Papers in Press. Published January 9, 2012 as doi:10.1373/clinchem.2011.172155
The latest version is at http://hwmaint.clinchem.org/cgi/doi/10.1373/clinchem.2011.172155

Clinical Chemistry 58:3
000–000 (2012)

State-of-the-Art Vitamin D Assays: A Comparison of Automated Immunoassays with Liquid Chromatography–Tandem Mass Spectrometry Methods

Christopher-John L. Farrell,[1,2] Steven Martin,[1] Brett McWhinney,[3] Isabella Straub,[4] Paul Williams,[5] and Markus Herrmann[1,4*]

BACKGROUND: Vitamin D testing is increasing worldwide. Recently several diagnostic manufacturers including Abbott and Siemens have launched automated 25-hydroxy vitamin D (25OH-D) immunoassays. Furthermore, preexisting assays from DiaSorin and Roche have recently been modified. We compared the performance of 5 automated immunoassays, an RIA and 2 liquid chromatography–tandem mass spectrometry (LC-MS/MS) methods.

METHODS: Aliquots of 170 randomly selected patient samples were prepared and 25OH-D was measured by 2 LC-MS/MS methods, an RIA (DiaSorin, and automated immunoassays from Abbott (Architect), DiaSorin (LIAISON), IDS (ISYS), Roche (E170, monoclonal 25OH-D_3 assay), and Siemens (Centaur). Within-run and between-run imprecision were evaluated by measurement of 5 replicates of 2 serum pools on 5 consecutive days.

RESULTS: The LC-MS/MS methods agreed, with a concordance correlation coefficient (CCC) of 0.99 and bias of 0.56 μg/L (1.4 nmol/L). The RIA assay showed a performance comparable to LC-MS/MS, with a CCC of 0.97 and a mean bias of 1.1 μg/L (2.7 nmo/L). All immunoassays measured total 25OH-D (including D_3 and D_2), with the exception of the Roche assay (D_3 only). Among the immunoassays detecting total 25OH-D, the CCCs varied between 0.85 (Abbott) to 0.95 (LIAISON). The mean (SD) bias ranged between 0.2 (0.5) (LIAISON) and 4.56 (11.4) (Abbott) μg/L (nmol/L). The Roche 25OH-D_3 assay demonstrated small mean bias (−2.7 μg/L to −6.7 nmol/L) but a low CCC of just 0.66. Most assays demonstrated good intra- and interassay precision, with CV <10%.

CONCLUSIONS: Automated immunoassays demonstrated variable performance and not all tests met our minimum performance goals. It is important that laboratories be aware of the limitations of their assay.

© 2011 American Association for Clinical Chemistry

An increasing recognition of the high prevalence and manifold consequences of vitamin D deficiency *(1–2)* has caused a massive rise in vitamin D testing worldwide. In Australia, for example, requests for vitamin D have escalated from approximately 23 000 tests in 2000 to 2.2 million in 2010 *(3)*. To cope with such a workload laboratories require reliable automated assays.

25-hydroxy vitamin D (25OH-D)[6] is the predominant circulating form of vitamin D and is generally considered to be the best single marker of vitamin D status *(1, 4)*. There are 2 types of 25OH-D that can be found in the circulation, the endogenously derived 25-OH vitamin D_3 (cholecalciferol, 25OH-D_3) and 25-OH vitamin D_2 (ergocalciferol, 25OH-D_2), which is derived from plant sources and fish *(4–5)*. Normally, 25OH-D_3 accounts for approximately 95% of the total circulating 25OH-D pool, whereas 25OH-D_2 represents a minor fraction unless vitamin D_2–containing supplements are used by the patient *(5–8)*.

The measurement of 25OH-D can be performed by use of immunoassay, HPLC, and liquid chromatography–tandem mass spectrometry (LC-MS/MS) *(9–10)*. Automated 25OH-D immunoassays from various manufacturers including DiaSorin and Roche Diagnostics have been available for some time, but until recently the accuracy and precision of some of these tests were unsatisfactory *(9, 11–13)*. There are 2 main difficulties in establishing an automated immunoassay for 25OH-D:

[1] Laverty Pathology, North Ryde, New South Wales, Australia; [2] PaLMS Pathology, Royal North Shore Hospital, New South Wales, Australia; [3] Pathology Queensland, Queensland, Australia; [4] University of Sydney, Central and Nepean Clinical School, Royal Prince Alfred Hospital, New South Wales, Australia; [5] University of Sydney, Central Clinical School and Sydney South West Pathology Service, Royal Prince Alfred Hospital, New South Wales, Australia.
* Address correspondence to this author at: University of Sydney, Department of Biochemistry, Royal Prince Alfred Hospital, Missenden Rd., Camperdown, NSW 2050, Australia. Fax +61-2-9515-7931; e-mail markusherr@aol.com.
Received July 22, 2011; accepted December 2, 2011.
Previously published online at DOI: 10.1373/clinchem.2011.172155
[6] Nonstandard abbreviations: 25OH-D, 25-hydroxy vitamin D; LC-MS/MS, liquid chromatography–tandem mass spectrometry; VDBP, vitamin D binding protein; SRM, standard reference material; PQ, Pathology Queensland; CCC, concordance correlation coefficient; Cb, bias correction factor.

Copyright (C) 2012 by The American Association for Clinical Chemistry

in blood the strongly hydrophobic 25OH-D is largely bound to vitamin D–binding protein (VDBP) *(14)*, so generating antibodies against small antigenic molecules, such as 25OH-D, is challenging. To add additional complexity the US Food and Drug Administration stipulates that 25OH-D assays must detect both 25OH-D_2 and 25OH-D_3. Current 25OH-D immunoassays employ polyclonal or monoclonal antibodies directed against 25OH-D. However, competition between the 25OH-D capture antibody and VDBP in patient samples makes these assays difficult to control. This is of particular relevance in regard to homogenous 1-step assays, in which 25OH-D and VDBP are not completely separated. Such assays have been shown to agree poorly with higher-order methods, such as LC-MS/MS *(12)*.

Although isotope dilution LC-MS/MS can be considered the gold standard for the analytical measurement of small molecules, a generally accepted reference method for 25OH-D was lacking until recently. In 2010 Tai et al. developed a candidate reference method that in 2011 has been recognized as a reference method by the Joint Committee for Traceability in Laboratory Medicine *(15)*. Another candidate reference method has recently been published by the Laboratory for Analytical Chemistry at Ghent University *(16)*. In addition, the first standard reference material for 25OH-D was not introduced until 2008. Consequently, in the past different LC-MS/MS methods have been shown not to be generally in agreement *(17)*. The recent release of an NIST standard reference material (SRM 972) is anticipated to improve the analytical performance of 25OH-D measurements and to facilitate harmonization across all forms of 25OH-D assays.

