An evaluation of total 25-hydroxyvitamin D assay standardization

Where are we today?
Bjerg, Lise Nørkjær; Halgreen, Jens Romlund; Hansen, Svend Høime; Morris, Howard A.; Jørgensen, Niklas Rye

Published in:
Journal of Steroid Biochemistry and Molecular Biology

DOI:
10.1016/j.jsbmb.2019.03.015

Publication date:
2019

Document version:
Accepted manuscript

Document license:
CC BY-NC-ND

Citation for published version (APA):
Accepted Manuscript

Title: An evaluation of total 25-hydroxyvitamin D assay standardization; where are we today?

Authors: Lise Nørkjær Bjerg, Jens Romlund Halgreen, Svend Høime Hansen, Howard A. Morris, Niklas Rye Jørgensen

PII: S0960-0760(18)30757-X
DOI: https://doi.org/10.1016/j.jsbmb.2019.03.015
Reference: SBMB 5338

To appear in: Journal of Steroid Biochemistry & Molecular Biology

Received date: 13 December 2018
Revised date: 12 March 2019
Accepted date: 23 March 2019

Please cite this article as: Bjerg LN, Halgreen JR, Hansen SH, Morris HA, Jørgensen NR, An evaluation of total 25-hydroxyvitamin D assay standardization; where are we today?, Journal of Steroid Biochemistry and Molecular Biology (2019), https://doi.org/10.1016/j.jsbmb.2019.03.015

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 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.
An evaluation of total 25-hydroxyvitamin D assay standardization; where are we today?

Lise Nørkjaer Bjerg¹, Jens Romlund Halgreen¹, Svend Haøme Hansen¹, Howard A. Morris², Niklas Rye Jørgensen¹,³

¹Department of Clinical Biochemistry, Rigshospitalet, Copenhagen, Denmark.
²School of Pharmacy and Medical Sciences, University of South Australia, Chemical Pathology Directorate and Hanson Institute, SA Pathology, Adelaide, Australia
³OPEN, Odense Patient data Explorative Network, Odense University Hospital/Institute of Clinical Research, University of Southern Denmark, Odense, Denmark.

Running head: Vitamin D standardization, 25-hydroxyvitamin D

Acknowledgements: This work was supported by Immunodiagnostic Systems, plc, Tyne and Wear, UK; Roche Diagnostics, Hvidovre, DK; DiaSorin, Saluggia, IT; Ortho Clinical Diagnostics, Claritan, NJ, USA; Alere A/S, Holstebro, DK; who all provided assays for the comparison.

Conflict of interest: Lise Nørkjaer Bjerg and Niklas Rye Jørgensen declare that they have received the assays for the study as a donation from the above-mentioned manufacturers, but have no further conflict of interest. Howard Morris has received support from IDS plc, Diasorin and Roche Diagnostics in the form of assay reagents for research.

Corresponding author:

Niklas Rye Jørgensen
Professor, MD, DMSc, PhD
Department of Clinical Biochemistry
Rigshospitalet, Denmark
Phone + 45 38 63 24 56
Email: niklas@dadlnet.dk

Word count: 4,225
Figures: 2
Tables: 7

Keywords: Vitamin D standardization program, 25-hydroxyvitamin D
Highlights:

- Only few 25(OH) D assays comply with requirements in the Vitamin D Standardization Program
- 25(OH) assays vary greatly in the categorization of patients in vitamin D deficiency/sufficiency
- Standardization has been obtained in some 25(OH)D immunoassays

ABSTRACT

Background: Serum total 25-hydroxyvitamin D is a measure of the total circulating 25-hydroxyvitamin D concentration and is the primary measurement for estimating vitamin D status. A number of automated immunoassays are commercially available, and in an attempt to standardize the assays the Vitamin D Standardization Program (VDSP) was established in 2010. Therefore, the aim of the current project is to evaluate the status of the standardization of routinely used 25-hydroxyvitamin D assays.

Methods: 200 patient serum samples were measured in Spring 2017 on seven different assays for 25-hydroxyvitamin D. Samples were measured in duplicate for the evaluation of precision. A certified standard reference material (SRM972a) from The National Institute of Standardization and Technology (NIST) was measured to evaluate the accuracy of the assays. Finally, the agreement of the assays of clinically categorizing patients into vitamin D deficiency, inadequacy or adequacy was evaluated.

Results: All seven assays achieved precision below the VDSP requirement of CV <10%. However, only two of the assays achieved an accuracy bias <5% when measuring the SRM972a. When comparing methods using Deming regression, substantial proportional and/or systematic bias was found between many of the assays. Finally, when evaluating the ability of the assays to categorize patients into “vitamin D deficiency” (25-hydroxyvitamin D concentration < 30 nmol/L (<12 ng/mL)), “vitamin D inadequacy” (30-50 nmol/L)
(12-20 ng/mL), “vitamin D adequacy” (50-250 nmol/L (20-100 ng/mL)) and “risk of toxicity” (>250 nmol/L (>100 ng/mL)), clinically relevant differences between assays were detected. Especially in the deficiency group, major discrepancies were found as the percentage of patients ranged from 1.5% to 14.3% between the assays.

Conclusions: In conclusion, some of the commercially available assays have been standardized with performance as required by the VDSP. However, several of the assays do still not comply with the VDSP requirements even eight years after the program was started. This may have clinical consequences for patients, and manufacturers are therefore encouraged to continue their work on standardizing serum 25-hydroxyvitamin D assays.

