallace AM, Gibson S, de la Hunty A, Lamberg-Allardt C, Ashwell M. Measurement of 25-hydroxyvitamin D in the clinical laboratory: Current procedures, performance characteristics and limitations. *Steroids.* 2010; 75:477–488. [PubMed: 20188118]
23. Institute of Medicine, Food and Nutrition Board. Dietary reference intakes for calcium and vitamin D. Washington, DC: National Academy Press; 2010.
24. Carter GD. 25-Hydroxyvitamin D assays: The quest for accuracy. *Clin Chem.* 2009; 55:1300–1302. [PubMed: 19423731]
25. Carter GD, Berry JL, Gunter E, et al. Proficiency testing of 25-hydroxyvitamin D (25-OHD) assays. *J Steroid Biochem Mol Biol.* 2010; 121:176–179. [PubMed: 20302938]
26. Carter GD. 25-Hydroxyvitamin D: A difficult analyte. *Clin Chem.* 2012; 58:486–488. [PubMed: 22238254]
27. Kamao M, Tatematsu S, Hatakeyama S, et al. C-3 epimerization of vitamin D-3 metabolites and further metabolism of C-3 epimers-25-hydroxyvitamin D-3 is metabolized to 3-epi-25-hydroxyvitamin D-3 and subsequently metabolized through C-1 alpha or C-24 hydroxylation. *J Biol Chem.* 2004; 279:15897–15907. [PubMed: 14757768]
28. Singh RJ, Taylor RL, Reddy GS, Grebe SKG, et al. C-3 epimers can account for a significant proportion of total circulating 25-hydroxyvitamin D in infants, complicating accurate measurement and interpretation of vitamin D status. *J Clin Endocrinol.* 2006; 91:3055–3061.
29. Schleicher RL, Encisco SE, Chaudhary-Webb M, Paliakov E, McCoy LF, Pfeiffer CM. Isotope dilution ultra performance liquid chromatography-tandem mass spectrometry method for simultaneous measurement of 25-hydroxyvitamin D2, 25-hydroxyvitamin D3 and 3-epi-25-hydroxyvitamin D3 in human serum. *Clin Chim Acta.* 2011; 412:1594–1599. [PubMed: 21601563]
30. Lensmeyer GL, Wiebe DA, Binkley N, Drezner MK. HPLC method for 25-hydroxyvitamin D measurement: Comparison with contemporary assays. *Clin Chem.* 2006; 52:1120–1126. [PubMed: 16574756]
31. Lensmeyer G, Wiebe D, Binkley N, Drezner M. Measurement of 25-hydroxyvitamin D revisited - Reply. *Clin Chem.* 2006; 52:2305–2306.
32. Stepman HCM, Vanderroost A, Stockl D, Thienpont LM. Full-scan mass spectral evidence for 3-epi-25-hydroxyvitamin D-3 in serum of infants and adults. *Clin Chem Lab Med.* 2011; 49:253–256. [PubMed: 21143012]
33. Strathmann FG, Sadilkova K, Laha TJ, et al. 3-epi-25 hydroxyvitamin D concentrations are not correlated with age in a cohort of infants and adults. *Clin Chim Acta.* 2012; 413:203–206. [PubMed: 21983164]
34. Lensmeyer G, Poquette M, Wiebe D, Binkley N. The C-3 epimer of 25-hydroxyvitamin D-3 is present in adult serum. *J Clin Endocrinol Metab.* 2012; 97:163–168. [PubMed: 22013102]

Highlights

- VitDQAP, the first accuracy-based program for 25(OH)D, used serum SRMs as study samples
- Summary statistics were determined for LC, immunoassay, and all method results
- Participant performance was consistent for SRMs containing mostly 25(OH)D₃
- Participant immunoassay results were biased low for serum SRMs with high 25(OH)D₂
- In general large within- and between-lab variability hindered accuracy assessments

Fig. 1.