Recently several diagnostic companies, including Siemens, IDS, and Abbott, have launched automated 25OH-D assays. Others, such as Roche and DiaSorin, have modified or are in the process of modifying their assays. Considering the difficulties encountered with automated 25OH-D assays in the past, it is of interest to see if the latest generation of 25OH-D assays represents an improvement and if their performance meets the needs of clinical laboratories. Therefore we compared 5 automated immunoassays, an established RIA, and 2 independent LC-MS/MS methods for the measurement of 25OH-D.

Material and Methods

STUDY DESIGN

We randomly selected 170 serum samples from routine vitamin D assay requests. These samples were observed to display an even dispersion of vitamin D concentrations between 2 and 60 μg/L (5 and 151 nmol/L) as measured by LC-MS/MS. Samples were divided into 6 aliquots, stored at −20 °C, and analyzed in batches, with a freshly thawed aliquot used for each analytical run. We used 2 LC-MS/MS methods to measure 25OH-D, a commercial RIA (DiaSorin) and 5 automated chemiluminescent immunoassays from Abbott Diagnostics (Architect), DiaSorin (LIAISON), IDS (ISYS), Roche Diagnostics (E170), and Siemens (Centaur). The DiaSorin LIAISON kit was a premarket evaluation assay with demonstrated performance similar to the currently marketed assay. The Roche assay used was the monoclonal vitamin D_3 assay, which specifically detected 25OH-D_3. At the time of the study this test was used in our laboratory and served to select the samples for this study. For full methodological details and performance characteristics of all assays as provided by the respective manufacturers/laboratories see Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol58/issue3.

In addition, 2 serum pools were prepared for the assessment of assay precision. The 2 LC-MS/MS methods demonstrated mean 25OH-D concentrations of 11.6 μg/L and 32.8 μg/L (29 nmol/L and 82 nmol/L), respectively, for the 2 pools, with no appreciable concentrations of 25OH-D_2. Multiple aliquots from both pools were prepared and stored at −20 °C. On 5 consecutive days a freshly thawed aliquot of each pool was assayed 5 times with all assays. The same experiment was repeated several months later for all except the Roche assay, which had been discontinued by that time.

Measurements were performed in a blinded fashion in 4 different laboratories with a different operator for each method. Measurements with the methods from IDS, Siemens, Abbott, and Roche were performed at Laverty Pathology (North Ryde, Australia). The DiaSorin LIAISON assay was performed at the department of biochemistry of the Royal Prince Alfred Hospital (Camperdown, Australia). The research and development laboratory of DiaSorin (Stillwater, MN) analyzed the samples with RIA and 1 of the 2 LC-MS/MS methods. Pathology Queensland (PQ) (Brisbane, Australia) provided the second LC-MS/MS method.

MEASUREMENT OF 25OH-D BY LC-MS/MS

The 2 LC-MS/MS methods were noncommercial assays from 2 independent laboratories in Australia (PQ) and the US (DiaSorin). The principles of both procedures are described below. Details regarding instrument settings, calibrators, controls, and internal standards are provided as supplemental data (see online Supplemental Data Table 2).

PQ LC-MS/MS METHOD

Serum samples, calibrators, and controls were placed on a robotic liquid-handling system (Tecan, Freedom EVO150). To precipitate protein and dissociate vitamin D from VDBP, 150 μL of sample, 150 μL of 0.2 mol/L zinc sulfate, and then 500 μL of 100% methanol containing 50 nmol/L of deuterated internal standards were added to a 2-mL square 96-well collection plate. After centrifugation the supernatant was transferred to a conditioned Oasis μElutionHLB solid-phase extraction plate. This was washed with 60% methanol and the retained analytes were eluted with a 2-step protocol that matched the organic strength of the initial conditions of chromatography. The elution plate was sealed and transferred to the autosampler. Chromatographic separation of the samples was performed on a Waters ACQUITY UPLC system equipped with a Waters ACQUITY BEH 2.1 × 50–mm C8 1.7-μm column. Mobile phase A was water with 2 mmol/L ammonium acetate and 0.1% formic acid, mobile phase B consisted of methanol with 2 mmol/L ammonium acetate and 0.1% formic acid. We injected 20 μL of sample into the UPLC system using a column temperature of 45 °C and a flow rate of 0.4 mL/min. After injection of the sample, the system was run isocratically for 1.2 min (mobile phase A 27%, mobile phase B 73%) after which a linear gradient with an increasing fraction of mobile phase B (from 73% to 98%) and decreasing fraction of mobile phase A (from 27% to 2%) was applied until 3.2 min. The eluent of the UPLC system was introduced in a Waters XE Premier mass spectrometer set in positive electrospray mode for the quantification of 25OH-D. Results were processed using MassLynx 4.1 software (Waters).

DIASORIN LC-MS/MS METHOD

We mixed 150 μL of samples, calibrators, and controls with 150 μL of 0.2-mol/L zinc sulfate followed by 300 μL methanol containing 25 μg/L of deuterated internal standards. After thorough mixing, 750 μL of hexane was added and samples were mixed again. Following centrifugation, 650 μL of the hexane phase was pipetted into glass vials and dried under nitrogen at 55 °C for 5 min. The dried residue was reconstituted in 75 μL methanol/water (65/35 v/v) and loaded onto a Waters ACQUITY UPLC system. Chromatographic separation was performed using a Waters ACQUITY BEH 2.1 × 50–mm phenyl 1.7-μm column with a precolumn filter. Mobile phase A was water with 2 mmol/L ammonium acetate and 0.1% formic acid; mobile phase B consisted of methanol with 2 mmol/L ammonium acetate and 0.1% formic acid. We injected 20 μL of sample into the UPLC system using a column temperature of 35 °C and a flow rate of 0.45 mL/min. After injection of the sample, a gradient with an increasing fraction of mobile phase B (from 65% to 85%) and decreasing fraction of mobile phase A (from 35% to 15%) was applied over a period of 3.0 min. The eluent was introduced into a Waters TQD tandem quadropole detection system set in positive electrospray mode. Results were processed by use of MassLynx 4.1 software (Waters).

STATISTICS

The results of the 170 serum samples were analyzed by concordance correlation coefficient (CCC), Passing–Bablok regression, Bland–Altman plots, multiline plots, and mountain plots.

CONCORDANCE CORRELATION COEFFICIENT

The concordance correlation coefficient (CCC) *(18, 19)* is used to evaluate the degree to which pairs of observations fall on the 45° line through the origin. It contains a measurement of precision (Pearson correlation coefficient r) and accuracy [bias correction factor (Cb)] and is calculated as follows: $CCC = r \times Cb$. The Pearson correlation coefficient measures how far each observation deviates from the best-fit line. The Cb measures how far the best-fit line deviates from the 45° line through the origin. Interpretation of CCC results was as follows: >0.99, excellent agreement; 0.99–0.95, substantial agreement; 0.90–0.94, moderate agreement; <0.9 poor agreement.