INTRODUCTION

Vitamin D is a hormone that has pivotal roles in calcium metabolism and in the preservation of skeletal health. It can be synthesized in the skin upon exposure to sunlight, but can also be ingested through nutritional sources, where especially oil-rich fish such as salmon, herring, and mackerel are the most common. Vitamin D exists in two isoforms, vitamin D<sub>2</sub> and vitamin D<sub>3</sub>. Except for certain types of mushrooms, vitamin D from animal sources consists of vitamin D<sub>3</sub>. However, fortified food products and vitamin D supplements can contain either vitamin D<sub>2</sub> or vitamin D<sub>3</sub>, though there are geographical differences in the use of vitamin D<sub>2</sub>. Vitamin D from the skin or through the diet is hydroxylated in the liver to 25-hydroxyvitamin D (25(OH)D) and further in the kidneys to 1,25-dihydroxyvitamin D to become biologically active.

Serum total 25-hydroxyvitamin D (25(OH)D) is a measure of the total circulating 25OHD concentration and is defined as the sum of 25(OH)D<sub>2</sub> and 25(OH)D<sub>3</sub>. Serum 25(OH)D concentration is the primary measurement for evaluating vitamin D status [1]. A number of clinical guidelines [2, 3] have recommendations for when 25(OH)D should be measured. These include patients with a) bone diseases that may be improved with vitamin D treatment, b) patients with bone...
diseases prior to specific treatment where correcting vitamin D deficiency is appropriate, c) patients with musculoskeletal symptoms that could be attributed to vitamin D deficiency (rickets, osteomalacia), and d) a number of diseases and concurrent treatments that are known to predispose to low vitamin D levels and impaired calcium uptake. In contrast, most guidelines do not recommend general screening for vitamin D deficiency. No evidence-based consensus exists defining hypovitaminosis D [1], though internationally it is recognized that serum 25(OH)D levels below 30 nmol/L (12 ng/mL) confers an increased risk of osteomalacia and rickets. Serum concentrations between 30 and 50 nmol/L (12 and 20 ng/mL) may be inadequate in some people and is therefore termed “vitamin D deficiency”, while concentrations >50 nmol/L (>20 ng/mL) are currently considered sufficient, at least for the general population though some guidelines suggest sufficient levels should be > 75 nmol/L (30 ng/mL). Recommendations were proposed by different expert committees [3, 4]. When exceeding 250 nmol/L (100 ng/mL) the risk of vitamin D toxicity increases. Even though many studies have tried to establish which 25(OH)D levels are the optimal for skeletal musculoskeletal health, no evidence-based consensus has been reached on clear cut-off values for optimal vitamin D status [1]. One contributor to this confusion may be the lack of standardization of vitamin D assays. Vitamin D is an extremely difficult analyte to measure and the use of different methods such as LC-MS/MS and a variety of immunoassays and underlying differences in traceability have contributed to the variability in the levels of measured 25(OH)D. Therefore, in 2010 the Vitamin D Standardization Program (VDSP) was founded with the focus of promoting standardized measurements of serum total 25(OH)D around the world [5, 6]. VDSP developed tools and methods to standardize the measurements prospectively as well as retrospectively so also 25(OH)D measurements performed in past studies could be standardized [7]. As part of the program VDSP provided National Institute of Standardization and Technology (NIST) standard reference materials (SRM) with assigned reference values determined by
measurements on gold standard analysis methods [8]. Several reports have evaluated the status of the 25(OH)D assay standardization in routine clinical laboratories over the past years. These have shown varying degrees of standardization, yet with still significant bias between routine assays and the NIST SRM. In 2017, the First International Conference on Controversies in Vitamin D was held in Pisa, Italy, where, among other vitamin D related issues, the lack of standardization of assays was emphasized [1].

In order to evaluate the current status of standardization of routinely used 25(OH)D assays in clinical laboratories the current study was initiated. Therefore, the aim of the study was to a) evaluate how seven common commercially available 25(OH)D assays performed in terms of requirements to precision and accuracy as set out in the VDSP and 2) to estimate the clinical consequences of any potential lack of standardization between assays.
MATERIALS AND METHODS

Study materials and control

Two hundred serum samples were collected consecutively in March 2017. These were collected as extra sample material from patients attending clinics from various medical specialties at Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark and from patients attending general practice clinics in Copenhagen who should already have measured 25-hydroxyvitamin D. The range of serum 25(OH)D levels was from 14.9 to 179.4 nmol/L (6.0 to 71.8 ng/mL) as measured on LC-MS/MS. Ethical approval was not required since aliquots were anonymized and the measurements were used for quality control.

Pre-analytical handling was performed according to standard procedures: venous blood was collected in serum tubes with gel and left for coagulation. Within 2 hours after collection samples were centrifuged at 2000 g for 10 minutes. Serum was aliquoted into 5 freeze tubes and stored at -20°C until thawing for analysis. Analysis was performed between May 2017 and June 2018.

A standard reference material was used as control of bias. This was the National Institute of Standards and Technology (NIST) Standard Reference Material 972a, (SRM972a) (National
Institute of Standards and Technology, Gaithersburg, MD, USA) [8]. The SRM972a contains four levels and it is based on human serum with native levels of 25(OH)D$_3$ and 25(OH)D$_2$ fortified with different concentrations of the epimer, 3-epi-25(OH)D$_3$. Certified and reference values are shown in Table 1.

$25$-hydroxyvitamin D immunoassay methods and compliance with VDSP

Samples were measured on the seven instruments and methods listed in Table 1 and were analyzed as batch analyses on 1-9 consecutive days. The samples were run in duplicate. However, the first set was analyzed in one run and the duplicate set was analyzed in a second run after the first full set was completed. Precision was calculated from the duplicate determinations of the 200 samples. Bias as compared to the SRM972a was determined from duplicate determinations of aliquots of the SRM972a which was purchased immediately prior to the study and kept on -20°C until the time of analysis. The SRM972a measurements were done on the same day as the analysis of patient samples as mentioned above.

$LC$-$MS/MS$ method for $25$-hydroxyvitamin D measurement

25(OH)D$_3$ and 25(OH)D$_2$ were analyzed by LC-MS/MS using a liquid chromatography system (Waters Acquity UPLC) connected to a triple quadrupole mass spectrometer (Waters Xevo TQ-S).
ClinCal® Calibrators from Recipe (reference number MS7013) were used. The calibrators contain a blank calibrator and 3 concentration levels. Calibrators are traceable to NIST SRM2972a.