Results for 25(OH)D_{Total} in SRM 1950 Metabolites in Human Plasma (A), SRM 972a Vitamin D Metabolites in Human Serum L1 (B), SRM 972 Vitamin D in Human Serum L3 (C), and SRM 972a L3 (D). The results from the individual methods are displayed with different symbols, including: CLIA (), EIA (), RIA (), LC-MSⁿ (), and LC-UV (). For each of the techniques within both graphs (IA and LC), the solid lines (—) represent the consensus median and the dashed lines (---) represent approximate 95% confidence intervals ($2 \times \text{MAD}_e$). The grey-shaded bars represent the ranges bound by the NIST values with $\pm$ estimated U_{95} uncertainty.

Fig. 2.

Box-and-whisker plot for the 25(OH)D_{Total} results in the human serum control, SRM 968d L1, in Study 2 through Study 6. Each box represents the 25% to 75% quartile range, with the IA and LC results depicted separately. The width of the boxes is proportional to the total number of results reported for each technique. The error bars represent the empirically-determined 95% range, and results outside of that range are designated with an “•”. The grey-shaded bar represents the range bound by the NIST values with $\pm$ estimated U_{95} uncertainty.

Fig. 3.

Participant results for 25(OH)D_{Total} in SRM 972a L2 in conjunction with the clinically-relevant serum 25(OH)D_{Total} concentration levels. Each point represents the mean value (N = 2), and the error bars represent ± 2 SD. The grey-shaded bars represent the ranges bound by the NIST values with $\pm$ estimated U_{95} uncertainty.

Study materials, NIST values with U_{95} expanded uncertainty, and participant results for the first six comparability studies of the VitDQAP. Participant results for 25(OH)D_{Total} are the median values.

Table 1

Material	Study	Participant Results										NIST Values ± U_{95} (ng/mL)			
		N		25(OH)D _{Total} (ng/mL)						CV (%)		25(OH)D _{total}		25(OH)D ₂	3-epi-25(OH)D ₃
		All	IA	LC	All	IA	LC	IA	LC	All	IA	25(OH)D _{total}		25(OH)D ₂	3-epi-25(OH)D ₃
SRM 1950*	1	17	7	10	26.7	27.0	26.6	7	17	6		25.30 ± 0.79	24.78 ± 0.77	0.52 ± 0.17 [‡]	n.a.v.
	5	57	17	40	27.2	28.0	27.0	10	13	9					
SRM 968d L1	2	39	16	23	14.0	14.2	13.2	15	12	11		12.4 ± 0.3	12.4 ± 0.3	< 0.5	n.a.v.
	3	35	14	21	13.8	14.5	13.4	15	13	12					
	4	43	17	26	14.1	15.9	13.6	14	17	8					
	5	57	17	40	14.4	14.6	14.0	11	8	11					
SRM 968d L2	6	56	18	38	13.4	14.0	13.1	10	11	10					
	2	39	16	23	11.8	10.3	12.2	16	15	11		10.3 ± 0.2	10.3 ± 0.2	< 0.5	n.a.v.
SRM 968e L1	4	45	17	28	7.9	8.0	7.9	19	19	18		7.09 ± 0.14	7.09 ± 0.14	< 0.5	< 0.5
SRM 968e L2	4	45	17	28	14.7	16.0	14.2	14	15	11		12.9 ± 0.3	12.9 ± 0.3	< 0.5	n.a.v.
SRM 968e L3	4	45	17	28	22.9	27.1	21.8	16	10	9		19.9 ± 0.4	19.9 ± 0.4	< 0.5	n.a.v.
SRM 972 L3 [‡]	3	35	14	21	41.6	30.1	46.7	28	21	10		44.9 ± 2.3	18.5 ± 1.1	26.4 ± 2.0	1.06 ± 0.03 [‡]
SRM 972a L1	6	56	18	38	31.3	30.9	31.5	7	10	7		29.3 ± 1.1	28.8 ± 1.1	0.54 ± 0.06 [‡]	1.84 ± 0.08
SRM 972a L2	5	57	17	40	19.5	19.6	19.5	9	10	9		18.9 ± 0.4	18.1 ± 0.4	0.81 ± 0.06	1.29 ± 0.06
SRM 972a L3	6	56	18	38	33.9	28.5	34.8	17	11	10		33.2 ± 0.6	19.8 ± 0.5	13.3 ± 0.3	1.18 ± 0.13 [‡]

* human plasma material;
[‡] fortified with 25(OH)D₂;
[‡] reference concentration value;
n.a.v. = no assigned value