PASSING BABLOK REGRESSION

Passing Bablok Regression *(20)* calculates a regression equation ($y = a + bx$) including 95% CIs for the constant (a) and proportional bias (b). This procedure requires continuous variables and a linear relationship between the 2 methods. We tested the assumption of linearity by using the cumulative sum linearity test (cusum linearity test). The cusum test is used to assess whether residuals are randomly scattered above and below the regression line and do not exhibit any distinct trend. A P value <0.05 indicates a significant deviation from linearity.

INTRAASSAY PRECISION

Intraassay precision was assessed by calculating the mean, SD, and CV of 5 replicates from each of the 2 serum pools measured on 5 consecutive days. Interassay precision was assessed by calculating the SD and CV of the daily means of the 5 testing days. The precision experiment was repeated on a second occasion several months after the first experiment and results of each experiment were analyzed individually and in a combined fashion.

We performed all calculations using Medcalc version 11.5.1.0.

Results

LC-MS/MS PQ vs LC-MS/MS DiaSorin

A comparison of the 2 LC-MS/MS methods against the mean values is shown in Table 1 and Fig. 1. The Passing–Bablok regression analysis illustrates close agreement between the 2 methods for the entire cohort (Fig. 1A) and samples with a 25OH-D concentration of <8 μg/L (20 nmol/L) (Fig. 1B). Both methods agreed, with a CCC of 0.99 (95% CI 0.993–0.996 for DiaSorin, 0.992–0.996 for PQ), a Pearson coefficient of 0.99 and a bias correction of 0.99. The mean bias with respect to the mean of the 2 LC-MS/MS methods was +0.56 (1.96 SD, −1.4 to 2.48) μg/L [+1.4 (1.96 SD, −3.5 to 6.2) nmol/L] for the DiaSorin LC-MS/MS method (Fig. 1C) and −0.56 (1.96 SD, −2.48 to 1.4) μg/L [−1.4 (1.96 SD, −6.2 to 3.5) nmol/L] for the PQ LC-MS/MS method (Fig. 1E). For values <8 μg/L (20 nmol/L) the mean bias was −0.12 (1.96 SD, −0.84 to 0.6) μg/L [−0.3 (1.96 SD, −2.1 to 1.5) nmol/L] and 0.12 (1.96 SD, −0.6 to 0.84) μg/L [0.3 (1.96 SD −1.5 to 2.1) nmol/L], for the DiaSorin (Fig. 1D) and PQ (Fig. 1F) methods, respectively.

Because of the agreement of the results from the 2 LC-MS/MS methods we used the mean value from the 2 assays for all the comparisons.

Immunoassays vs LC-MS/MS

The mean values for the sample cohort (over all 170 samples) of all immunoassays varied from 14.4 to 21.6 μg/L (36 to 54 nmol/L). When compared with the LC-MS/MS run mean of 16.8 μg/L (42 nmol/L) the Abbott assay showed the greatest deviation (+28%; Fig. 2). The individual distribution of the data points illustrates significant differences between assays for these samples (Fig. 2). The Bland–Altman plots in Fig. 3 show that the mean bias over all samples was lowest for the LIAISON assay [0.2 μg/L (0.5 nmol/L)] and highest for the Abbott assay [4.6 μg/L (11.4 nmol/L)] compared to LC-MS/MS. The Passing–Bablok regression analyses revealed proportional bias for all immunoassays, with Siemens and Roche assays showing the most obvious effects (Fig. 3, Table 1). In addition, constant bias was observed in all immunoassays, with Siemens having the highest constant bias of 5.9 μg/L (14.8 nmol/L). The Siemens assay was the only immunoassay for which the line of best fit crossed the line of identity, indicating a high bias at concentrations below 16 μg/L (40 nmol/L) and a low bias at concentrations above this threshold. CCC, Pearson coefficients for precision, and the bias correction coefficients were ≥0.90 for all immunoassays except for Abbott (CCC = 0.85) and Roche (CCC = 0.66, Pearson coefficient = 0.68).

The performance of the different immunoassays was most variable at 25OH-D concentrations <8 μg/L (20 nmol/L) (Fig. 4). The RIA showed the lowest mean bias and the narrowest 95% CI compared to LC-MS/MS (Fig. 4, A and G). Among the automated immunoassays the mean bias ranged from 1.0 to 5.2 μg/L (2.4–13.0 nmol/L) (17% to 118%) with 95% CIs varying from 6.2 to 10.8 μg/L (15.4–27.0 nmol/L) (123% to 285%).

Total 25OH-D concentrations in the 2 serum pools used for precision studies were 11.6 and 32.8 μg/L (29 and 82 nmol/L) with no appreciable quantities of 25OH-D_2, as measured by LC-MS/MS. The best overall precision was achieved by the PQ LC-MS/MS method (Table 2). All assays showed a within-run precision of ≤10% with the exception of Roche (12.1% for the low pool). The combined between-run precision was ≤15% for all assays except for Roche (19%), for which only 1 set of results was available. For the first experiment the between run-precision of the LIAISON test also exceeded 15% but improved to 6.5% in the second experiment.

Discussion

The present study showed the excellent concordance of the values obtained by the 2 LC-MS/MS methods despite having different sample preparation and extraction procedures. Immunoassays demonstrated variable performance and not all assays demonstrated the ability to meet the needs of clinical laboratories. Only the RIA assay achieved a performance that was comparable to LC-MS/MS.

Clinical laboratories can apply performance goals based on biological variation to decide if a 25OH-D assay is analytically acceptable *(21)*. With this approach, the minimum requirements can be calculated as a mean bias ≤15.8% and imprecision ≤9.1% *(22)*. At 50 nmol/L, the recommended cutoff for vitamin D deficiency *(23)*, this corresponds to a bias of approximately 8 nmol/L. In addition, our assessment of assay acceptability also included the CCC with an arbitrary cutoff of 0.9. Among the automated immunoassays the LIAISON showed superior agreement (mean bias +6.4%, CCC = 0.95), and the IDS assay (bias of 14% and a CCC of 0.90) showed acceptable agreement with the 2 LC-MS/MS methods. Although Roche had a low mean bias (−6.9%), the CCC of 0.66 indicates poor concordance with LC-MS/MS. Both Abbott (+41%) and Siemens (+27%) showed excessive bias. In addition the Abbott assay had an unacceptable concordance with LC-MS/MS (CCC = 0.85).

Before the release of the NIST SRM 972 for 25OH-D in 2008, LC-MS/MS methods showed substantial disagreement. This was illustrated by the results from the DEQAS (Vitamin D Quality Assessment Scheme) external quality assurance scheme (October

Table 1. Passing–Bablok and concordance correlation analysis of all 25OH-D methods against the mean of the 2 LC-MS/MS methods.