Samples were prepared by adding patient serum (100 µL) to acetonitrile (300 µL) with isotope-labelled internal standards 26, 27-hexadeuterium-25-hydroxyvitamin D₃ (Synthetica AS, Norway), and 25-hydroxyvitamin D₂-[25,26,27-¹³C₃] (IscSciences, USA). After mixing it was purified by application of a Biotage Isolute PLD+ plate for removal of phospholipids. The eluate from purification was used for injection (injection volume 5 µL).

For chromatography an isocratic solvent gradient 82% methanol (2mM ammonium acetate, 0.1% formic acid) and 18% water (2mM ammonium acetate, 0.1% formic acid) were applied on the UPLC column: Waters HSS-T3, particle size 1.8 µm, column dimensions 2.1 mm X 100 mm, flow 0.6 ml/min. Total runtime including wash procedure was 5 minutes. In this method, 25(OH)D₃ and 25(OH)D₂ are chromatographically separated, whereas 25(OH)D₃ and epi-25-(OH)D₃ co-elutes.

For mass spectrometry quantifications were based on the positive mass transitions 383.5 > 257.2 for 25(OH)D₃, and 395.5 > 269.2 for 25(OH)D₂. Both with corresponding transitions for the internal standard compounds.

Clinical evaluation

To check whether patients were homogeneously classified into 25(OH)D deficiency/sufficiency categories, patients were categorized according to the thresholds of serum 25-hydroxyvitamin D levels as recommended by the Institute of Medicine (IOM)[4]: <30 nmol/L (<12 ng/mL) (deficiency), 30-50 nmol/L (12-20 ng/mL) (inadequacy), 50-250 (20-100 ng/mL)(sufficiency), and >250 nmol/L (>100 ng/mL) (risk of toxicity).
Statistical analyses

Precision of the assays was calculated from the duplicate measurements of the 200 samples using the Root-Mean-Square method. The assays were compared using Deming regression and Bland-Altman analysis [9]. For classification of patients into the clinically relevant categories of 25(OH)D deficiency/sufficiency, the strength of the agreement between methods was tested using the inter-rater agreement statistic weighted Kappa [10]. The strength of agreement was then interpreted based on the κ value as follows [11]: κ <0.20 (poor), 0.21-0.4 (fair), 0.41-0.60 (moderate), 0.61-0.80 (good), 0.81-1.00 (very good). All statistical analyses were performed using MedCalc Statistical Software version 16.2.1 (MedCalc Software bvba, Ostend, Belgium). A p-value < 0.05 was used as criterion for significance throughout the study.
RESULTS

Compliance with the VDSP requirements for assay performance

First, we wanted to establish whether the routinely used assays complied with the VDSP requirement to bias. We therefore measured the SRM972a in duplicate on all instruments. The VDSP requirement is set at 5%. The calculated bias ranged from 0.3% to 87.6%, based on the certified value of 25(OH)D$_2$ and 25(OH)D$_3$ in the SRM972a. However, none of the instruments fulfilled this requirement completely for all of the four SRM972a levels, though two assay (iSYS and Liaison XL) fulfilled the requirement if looking at the mean bias for all four levels (Table 3).

Next, we wanted to determine whether the methods complied with the VDSP requirements for assay performance. Therefore, we determined the precision of the assays based on the duplicate measurements of the patient samples. The VDSP requirements are 10%. The CV% ranged from 2.4 to 9.0%. Thus, all assays fulfilled the requirement (Table 4). This was also the case if CV% was calculated from the duplicate measurements of the SRM972a (Table 3).

Method comparisons

Next, we wanted to determine the correlation between measurements done using the different assays. Thus, we performed a Deming regression analysis on the results from the 200 samples by correlating results from the LC-MS/MS with the other assays (Figure 1). Significant proportional bias was demonstrated between the LC-MS/MS and AIA900, Centauer, and Vitros, while significant systematic bias was found between the LC-MS/MS and AIA900, Cobas, Vitros. Next, we correlated results between the other assays and found striking differences between the assays both in terms of slope and in terms of intercept, demonstrating both systematic and proportional
bias between the methods (Table 5). Figure 2 shows the correlations between the method with the least bias (iSYS) compared to the SRM972a with the methods with the second least bias (Liaison XL) and the methods with the biggest bias compared to the SRM972a (Centauer, Vitros, and Cobas). Thus, the methods showing the smallest bias compared to the SRM972a also correlated best in the Deming regression analysis, while substantial proportional and systematic differences were present between the other assays.

Clinical evaluation and agreement of classification

The 25-hydroxyvitamin D levels covered the full clinical measuring range. As mentioned above the IOM proposed definitions for deficiency, inadequacy, adequacy, and risk of toxicity based on serum 25-hydroxyvitamin D levels. Therefore, we wanted to determine how the different assays classified the patients into these categories. Despite the apparent comparability between assays, we found big differences in the clinical classification of patients. In the “deficiency” group (25-hydroxyvitamin D levels < 30 nmol/L (12 ng/mL) the percentage of the patients ranged from 1.5% to 14.3%. The disagreement was less in the “inadequacy” (30-50 nmol/L (12-20 ng/mL)) group where the percentage of patients ranged from 11.5% to 18%. For patients with adequate 25-hydroxyvitamin D levels (50-250 nmol/L (20-100 ng/mL)) the percentage of patients ranged from 69% to 87% (Table 6). Finally, only two of the assays classified one patient (0.5%) as having toxic levels above 250 nmol/L (100 ng/mL). When analyzing the inter-assay agreement of classification using weighted k statistics, we found agreements with k-values that ranged from 0.366 (“fair”) to 0.850 (“very good”) (Table 7). None achieved very good agreement with LC-MS/MS, however, two of the assays had an agreement of very good (Liaison XL and iSYS), which were also the two complying best with the VDSP requirements.
DISCUSSION

Brief resumé of the results

This study showed that most of the methods tested still do not comply with the requirements from the VDSP eight years after the work towards a standardization of 25(OH)D assays was started. While all methods kept within the recommended limits for CV%, no instrument fully complied with the accuracy requirement of <5% bias when measuring the SRM972a on all levels. However, two instruments fulfilled the requirements when looking at the mean bias of all four levels. This was also reflected in a highly varying correlation of results when samples were measured on the different instruments. Both systematic and proportional differences between assays where found leading to the misclassification of patients in terms of vitamin D deficiency status.

Several studies have evaluated and compared 25(OH)D assays. Though the methods overall correlated well, substantial variability between assays was found [12-16]. Some of these studies were performed early after the initiation of the VDSP initiative where standardization of assays could not yet have taken place. However, several of the studies were performed within the last three to four years where manufacturers would have had the possibility to standardize their assays according to the procedures recommended by the VDSP. Recent reports on studies similar to ours have demonstrated that current, commercially available immunoassays for 25(OH)D continue to lack standardization. One evaluated several LC-MS/MS and immunoassays in hospital and university laboratories in Japan [17]. Another was coordinated by the VDSP with samples distributed to 15 different laboratories. The results clearly demonstrated that overall, assay standardization has not been achieved for all immunoassays. In the latter study, almost all LC-MS/MS methods complied with the VDSP performance criteria of CV<10% and bias<5%. In contrast, only 50% of immunoassays met the criterion of <10% CV and only three out of eight achieved a bias <5% [18]. This is in line with the findings in our study, where none of the 7 assays
achieved a bias <5%. While demonstrating an excellent precision, the LC-MS/MS used in our study did not comply with the VDSP bias requirement. However, while LC-MS/MS is considered the gold standard it does not necessarily separate 25(OH)D₃ and its epi-isomer as here demonstrated from the results from Level 4 of SRM972a. Selection of a specific chromatographic column will be needed to obtain such separation. It may also be due to the fact that the LC-MS/MS is not directly calibrated using the SRM2972a. Thus, even though LC-MS/MS is considered the gold standard for measuring 25(OH)D it only meets the VDSP standards if procedures are according to the VDSP guidelines. Thus, lack of standardization could be caused by lack of traceability to a higher order reference standard. However, most assays in this study are traceable to the VDSP NIST standard, either the SRM 972 or the SRM 2972/2972a (Table 2).

Several methods have now been reformulated for standardization. A very recent study demonstrated that the restandardized 25(OH)D IDS-iSYS assay successfully complied with the VDSP requirements for precision and accuracy bias when evaluating DEQAS samples and comparing measurements on the iSYS platform with a VDSP traceable LC-MS/MS method [19]. This is in line with the findings that we obtained in the current study where the iSYS and the Liaison XL were the only methods fulfilling the VDSP criteria for both precision and accuracy.

The need for standardization is obvious but then the question arises whether results obtained on the new re-standardized assays diverge significantly from results obtained on pre-alignment assays as this could have substantial influence on the interpretation of patient data. This was addressed in a study comparing samples that were both measured on the pre-standardization IDS-iSYS assay and on the standardized assay [20]. Significant associations between samples were found and only negligible bias was seen. Thus, standardization of the assay appears to have little impact on the results. However, this has only been examined for the iSYS assay. Similar analyses should be done for each individual assay after standardization has been done as substantial differences between the
assays would be expected. However, this could also suggest that even though standardization of 25(OH)D assays is highly warranted for a number of reasons, lack of standardization per se will not necessarily explain the confusion between studies on which 25(OH)D levels are the optimal for skeletal musculoskeletal health. Thus, standardization of assays alone will not necessarily lead to evidence-based consensus on clear cut-off values for optimal vitamin D status.

Serum 25(OH)D is a very difficult analyte to measure [5]. A number of factors contribute to the assay variation; Antibody affinity and specificity for 25(OH)D$_2$ and 25(OH)D$_3$, cross-reactivity with other vitamin D metabolites (24,25-(OH)$_2$D), cross-reactivity with the 3-epi-25-hydroxyvitamin D$_3$ (3-epi-25-(OH)D$_3$), matrix effects, as well as the levels of vitamin D binding protein. Some of the assays evaluated in this study clearly use binding agents with affinity for 24,25-(OH)$_2$D (Roche) as well as some assays cross-react with 3-epi-25-(OH)D$_3$ (Table 2). Whether assays should include measurements of the different metabolites and the epimer has been the subject of many studies. However, no clear understanding of the role in the assessment of vitamin D status has been reached for the metabolites such as 24,25-(OH)2D and for 3-epi-25-(OH)D$_3$ and it is recommended that their measurements are standardized before any firm conclusions can be done.