	Passing–Bablok regression analysis								Concordance correlation analysis			
Method	Intercept	95% CI	Slope	95% CI	Minimum	Maximum	SD	Cusum test	CCC	95% CI	*r*	Cb
All samples												
RIA	2.8	2.1 to 3.6	0.87	0.85 to 0.90	6.4	123.4	22.6	Yes	0.968	0.959 to 0.974	0.985	0.983
Abbott	5.8	2.9 to 8.5	1.1	1.03 to 1.19	16.5	195.1	32.4	No	0.848	0.809 to 0.880	0.931	0.910
IDS	−2.6	−4.5 to −0.6	1.2	1.15 to 1.27	5	246.3	34.7	No	0.898	0.874 to 0.919	0.954	0.942
LIAISON	4.6	2.8 to 5.8	0.88	0.84 to 0.93	5	136	24	Yes	0.949	0.933 to 0.962	0.954	0.995
Siemens	14.8	13.2 to 16.8	0.68	0.64 to 0.73	14.5	138.2	20.5	No	0.908	0.885 to 0.928	0.942	0.955
Roche	3.0	−1.1 to 5.1	0.75	0.67 to 0.85	5	209	28.5	No	0.657	0.566 to 0.733	0.679	0.968
LC-MS/MS DiaSorin	−0.79	−1.3 to −0.4	1.05	1.04 to 1.06	4.5	153.5	27.6	Yes	0.995	0.993 to 0.996	0.997	0.998
LC-MS/MS PQ	0.6	0.3 to 1.1	0.96	0.94 to 0.97	5	149.5	25.6	Yes	0.994	0.992 to 0.996	0.996	0.998
Samples >8 μg/L (20 nmol/L)												
RIA	2.9	1.5 to 4.4	0.87	0.83 to 0.90	17.9	123.4	20.6	Yes	0.952	0.937 to 0.964	0.978	0.974
Abbott	−2	−7.4 to 2.5	1.25	1.14 to 1.36	18.4	195.1	32.0	No	0.820	0.771 to 0.860	0.922	0.890
IDS	−6.8	−10.9 to −3.5	1.3	1.20 to 1.37	17.6	246.3	33.6	Yes	0.863	0.827 to 0.892	0.943	0.916
LIAISON	2.1	−0.8 to 5.1	0.93	0.86 to 1.00	18.9	136	22.6	Yes	0.932	0.906 to 0.951	0.937	0.995
Siemens	10.70	7.7 to 13.1	0.77	0.70 to 0.83	22.7	138.2	20.3	No	0.918	0.890 to 0.939	0.934	0.983
Roche	6.9	−14.0 to 1.5	0.93	0.80 to 1.07	5	209	29.9	No	0.578	0.464 to 0.673	0.625	0.924
LC-MS/MS DiaSorin	−0.79	−1.3 to −0.4	1.05	1.04 to 1.06	4.5	153.5	27.6	Yes	0.992	0.989 to 0.994	0.995	0.997
LC-MS/MS PQ	0.6	0.3 to 1.1	0.96	0.94 to 0.97	5	149.5	25.6	Yes	0.991	0.988 to 0.994	0.995	0.997

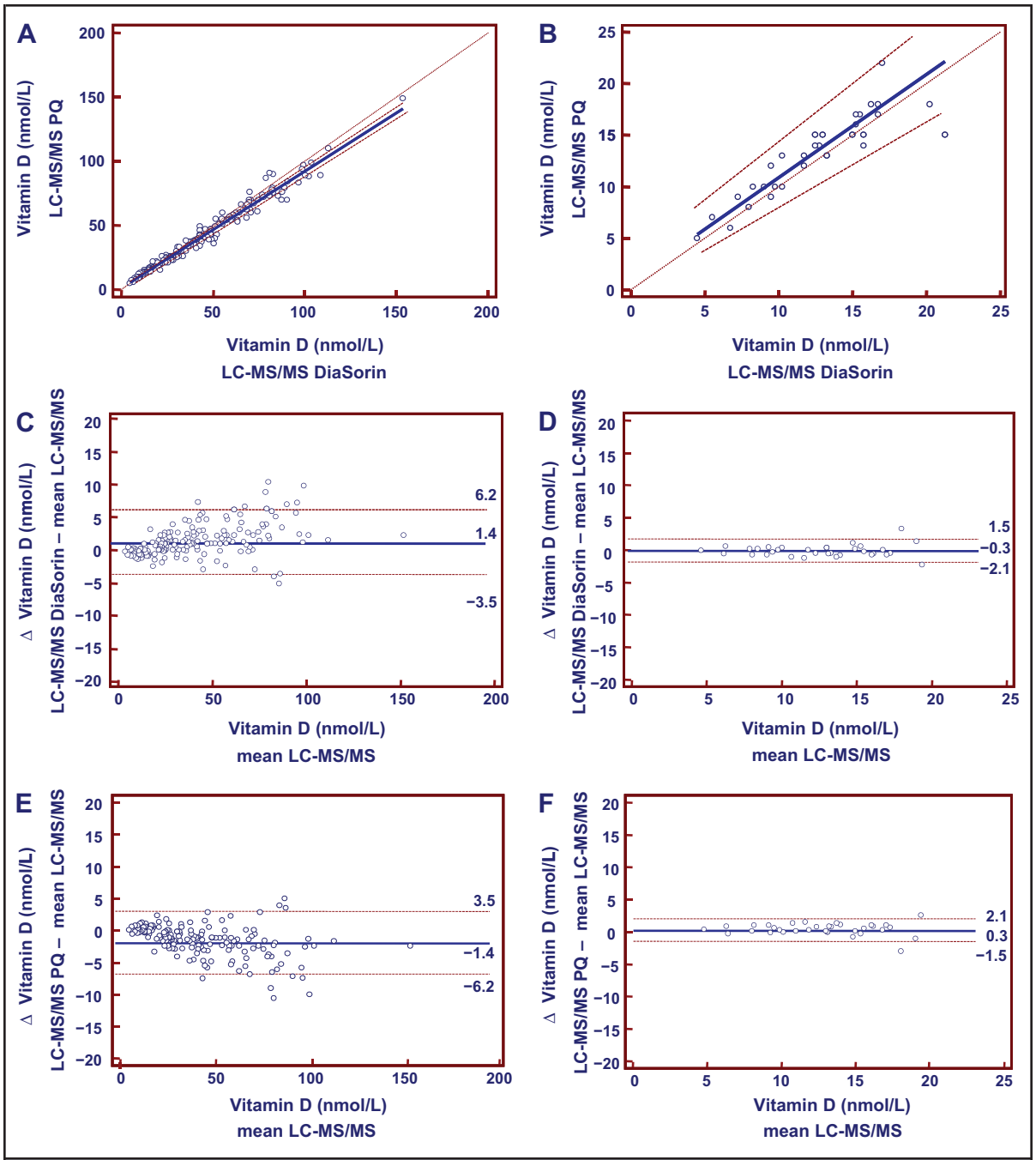

Fig. 1. Comparison of the 2 LC-MS/MS methods from DiaSorin and PQ.