Recently, the First International Conference on Controversies in Vitamin D was held in Pisa in Italy [1]. One of the topics discussed was the standardization of 25(OH)D assays and it was clearly stated that 25(OH)D assays should not include measurements of 3-epi-25-(OH)D$_3$ or any other vitamin D metabolite, as this would lead to an overestimation of 25(OH)D. Especially in children, 3-epi-25-(OH)D$_3$ is present in higher concentrations but it can also be present in adults with 5-10% of the total 25(OH)D concentration. As the assays used in this study have different affinities for 3-epi-25-(OH)D$_3$ and for 24,25-(OH)2D as stated in the information from the manufacturers, this may have contributed to the variability and bias between assays.
Clinical decision making depends on the results of measurements of 25(OH)D. Evidence-based consensus has not been reached on fixed limits for vitamin D deficiency which is partially due to the variability of 25(OH)D assays used in epidemiological/cohort studies. However, several guidelines recommend certain cut-off values for a) the general population [4] b) patients with current bone disease or risk of bone disease [2, 3]. Controversies exist to what the optimal 25(OH)D level is as the US Institute of Medicine (IOM) and the UK National Osteoporosis Society (NOS) consider serum levels below 30 nmol/L (12 ng/mL) as a good indicator of deficient vitamin D status and that levels between 30 and 50 nmol/L (12 and 20 ng/mL) in some patients may be insufficient in relation to bone health. Moreover, levels above 50 nmol/L (20 ng/mL) are considered sufficient for bone health [2, 4]. The US Endocrine Society Clinical Practice Guideline recommends somewhat higher levels of serum 25(OH)D, where “deficiency” is defined as levels below 50 nmol/L (20 ng/mL), insufficiency between 50-75 nmol/L (20-30 ng/mL) and sufficient levels above 75 nmol/L (30 ng/mL)[3]. However, as agreement among assays is varying the classification of patients/persons into these categories highly depends on the assay used for determining the individual patients’ vitamin D status. In the present study, we found poor concordance in the ability of assays to classify patients. iSYS was the method that classified most patients (14.3%) in the “deficiency” category (according to IOM and NOS classification) while Vitros classified the least in this category (1.5%). Two assays (Cobas and Centauer) each classified one patient (0.5%) as in the “risk of toxicity” category (Table 6). This was also reflected in the kappa statistics as a measure of the strength of agreement between assays’ classification of patients. Kappa values ranged from 0.366 (fair agreement) to 0.850 (very good agreement) with only the agreement between Liaison and iSYS characterized as “very good”. As the decision of starting vitamin D treatment in patients relies on the measurement of 25(OH)D, disagreement between assays has obvious clinical consequences. In agreement with a previous study by Cavalier et al., which concluded that
discrepancies in classification of vitamin D levels between assays might have clinical consequences for the individual patient [12] we also found that in many cases the disagreement was not only from one category to the next but also that several results “jumped” one category, meaning that results from one assay could fall in the “deficiency” category and in the “sufficient” category on another assay and vice versa.

The implications of lack of standardization between commercially available assays are not only on the clinical level. It also affects the comparability of results between studies on vitamin D status in the general population. These studies are often used for decision-making in terms of enrichment of food, campaigns for increased awareness of vitamin D intake, determination of sufficient/in-sufficient vitamin D levels etc. However, assay variation has made it difficult or even impossible to compare or pool data from the many studies. For the purpose of being able to compare existing studies where 25(OH)D was measured on non-standardized assay, VDSP provides a manual for retrospectively standardizing laboratory results of 25(OH) [7]. It has become clear that interpretation of data from large cohort-based studies is highly dependent and influenced by the assay used and whether there has been a re-standardization of results. This was clearly demonstrated in the NHANES study. Before standardization of data for 25(OH)D, it appeared that there was a decline in mean serum 25(OH)D from NHANES III (1988-1994) to NHANES (2005-2006) and an increase in the numbers of individuals with 25(OH)D values below 30 nmol/L (12 ng/mL) (4% and 6%, respectively) and individuals below 50 nmol/L (<20 ng/mL) (22% and 31%, respectively)[21]. However, after standardizing results it became clear that there was no change in mean 25(OH)D levels between the two timepoints [21]. Similar findings were made in the German KIGGS study, where post hoc standardization of results changed the percentage of individuals with 25(OH)D levels <30 nmol/L (<12 ng/mL) from 28% to 13%, and <50 nmol/L (<20 ng/mL) from 64% to 47% [22]. In the Danish Health 2006 survey the mean 25(OH)D levels increased from 44
nmol/L to 65 nmol/L (17.6 ng/mL to 26 ng/mL) after standardizing the results [22]. Thus, the effect that retrospective standardization has on the results depends very much on the assays used in the different studies, emphasizing that standardization of assays is highly warranted in order to facilitate comparison of measurements and studies [23].

The strengths of this study are the inclusion of seven different methods that are routinely used for the determination of 25(OH)D in clinical practice whereof one of these (AIA900) has not previously been compared with the other widely used assays. Also, the inclusion of a relatively large number of samples (N=200) from a randomly selected population of patients attending the hospital clinics is a strength thereby increasing validity of the evaluations as compared to only using control samples. Moreover, the inclusion of SRM972a permitted the evaluation of accuracy and bias as compared to the higher order standardized reference material provided by NIST and which is part of the VDSP. The limitations of the study are that the samples were not measured exactly according to the recommendations from the VDSP which states that duplicate measurements should be done on two consecutive days. Our samples were measured in duplicate though not necessarily on two consecutive days as some duplicate determinations were done on the same day. This might have affected the CV% in a positive way. We found that all methods complied with the VDSP requirement for CV%<10%, but the true CV% might be higher if samples were run on separate days. Also, for some of the instruments the assays were not running routinely but were set up on the instrument for this study (Vitros, Liaison XL, AIA900, iSYS) while the other assays were running routinely in the participating labs. Next, the performance of the LC-MS/MS used in this study has not been directly evaluated by a VDSP reference laboratory and may therefore not be considered a true “gold standard”. However, it performs nicely in the DEQAS EQC program (bias of app. 2.5%) and has therefore been used as a comparator in this study though a direct comparison with a VDSP certified LC-MS/MS would have been preferred. This has also affected the possibility of using the
200 patient samples for determining bias for the individual instruments. If the LC-MS/MS had been calibrated using the SRM2972a standard values from the other instruments could have been calibrated to the standard and the bias could have been calculated using all these samples. However, this was not the case, and bias was therefore calculated using the measurements of the human serum-based standard SRM972a.

In conclusion, the majority of commercially available 25(OH)D assays included in this study are still not standardized according to the VDSP protocol only a few assays have achieved to be standardized. As this has potentially impact on both the clinical management of patients and on research and surveillance of vitamin D status in the population, manufacturers of 25(OH)D assays should increase their efforts to optimize the current and future assays to comply with the standardization protocol from the VDSP for the benefit of the patients and the society. However, standardization of assays alone will not necessarily lead to evidence-based consensus on clear cut-off values for optimal vitamin D status as other factors such as lack of understanding of the role of vitamin D in the human physiology may contribute to the confusion.