(A), Passing–Bablok regression analysis for all samples; (B), Passing–Bablok regression analysis for samples with a 25OH-D concentration of less than 8 μg/L (20 nmol/L); (C), Bland–Altman plot showing the bias between the DiaSorin LC-MS/MS method and the mean of the 2 LC-MS/MS methods for all samples; (D), Bland–Altman plot showing the bias between the DiaSorin LC-MS/MS method and the mean of the 2 LC-MS/MS methods for samples with a 25OH-D concentration of less than 8 μg/L (20 nmol/L); (E) Bland–Altman plot showing the bias between the PQ LC-MS/MS method and the mean of the 2 LC-MS/MS methods for all samples; (F), Bland–Altman plot showing the bias between the PQ LC-MS/MS method and the mean of the 2 LC-MS/MS methods for samples with a 25OH-D concentration of less than 8 μg/L (20 nmol/L). To convert 25OH-D concentrations to micrograms per liter, multiply by 0.4.

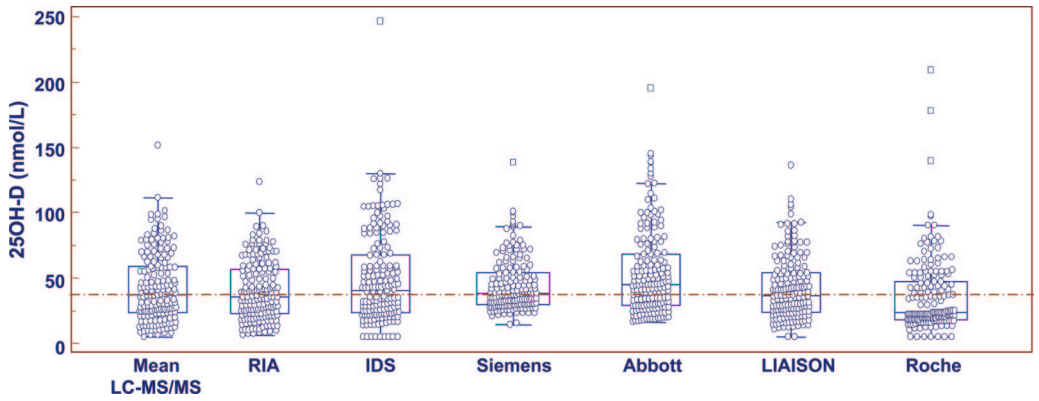

Fig. 2. Box-and-whisker plot showing the distribution of results for all assays tested.

The central boxes represent the 25th to 75th percentile range. The lines inside the boxes show the median value for each method. The whiskers extend from the minimum to the maximum value, excluding outliers. An outlier value is defined as a value that exceeds the upper or lower quartile plus or minus 1.5 times the interquartile range. The dashed horizontal line shows the median value as measured by LC-MS/MS. To convert 25OH-D concentrations to micrograms per liter, multiply by 0.4.

2008 distribution), which demonstrated interlaboratory CVs for LC-MS/MS methods commensurate with that of immunoassays *(24)*. This observation was further supported by a comparison of 3 different LC-MS/MS methods reported by Binkley et al. *(17)*. Both of the LC-MS/MS methods used in this study were aligned to the NIST SRM 972, and their excellent concordance across the range of 2–60.4 μg/L (5–151 nmol/L) provides good evidence that the release of this standard had indeed helped to improve assay comparability.

The outstanding performance of the DiaSorin RIA confirmed the findings of a previous study *(12)* and was most likely attributable to the acetonitrile extraction step used in this assay. This step released all 25OH-D from VDBP and eliminated important sources of interference, including heterophile antibodies, before incubation with the capture antibody. Automated immunoassays cannot use such an aggressive extraction step, which may explain the variable performance of these assays. A good example was the LIAISON assay, which employed exactly the same capture antibody as the RIA, but the LIAISON results showed more scatter compared to LC-MS/MS. However, compared to the other automated immunoassays, the LIAISON exhibited better concordance with LC-MS/MS results. The substantial differences in assay performance among the automated immunoassays clearly demonstrated that not all manufacturers were equally successful in addressing the technical challenges related to the measurement of 25OH-D.

Most immunoassays had difficulties measuring low concentrations. For example, below 8 μg/L (20 nmol/L) Siemens and Abbott showed excessive bias, with means of 118% and 105%, respectively. Given that Abbott recommends not reporting quantitative results of <8 μg/L (20 nmol/L), this excessive bias does not present a serious problem when the assay is used according to the manufacturer's specifications. Roche (+35%), LIAISON (+35%), and IDS (+17%) also failed to meet the minimum performance goal for bias (≤15.8%). Only the RIA (mean bias +10%) was able to provide reliable results down to 2 μg/L (5 nmol/L).

The relatively poor performance of the automated immunoassays at 25OH-D concentrations <8 μg/L (<20 nmol/L) had an adverse effect on their overall performance. When we considered only samples >8 μg/L (>20 nmol/L) the mean bias of most assays improved, but only LIAISON (+1%), Siemens (+4%), and IDS (+13%) met the minimum performance goal (≤15.8%). From a clinical point of view 25OH-D concentrations <8 μg/L (<20 nmol/L) are clearly deficient. Therefore inaccuracies at very low concentrations have a limited impact on treatment decisions but may influence the lower limit to which laboratories report.

Based on the biological variation–derived minimum requirement for assay precision the within-run precision was satisfactory for all immunoassays with the exception of the Roche assay, which showed excessive imprecision for the high pool serum. In contrast, several immunoassays did not meet the minimum requirement for between-run precision. During the first precision experiment the between-run precision of LIAISON, Siemens, and Roche exceeded 9.1%. However, the second experiment demonstrated that all im-

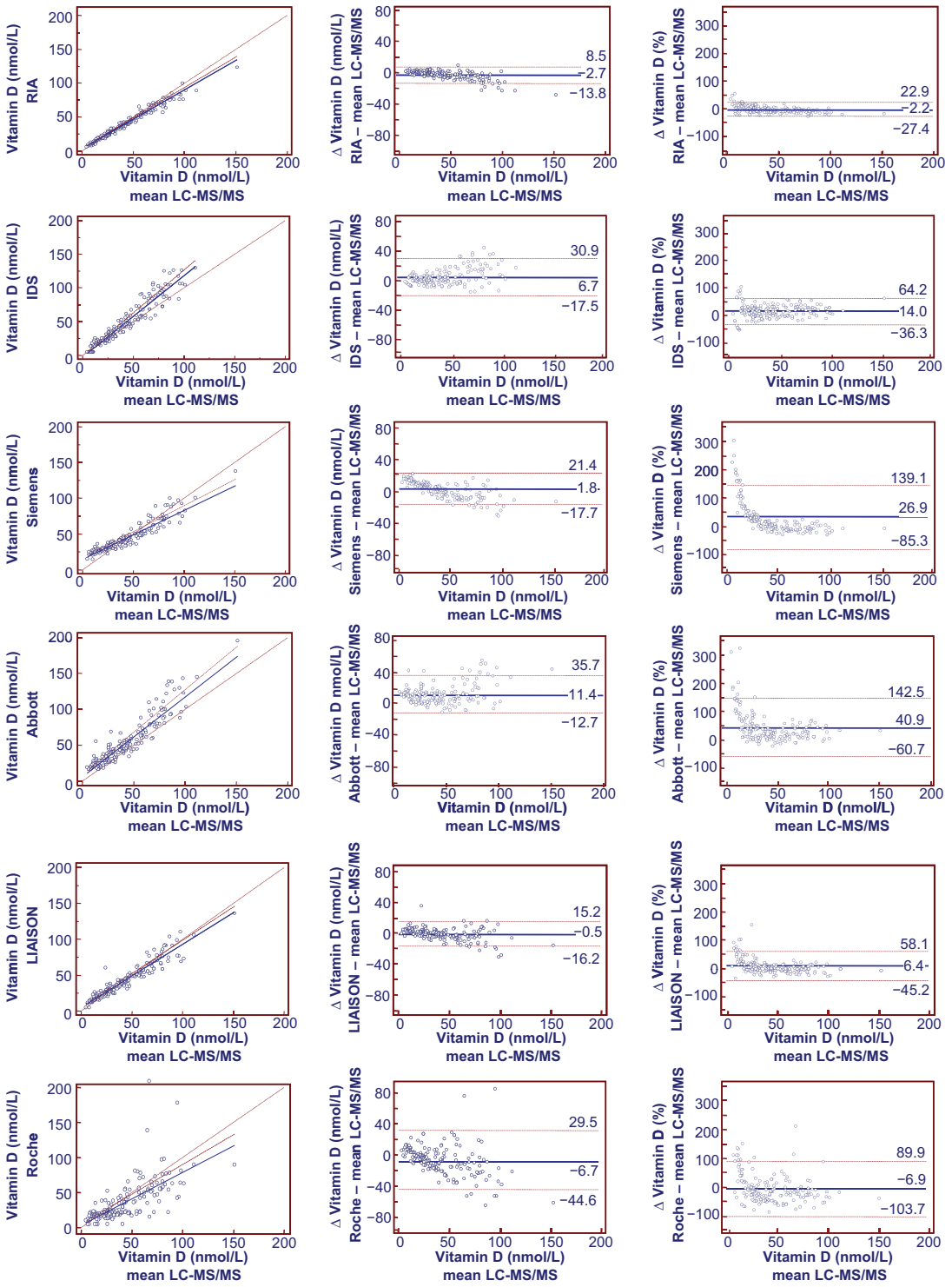

Fig. 3. Comparison of all immunoassays against LC-MS/MS by Passing–Bablok regression analysis (left panels) and Bland–Altman plots (middle and right panels).

The Bland–Altman plots show bias in nanomoles per liter (middle panels) and percentage (right panels). To convert 25OH-D concentration to micrograms per liter, multiply by 0.4.

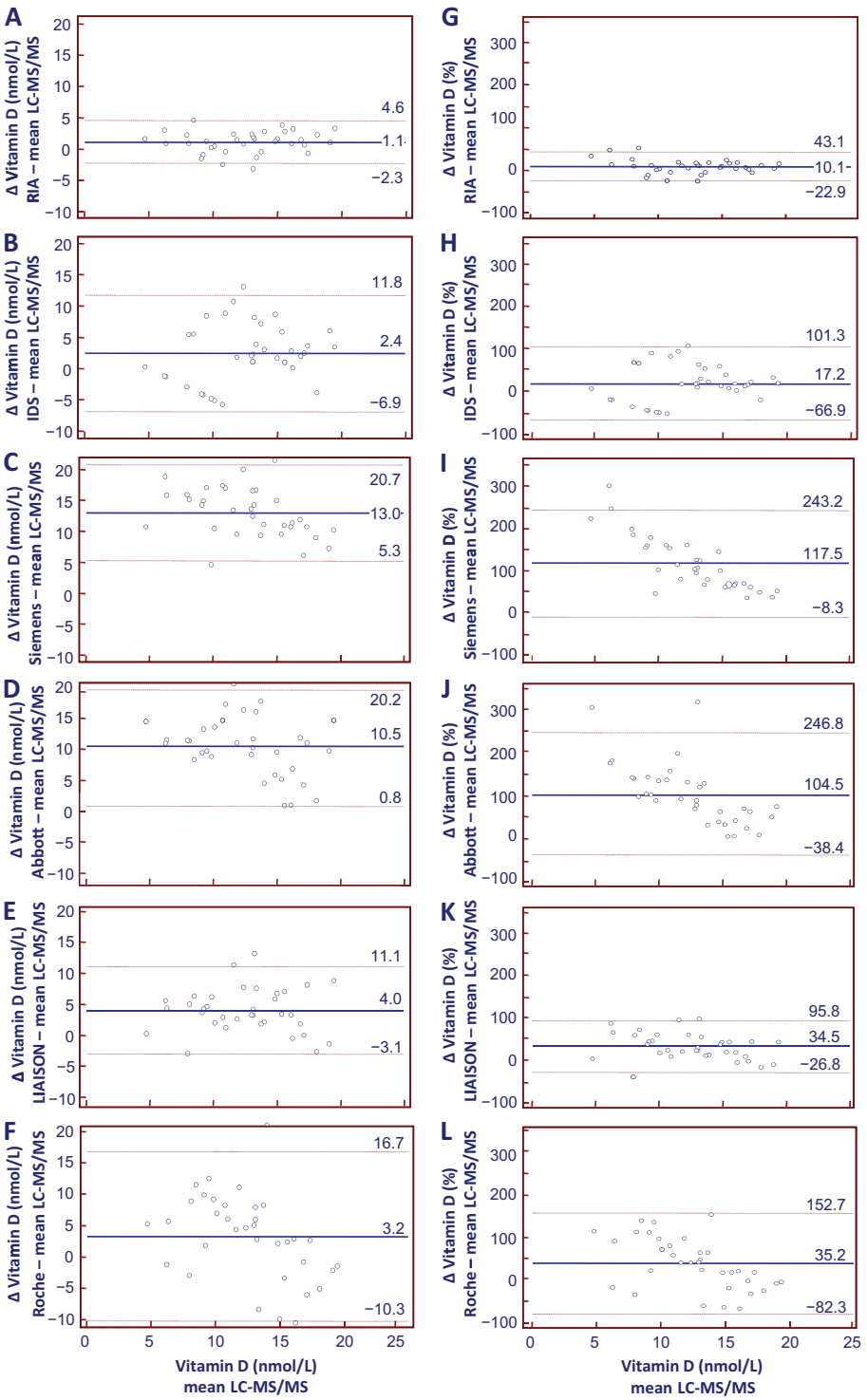

Fig. 4. Bland–Altman plots showing the bias of all assays against the mean LC-MS/MS result for samples with a 25OH-D concentration <8 μg/L (<20 nmol/L) as assessed by LC-MS/MS.