AUTHOR CONTRIBUTIONS

Concept and design of the study: LNB, NRJ. Conducting experiments: JRH. Analyzing data: LNB, NRJ. Writing, critically reviewing and approving the manuscript: LNB, JRH, SHH, HAM, NRJ.
REFERENCES


FIGURE LEGENDS
Figure 1: Deming regressions/correlations and Bland-Altman plots of comparisons between LC-MS/MS and the other assays
Figure 2: Deming regression/correlations between the assays that complied best/worst with VDSP requirements for bias when analyzing the SRM972a
Table 1. Certified* and reference# values for 25-hydroxyvitamin D in the SRM972a. Values are means from analyses at NIST using ID-LC-MS and ID-LC-MS/MS and from Center for Disease Control (CDC) using ID-LC-MS/MS. The uncertainty listed is an expanded uncertainty about the mean to cover the measurand with 95% confidence.

<table>
<thead>
<tr>
<th>Level 1</th>
<th>ng/mL</th>
<th>nmol/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total 25-hydroxyvitamin D</td>
<td>29.3 ± 1.1</td>
<td>73.3 ± 2.8</td>
</tr>
<tr>
<td>25-hydroxyvitamin D₂</td>
<td>0.54 ± 0.06</td>
<td>1.3 ± 0.2</td>
</tr>
<tr>
<td>25-hydroxyvitamin D₃</td>
<td>28.8 ± 1.1</td>
<td>71.8 ± 2.7</td>
</tr>
<tr>
<td>3-epi-25-hydroxyvitamin D₃</td>
<td>1.81 ± 0.10</td>
<td>4.5 ± 0.2</td>
</tr>
</tbody>
</table>

| Level 2                       |        |         |
| Total 25-hydroxyvitamin D     | 18.9 ± 0.4 | 47.3 ± 1.0 |
| 25-hydroxyvitamin D₂          | 0.81 ± 0.06 | 2.0 ± 0.2 |
| 25-hydroxyvitamin D₃          | 18.1 ± 0.4 | 45.1 ± 1.0 |
| 3-epi-25-hydroxyvitamin D₃    | 1.28 ± 0.09 | 3.2 ± 0.2 |

| Level 3                       |        |         |
| Total 25-hydroxyvitamin D     | 33.2 ± 0.5 | 83.0 ± 1.3 |
| 25-hydroxyvitamin D₂          | 13.3 ± 0.3 | 32.3 ± 0.8 |
| 25-hydroxyvitamin D₃          | 19.8 ± 0.4 | 49.5 ± 1.1 |
| 3-epi-25-hydroxyvitamin D₃    | 1.17 ± 0.14 | 2.9 ± 0.4 |

| Level 4                       |        |         |
| Total 25-hydroxyvitamin D     | 30.0 ± 0.9 | 75.0 ± 2.3 |
| 25-hydroxyvitamin D₂          | 0.55 ± 0.10 | 1.4 ± 0.2 |
| 25-hydroxyvitamin D₃          | 29.4 ± 0.9 | 73.4 ± 2.3 |
| 3-epi-25-hydroxyvitamin D₃    | 26.0 ± 2.2 | 64.8 ± 5.4 |
Table 2: Performance characteristics and analytical specificities for assays evaluated

<table>
<thead>
<tr>
<th>Manufacturer and Instrument</th>
<th>Measuring Principle</th>
<th>Measuring Range$^1$</th>
<th>LLOQ$^{1,2}$</th>
<th>Cross-reactivity of 25(OH)D metabolite</th>
<th>Cross-reactivity of 25(OH)D metabolite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siemens Healthineers ADVIA Centaur XP</td>
<td>Chemiluminescence immunoassay (CLIA)</td>
<td>4.2 - 150</td>
<td>10.5 - 375</td>
<td>10.5</td>
<td>104.5</td>
</tr>
<tr>
<td>TOSOH Bioscience AIA900</td>
<td>Fluorescent enzyme immunoassay (FEIA)</td>
<td>4.0 - 120</td>
<td>10.0 - 300</td>
<td>NA</td>
<td>101.2</td>
</tr>
<tr>
<td>Roche Elecsys Cobas e411 1st generation</td>
<td>Electrochemiluminescence immunoassay (ECLIA)</td>
<td>3.0 - 70</td>
<td>7.50 - 175</td>
<td>12.5</td>
<td>92</td>
</tr>
<tr>
<td>Ortho Clinical Diagnostics Vitros 5600</td>
<td>Chemiluminescence immunoassay (CLIA)</td>
<td>8.0 - 150</td>
<td>20.0 - 375</td>
<td>20.0</td>
<td>104.9</td>
</tr>
<tr>
<td>DiaSorin Liaison XL</td>
<td>Chemiluminescence immunoassay (CLIA)</td>
<td>4.0 - 150</td>
<td>10.0 - 375</td>
<td>NA</td>
<td>100</td>
</tr>
<tr>
<td>IDS-iSYS</td>
<td>Chemiluminescence immunoassay (CLIA)</td>
<td>4.0 - 110</td>
<td>10.0 - 275</td>
<td>8.83</td>
<td>105</td>
</tr>
<tr>
<td>Waters UPLC triple quadrupole LC-MS/MS</td>
<td>Liquid chromatography / tandem mass spectrometry</td>
<td>4.0 - 200</td>
<td>10.0 - 500</td>
<td>10.0</td>
<td>Not relevant</td>
</tr>
</tbody>
</table>

$^1$Adopted from assay inserts.

$^2$Evaluated according to CLSI-protocol (Clinical and Laboratory Standards Institute) EP17-A2.