Bias is shown in nanomoles per liter (left panels) and percentage (right panels). To convert 25OH-D concentrations to micrograms per liter, multiply by 0.4.

Table 2. Precision studies for all 25OH-D methods performed by using a low and a high serum pool.

Assay	Mean 25OH-D, nmol/L[a]	Range of daily 25OH-D means, nmol/L[a]	Within-run CV, %			Between-run CV, %		
			Experiment 1	Experiment 2	Experiment 1 + 2 combined	Experiment 1	Experiment 2	Experiment 1 + 2 combined
Low pool								
Abbott	34	29–37	5.4	3.0	4.2	6.3	4.7	6.9
IDS	30	26–34	2.0	5.6	3.8	7.4	6.7	8.8
Liaison	25	22–35	9.7	7.9	8.8	18.3[b]	6.5	*15.2*[a]
RIA	28	23–36	5.5	5.4	5.5	7.9	6.3	14.7
Roche	15	10–17	12.1	—	—	19.0[b]	—	—
Siemens	30	24–36	6.1	10.2	8.2	8.8	7.2	12.8
LC-MS/MS DiaSorin	30	29–33	4.7	2.9	3.8	2.1	2.1	3.8
PQ LC-MS/MS	28	27–29	2.5	3.3	2.9	2.2	1.9	3.8
High pool								
Abbott	92	89–96	2.0	1.8	1.9	1.6	2.8	2.5
IDS	88	77–96	1.7	4.4	3.1	3.9	5.3	6.3
Liaison	78	71–94	8.6	4.2	6.4	10.1	2.3	9.6
RIA	72	64–75	4.6	5.6	5.1	1.5	6.7	5.1
Roche	50	47–54	2.3	—	—	5.5	—	—
Siemens	71	63–82	8.4	6.0	7.2	11.0	5.9	8.9
LC-MS/MS DiaSorin	88	78–95	4.1	4.1	4.1	4.5	3.2	6.1
PQ LC-MS/MS	79	76–81	1.6	1.6	1.6	2.4	1.0	2.0

[a] To convert 25OH-D concentrations to micrograms per liter, multiply by 0.4.
[b] This between-run precision was affected by a 4-μg/L (10-nmol/L) shift on day 5 of the first precision experiment. The second experiment better reflects our experience with this assay.
[c] The poor precision of the Roche does not reflect our experience with this assay and may be attributable to the instrument used for this study [Herrmann et al *(12)*]. At the time of the second precision experiment the Roche assay had been discontinued and a repetition of the precision experiment was not possible.

munoassays are technically capable of reaching this goal. Pooling the results from the 2 precision experiments best reflects long-term precision of the assays in routine practice and confirms that not all immunoassays meet the precision goal of 9.1%. In fact, even the RIA did not meet this goal at the lower 25OH-D concentration. Unfortunately, the poor precision of the Roche test at the higher concentration could not be reevaluated owing to discontinuation of this test between the 2 precision experiments. The poor precision seen with the Roche assay may be specific to the analyzer used in our study, because a previous study had shown excellent precision for this test *(12)*. During the first experiment the LIAISON assay also showed suboptimal precision for the low serum pool but improved markedly during the second experiment. The poor precision of the LIAISON assay during the first experiment was caused by a positive shift of 4 μg/L (10 nmol/L) on day 5.

Similarly to their polyclonal 25OH-D_3 assay *(12)*, Roche's monoclonal assay shows significant analytical deficits. Roche has acknowledged this problem and replaced the assay in 2011. Abbott has also addressed the limitations of their tests and restandardized the Architect assay to correct for an over-recovery of approximately 15% to 25% compared to LC-MS/MS. It remains to be seen if these adjustments rectify these issues. Our results highlight the need for laboratories to evaluate carefully all tests that are used in clinical practice.

It is also worth mentioning that among the entire study cohort, only 4 study participants were found to have appreciable concentrations of 25OH-D_2. In all of these individuals the 25OH-D_2 concentration represented <25% of the total 25OH-D. Therefore, variable cross-reactivity with 25OH-D_2 in the assays is unlikely to be responsible for the differences in 25OH-D detected in certain assays. This result also confirms that 25OH-D_2 is not an analytical issue in Australia, where most 25OH-D supplements contain vitamin D_3. However, vitamin D_2-containing supplements are commonly prescribed in some countries, such as the US.

For laboratories in these countries it may thus be prudent to confirm the performance of their assay on a set of 25OH-D_2–containing samples.

The presence of less biologically active vitamin D epimers, particularly 3-epi-25OH-D_3 and 3-epi-25OH-D_2, is another potential confounder in 25OH-D measurement. Initial reports suggested that this issue is confined to children younger than 1 year *(25)*. However, a recent study by Stepman et al. showed that even in adults 3-epi forms are present in variable concentrations and may represent up to 17% of total 25OH-D *(26)*. A limitation of the LC-MS/MS methods used in this study is that they did not separate the epimer forms. All of the methods used in this study, including LC-MS/MS, detect epimers to a variable degree and this characteristic may have contributed to the scatter and bias observed for some assays. Therefore, the influence of 25OH-D epimer forms on the performance of 25OH-D assays remains unclear, and more detailed studies using LC-MS/MS methods that can separate vitamin D epimers are needed.

Another potential limitation of this study was the use of frozen samples for both the method comparison and precision studies. Because of the wide geographic dispersion of the participating laboratories freezing of samples was necessary to ensure identical preanalytical conditions for all analyses. Samples were frozen and thawed only once, which is acceptable according to the manufacturers' pack inserts. Furthermore, previous studies have shown that multiple freeze and thaw cycles have no significant effect on 25OH-D *(27–28)*. On the other hand, rare effects on the sample matrix including VDBP cannot completely be excluded.