$^3$SRM 2972: NIST SRM 2972 - 25-Hydroxyvitamin D Calibration Solutions parameters: 25-hydroxyvitamin D$_2$; 25-hydroxyvitamin D$_3$ (2 levels); 3-epi-25-hydroxyvitamin D$_3$.

$^4$SRM 972: NIST SRM 972 - Vitamin D Metabolites in Frozen Human Serum: 25-hydroxyvitamin D$_2$; 25-hydroxyvitamin D$_3$; 24R,25-dihydroxyvitamin D$_3$; 3-epi-25-hydroxyvitamin D$_3$.

$^5$The assay is traceable to the isotope dilution-liquid chromatography/tandem mass spectrometry (ID-LC/MS/MS) 25(OH)D Reference Method Procedure (RMP). The ID-LC/MS/MS is traceable to the NIST SRM 2972.

$^6$The assay is standardized to LC-MS/MS, which has been standardized to the NIST SRM.

$^7$The assay is traceable to in house reference calibrators, which have been value assigned to correlate to samples measured by LC-MS/MS.

$^8$The assay uses Recipe ClinCal® Calibrators which are traceable to NIST SRM972a
Table 3. Bias and coefficient of variation (CV) for the different instruments when compared to certified and reference values of the SRM 972a (NIST) standard, which includes the sum of values for 25-hydroxyvitamin D2 and 25-hydroxyvitamin D3. The “measured” value represents the mean value of duplicate measurements of each reference sample.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Total</th>
<th>VDSP requirement fulfilled?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measured value (nmol/L)</td>
<td>Bias (%)</td>
<td>CV (%)</td>
<td>Measured value (nmol/L)</td>
<td>Bias (%)</td>
<td>CV (%)</td>
</tr>
<tr>
<td>iSYS</td>
<td>74.5</td>
<td>1.6</td>
<td>3.1</td>
<td>49.1</td>
<td>3.7</td>
<td>1.6</td>
</tr>
<tr>
<td>Liaison XL</td>
<td>75.6</td>
<td>3.1</td>
<td>0.5</td>
<td>47.2</td>
<td>-0.3</td>
<td>2.0</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>79.6</td>
<td>8.5</td>
<td>1.3</td>
<td>49.3</td>
<td>4.1</td>
<td>3.6</td>
</tr>
<tr>
<td>AIA900</td>
<td>85.4</td>
<td>16.5</td>
<td>2.3</td>
<td>49.9</td>
<td>5.5</td>
<td>3.6</td>
</tr>
<tr>
<td>Vitros</td>
<td>83.1</td>
<td>13.4</td>
<td>1.6</td>
<td>56.1</td>
<td>18.6</td>
<td>4.1</td>
</tr>
<tr>
<td>Centauer</td>
<td>66.5</td>
<td>-9.2</td>
<td>4.0</td>
<td>34.1</td>
<td>-27.8</td>
<td>9.1</td>
</tr>
<tr>
<td>Cobas</td>
<td>98.9</td>
<td>34.9</td>
<td>0.8</td>
<td>61.3</td>
<td>29.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Total 25(OH)D SRM 972a value (25(OH)D₂+25(OH)D₃)</td>
<td>73.3 ± 2.8 nmol/L</td>
<td>47.3 ± 1.0 nmol/L</td>
<td>83.0 ± 1.3 nmol/L</td>
<td>75.0 ± 2.3 nmol/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRM 972a value for 3-epi-25-(OH)D₃</td>
<td>4.5 ± 0.2 nmol/L</td>
<td>3.2 ± 0.2 nmol/L</td>
<td>2.9 ± 0.4 nmol/L</td>
<td>64.8 ± 5.4 nmol/L</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VDSP: Vitamin D Standardization Program; 25(OH)D₂: 25-hydroxyvitamin D₂; 25(OH)D₃: 25-hydroxyvitamin D₃; 3-epi-25(OH)D₃: 3-epi-25-hydroxyvitamin D₃
Table 4. Coefficients of variation (CV%) for the individual assays.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>CV (%)</th>
<th>VDSP requirement fulfilled?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liaison XL</td>
<td>2.4</td>
<td>Yes</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>3.0</td>
<td>Yes</td>
</tr>
<tr>
<td>iSYS</td>
<td>4.7</td>
<td>Yes</td>
</tr>
<tr>
<td>Vitros</td>
<td>5.2</td>
<td>Yes</td>
</tr>
<tr>
<td>AIA900</td>
<td>5.5</td>
<td>Yes</td>
</tr>
<tr>
<td>Cobas</td>
<td>6.9</td>
<td>Yes</td>
</tr>
<tr>
<td>Centauer</td>
<td>9.0</td>
<td>Yes</td>
</tr>
</tbody>
</table>

VDSP: Vitamin D Standardization Program.
Table 5. Deming regression analysis between methods. Intercept and slope are shown with 95% confidence intervals in brackets. Figures in bold represent significant bias between methods.

<table>
<thead>
<tr>
<th></th>
<th>AIA900</th>
<th>Centauer</th>
<th>Cobas</th>
<th>iSYS</th>
<th>LC-MS/MS</th>
<th>Liaison</th>
<th>Vitros</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIA900</td>
<td>1.30</td>
<td>0.93</td>
<td>1.28</td>
<td>1.21</td>
<td>1.30</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.04 to 1.56)</td>
<td>(0.81 to 1.04)</td>
<td>(1.13 to 1.44)</td>
<td>(1.13 to 1.29)</td>
<td>(1.18 to 1.43)</td>
<td>(1.28 to 1.49)</td>
<td></td>
</tr>
<tr>
<td>Centauer</td>
<td>-29.6</td>
<td>0.74</td>
<td>1.00</td>
<td>0.89</td>
<td>1.01</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-48.1 to -11.1)</td>
<td>(0.64 to 0.84)</td>
<td>(0.87 to 1.13)</td>
<td>(0.73 to 1.05)</td>
<td>(0.92 to 1.12)</td>
<td>(0.83 to 1.25)</td>
<td></td>
</tr>
<tr>
<td>Cobas</td>
<td>-13.8</td>
<td>9.5</td>
<td>1.36</td>
<td>1.29</td>
<td>0.70</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-23.9 to -3.6)</td>
<td>(1.9 to 17.1)</td>
<td>(1.27 to 1.46)</td>
<td>(1.11 to 1.47)</td>
<td>(0.66 to 0.75)</td>
<td>(0.60 to 0.77)</td>
<td></td>
</tr>
<tr>
<td>iSYS</td>
<td>-9.43</td>
<td>14.7</td>
<td>5.8</td>
<td>0.96</td>
<td>1.04</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-18.3 to -0.5)</td>
<td>(7.7 to 21.6)</td>
<td>(0.1 to 11.5)</td>
<td>(0.83 to 1.08)</td>
<td>(0.99 to 1.08)</td>
<td>(0.98 to 1.19)</td>
<td></td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>-13.0</td>
<td>15.4</td>
<td>1.5</td>
<td>-3.8</td>
<td>0.92</td>
<td>1.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-18.3 to -7.6)</td>
<td>(4.6 to 26.3)</td>
<td>(-10.3 to 13.3)</td>
<td>(-11.7 to 4.1)</td>
<td>(0.81 to 1.03)</td>
<td>(1.04 to 1.28)</td>
<td></td>
</tr>
<tr>
<td>Liaison</td>
<td>-12.5</td>
<td>12.5</td>
<td>-0.3</td>
<td>-4.0</td>
<td>-0.0</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-20.2 to -4.8)</td>
<td>(5.2 to 19.7)</td>
<td>(-4.6 to 4.0)</td>
<td>(-6.5 to -1.4)</td>
<td>(-7.0 to 7.0)</td>
<td>(0.98 to 1.15)</td>
<td></td>
</tr>
<tr>
<td>Vitros</td>
<td>-37.3</td>
<td>-3.7</td>
<td>15.2</td>
<td>-22.3</td>
<td>-21.0</td>
<td>-19.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-45.8 to -28.8)</td>
<td>(-19.3 to 11.9)</td>
<td>(7.3 to 23.1)</td>
<td>(-30.2 to -14.5)</td>
<td>(-30.2 to -11.7)</td>
<td>(-25.7 to -12.4)</td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Classification of patients into different categories of vitamin D levels. Green denotes the instrument in each category that categorizes the least patients into that particular category, while red denotes the instrument which categorizes most patients into that particular category.

<table>
<thead>
<tr>
<th>Vitamin D level</th>
<th>AIA900</th>
<th>Centauer</th>
<th>Cobas</th>
<th>iSYS</th>
<th>LC-MS/MS</th>
<th>Liaison</th>
<th>Vitros</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 30 nmol/L</td>
<td>23 (11.5%)</td>
<td>13 (6.7%)</td>
<td>7 (3.5%)</td>
<td>28 (14.3%)</td>
<td>20 (10.0%)</td>
<td>22 (11.0%)</td>
<td>3 (1.5%)</td>
</tr>
<tr>
<td>30-50 nmol/L</td>
<td>35 (17.5%)</td>
<td>23 (11.9%)</td>
<td>23 (11.5%)</td>
<td>33 (16.9%)</td>
<td>26 (13.0%)</td>
<td>36 (18.0%)</td>
<td>23 (11.5%)</td>
</tr>
<tr>
<td>50-250 nmol/L</td>
<td>142 (71%)</td>
<td>157 (80.9%)</td>
<td>169 (84.5%)</td>
<td>135 (68.8%)</td>
<td>154 (77.0%)</td>
<td>142 (71%)</td>
<td>174 (87.0%)</td>
</tr>
<tr>
<td>&gt;250 nmol/L</td>
<td>0 (0%)</td>
<td>1 (0.5%)</td>
<td>1 (0.5%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>
Table 7. \( \kappa \) values for the strength of agreement between assays. \( \kappa <0.20 \) (poor), 0.21-0.4 (fair), 0.41-0.60 (moderate), 0.61-0.80 (good), 0.81-1.00 (very good). Figures in bold represent very good agreement.

<table>
<thead>
<tr>
<th></th>
<th>AIA900</th>
<th>Centauer</th>
<th>Cobas</th>
<th>iSYS</th>
<th>LC-MS/MS</th>
<th>Liaison</th>
<th>Vitros</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIA900</td>
<td></td>
<td>0.561</td>
<td>0.530</td>
<td>0.780</td>
<td>0.749</td>
<td>0.828</td>
<td>0.406</td>
</tr>
<tr>
<td>Centauer</td>
<td>0.561</td>
<td></td>
<td>0.794</td>
<td>0.512</td>
<td>0.653</td>
<td>0.567</td>
<td>0.619</td>
</tr>
<tr>
<td>Cobas</td>
<td>0.530</td>
<td>0.794</td>
<td></td>
<td>0.483</td>
<td>0.595</td>
<td>0.431</td>
<td>0.610</td>
</tr>
<tr>
<td>iSYS</td>
<td>0.780</td>
<td>0.512</td>
<td>0.483</td>
<td></td>
<td>0.743</td>
<td>0.850</td>
<td>0.366</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>0.749</td>
<td>0.653</td>
<td>0.595</td>
<td>0.743</td>
<td></td>
<td>0.791</td>
<td>0.478</td>
</tr>
<tr>
<td>Liaison</td>
<td>0.828</td>
<td>0.567</td>
<td>0.556</td>
<td>0.850</td>
<td>0.791</td>
<td></td>
<td>0.367</td>
</tr>
<tr>
<td>Vitros</td>
<td>0.406</td>
<td>0.619</td>
<td>0.610</td>
<td>0.366</td>
<td>0.478</td>
<td>0.367</td>
<td></td>
</tr>
</tbody>
</table>