In conclusion, several automated 25OH-D immunoassays have recently been launched. The DiaSorin LIAISON premarket evaluation assay demonstrated the best performance characteristics. LIAISON, IDS, and Siemens met minimum performance goals for the measurement of 25OH-D at concentrations >8 μg/L (>20 nmol/L) and can be recommended for routine use. None of the automated immunoassays can reliably quantify 25OH-D concentrations <8 μg/L (<20 nmol/L). Regardless of the assay employed, it is of considerable importance that clinical laboratories be aware of the limitations of their particular assay. Finally, it is expected that vitamin D analysis will continue to be an evolving field because a number of manufacturers have recognized the limitations of their assay and are in the process of attempting to address these issues.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: C.-J. Lancaster Farrell, Siemens Diagnostics.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: M. Herrmann, Siemens Healthcare.
Research Funding: National Health and Medical Research Project Grant (not related to this report) and DiaSorin; M. Herrmann, Siemens.
Expert Testimony: None declared.
Other Remuneration: C.-J. Lancaster Farrell, Siemens.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: We are indebted to David McDonald (Laverty Pathology, Australia), Joshua Soldo (DiaSorin, US), Stephen Paull (DiaSorin, Australia), Steve Cummings (DiaSorin, US), Kim Gentle (Siemens Healthcare, Australia), and Basil Daher (Abbott Diagnostics, Australia) for their assistance during the completion of this study.

References

1. Holick MF. Vitamin D deficiency. N Engl J Med 2007;357:266–81.
2. Rosen CJ. Clinical practice: vitamin D insufficiency. N Engl J Med 2011;364:248–54.
3. Medical Benefits Rev Task Group. Review of the funding arrangements for pathology services. http://www.health.gov.au/internet/main/publishing.nsf/Content/A7F7F7F944757607CA2576A2001C1725/$File/Pathology%20Review%20Discussion%20Paper.pdf. (Accessed December 2011).
4. DeLuca HF. Overview of general physiologic features and functions of vitamin D. Am J Clin Nutr 2004;80:1689S–96S.
5. Holick MF, Biancuzzo RM, Chen TC, Klein EK, Young A, Bibuld D, et al. Vitamin D2 is as effective as vitamin D3 in maintaining circulating concentrations of 25-hydroxyvitamin D. J Clin Endocrinol Metab 2008;93:677–81.
6. Clemens TL, Zhou XY, Myles M, Endres D, Lindsay R. Serum vitamin D2 and vitamin D3 metabolite concentrations and absorption of vitamin D2 in elderly subjects. J Clin Endocrinol Metab 1986;63:656–60.
7. Leventis P, Kiely PD. The tolerability and biochemical effects of high-dose bolus vitamin D2 and D3 supplementation in patients with vitamin D insufficiency. Scand J Rheumatol 2009;38:149–53.
8. Romagnoli E, Mascia ML, Cipriani C, Fassino V, Mazzei F, D'Erasmo E, et al. Short and long-term variations in serum calciotropic hormones after a single very large dose of ergocalciferol (vitamin D2) or cholecalciferol (vitamin D3) in the elderly. J Clin Endocrinol Metab 2008;93:3015–20.
9. Herrmann M, Williams P. Measurement of vitamin D. In: Herrmann W, Obeid R, eds. Vitamins in the prevention of human disease. 1st ed. Berlin/New York: Walter de Gruyter GmbH & Co KG; 2011. P 395–9.
10. Hart GR, Furniss JL, Laurie D, Durham SK. Measurement of vitamin D status: background, clinical use, and methodologies. Clin Lab 2006;52:335–43.
11. Lensmeyer GL, Wiebe DA, Binkley N, Drezner MK. HPLC method for 25-hydroxyvitamin D measurement: comparison with contemporary assays. Clin Chem 2006;52:1120–6.
12. Herrmann M, Harwood T, Gaston-Parry O, Kouzios D, Wong T, Lih A, et al. A new quantitative LC tandem mass spectrometry assay for serum 25-hydroxy vitamin D. Steroids 2010;75:1106–

12.
13. Glendenning P, Taranto M, Noble JM, Musk AA, Hammond C, Goldswain PR, et al. Current assays overestimate 25-hydroxyvitamin D3 and underestimate 25-hydroxyvitamin D2 compared with HPLC: need for assay-specific decision limits and metabolite-specific assays. Ann Clin Biochem 2006;43:23–30.
14. Christakos S, Ajibade DV, Dhawan P, Fechner AJ, Mady LJ. Vitamin D: metabolism. Endocrinol Metab Clin North Am 2010;39:243–53.
15. Tai SS, Bedner M, Phinney KW. Development of a candidate reference measurement procedure for the determination of 25-hydroxyvitamin D3 and 25-hydroxyvitamin D2 in human serum using isotope-dilution liquid chromatography-tandem mass spectrometry. Anal Chem 2010;82:1942–8.
16. Stepman HC, Vanderroost A, Van Uytfanghe K, Thienpont LM. Candidate reference measurement procedures for serum 25-hydroxyvitamin D3 and 25-hydroxyvitamin D2 by using isotope-dilution liquid chromatography-tandem mass spectrometry. Clin Chem 2011;57:441–8.
17. 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;411:1976–82.
18. Liao JJ, Lewis JW. A note on concordance correlation coefficient. PDA J Pharm Sci Technol 2000;54:23–6.
19. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255–68.
20. Passing H, Bablok. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. J Clin Chem Clin Biochem 1983;21:709–20.
21. Fraser CG. Biological variation: from principles to practice. Washington (DC): AACC Press; 2001.
22. Viljoen A, Singh DK, Farrington K, Twomey PJ. Analytical quality goals for 25-vitamin D based on biological variation. J Clin Lab Anal 2011;25:130–3.
23. Holick MF, Binkley NC, Bischoff-Ferrari H, Gordon CM, Hanley DA, Heaney RP, et al. Evaluation, treatment, and prevention of vitamin D deficiency: An endocrine society clinical practice guideline. J Clin Endocrinol Metab 2011;96:1911–30.
24. Carter GD, Jones JC. Use of a common standard improves the performance of liquid chromatography-tandem mass spectrometry methods for serum 25-hydroxyvitamin-D. Ann Clin Biochem 2009;46:79–81.
25. Singh RJ, Taylor RL, Reddy GS, Grebe SKG. 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 Metab 2006;91:3055–61.
26. Stepman HCM, Vanderrost A, Stöckl D, Thienpont LM. Full-scan mass spectral evidence for 3-epi-25-hydroxyvitamin D3 in serum of infants and adults. Clin Chem Lab Med 2011;49:253–6.
27. Lewis JG, Elder PA. Serum 25-OH vitamin D2 and D3 are stable under exaggerated conditions. Clin Chem 2008;54:1931–2.
28. Antoniucci DM, Black DM, Sellmeyer DE. Serum 25-hydroxyvitamin D is unaffected by multiple freeze-thaw cycles. Clin Chem 2005;51:258–61.