Evaluation of Vitamin D Assays in Relation to the Measurement Range and the Statistical Method

Ölçüm Aralığı ve Kullanılan İstatistiksel Yöntemlere Göre D Vitamini Ölçüm Yöntemlerinin Değerlendirilmesi

Nihal YÜCEL,[a]
Nilgün UZUN KARA,[b]
Özlem ÇAKIR MADENCİ,[a]
Lale DAĞDELEN,[a]
Derya AKBABA,[a]
Asuman ORÇUN KAPTANAĞASI[a]

[a]Department of Clinical Biochemistry, Dr. Lütfi Kırdar Kartal Training and Research Hospital,
[b]Kaanmed Medical Toxicology Laboratories, İstanbul

Geliş Tarihi/*Received:* 22.04.2014
Kabul Tarihi/*Accepted:* 01.09.2014

This study was presented as a poster in XIII. National Clinical Chemistry Congress, 25-28 April 2013, İzmir, Turkey.

Yazışma Adresi/*Correspondence:*
Nihal YÜCEL
Dr. Lütfi Kırdar Kartal Training and Research Hospital,
Department of Clinical Biochemistry,
İstanbul,
TÜRKİYE/TURKEY
yucel.nihal@yahoo.com

ABSTRACT Objective: The aim of this study was to compare three vitamin D measuring assays with a reference method to analyze whether their performance varied according to different statistical procedures and measurement ranges used. **Material and Methods:** Three vitamin D measurement methods, a high-performance liquid chromatography (HPLC) and two immunoassay systems (Architect, Abbott Diagnostics and Cobas, Roche Diagnostics) were compared with liquid chromatography tandem mass spectrometry (LCMS/MS) as the reference method in a study population of 141 individuals. The study group was investigated as whole group and two different subgroups of <50 ng/ml and <20 ng/ml, according to LCMS/MS measurements. The results were analyzed by Passing Bablok regression analysis, concordance correlation coefficient, interrater agreement, Bland Altman plots. **Results:** Data generated from the various statistical analyses revealed that all of the methods investigated in this study performed worse for lower concentrations (<50 ng/ml and <20 ng/ml). The best overall performance was obtained by HPLC method. The methods showed variable performance with the statistical analysis used. **Conclusion:** Method comparison studies of vitamin D are prone to variable results according to the statistical method and measurement range chosen. Method comparison studies should be based on clinical requirements. Laboratory professionals should determine the range that is important from clinical point of view, and should evaluate the results according to total allowable error for that range.

Key Words: Vitamin D; immunoassay; chromatography, high pressure liquid; mass spectrometry

ÖZET Amaç: Çalışmanın amacı üç D vitamini ölçüm yöntemini, referans kabul edilen bir yöntemle karşılaştırmak ve performanslarının kullanılan istatistiksel yöntem ve seçilen ölçüm aralığına göre değişim gösterip göstermediğini değerlendirmektir. **Gereç ve Yöntemler:** Yüz kırk bir bireyden oluşan bir çalışma grubunda, bir yüksek performans likid kromatografi (HPLC) ve iki immunoassay yöntemi (Architect, Abbott Diagnostics ve Cobas, Roche Diagnostics) referans yöntem olarak alınan bir sıvı kromatografi-ardşık kütle spektrometresi (LCMS/MS) yöntemi ile karşılaştırıldı. Sonuçlar tüm çalışma grubu, D vitamini konsantrasyonu <50 ng/ml olan ve <20 ng/ml olan üç ayrı alt grupta incelendi. Elde edilen veriler Passing Bablok regresyon analizi, konkordans uyum katsayısı, interrater uyum, Bland Altman grafikleri ile değerlendirildi. **Bulgular:** Farklı istatistiksel analizlerin sonucuna göre, tüm yöntemler daha düşük konsantrasyonlarda (<50 ng/ml ve <20 ng/ml) daha düşük performans gösterdi. Ayrıca yöntemler, kullanılan istatistiksel analize bağlı olarak değişen performanslar sergilediler. En iyi toplam performans HPLC yöntemi ile elde edildi. **Sonuç:** Yöntem karşılaştırma çalışmalarında elde edilen performansların kullanılan istatistiksel analize ve kullanılan konsantrasyon aralığına bağlı olarak değiştiği tespit edilmiştir. Bu nedenle, bu çalışmalar klinik ihtiyaçlara göre yapılmalıdır; laboratuarlar yöntem performanslarını, klinik olarak önemli olan aralıklarda ve toplam kabul edilebilir hataya göre değerlendirmelidir.

Anahtar Kelimeler: D vitamini; bağışıklık testi; kromatografi, yüksek basınçlı sıvı; kitle spektrometri

doi: 10.5336/medsci.2014-40251

Turkiye Klinikleri J Med Sci 2014;34(4):406-16

The clinical outcome of vitamin D (vit D) deficiency has been one of the most popular subjects of the past decades. The endocrinological effects of vit D on bone metabolism are well known.[1] Moreover, vit D is involved in many other clinical situations, such as immunomodulation, regulation of blood pressure, some metabolic processes, cell growth, and apoptosis.[2] Based on this fact, in recent years we witnessed a large number of studies showing the association of vit D deficiency with various clinical situations; these were overall mortality, solid cancers, cardiovascular diseases, insulin resistance, obesity, diabetes, and autoimmune diseases like multiple sclerosis, dysregulated immune function, asthma, and increased infection rates.[3-12] As a result, the number of vit D requests by clinicians has increased exponentially during the last decade.

Vit D is the name of a group of compounds known as secosteroids, the most important of which are cholecalciferol (vit D_3) and ergocalciferol (vit D_2).[13] Vit D_3 is mainly derived on the skin by the action of UV-B radiation. The source of vit D_2 is the ingestion of vit D-containing plants and fungi. The subsequent steps of vit D metabolism are two successive hydroxylations; the first occurs in the liver producing 25-hydroxyvitamin D_3 (25-OHD_3) and 25-hydroxyvitamin D_2 (25-OHD_2), and the second, in the kidney, producing the biologically active forms 1,25-hydroxyvitamin D_3 (1,25-OHD_3) and 1,25-hydroxyvitamin D_2 (1,25-OHD_2). Although 1,25-OHD is the active form, measurement of 25-OHD is preferred to evaluate the vit D status of the body: 25-OHD has a longer half-life. The quantity of this metabolite circulating in the blood is hundreds of times greater (approximately 95% of the circulating 25-OHD pool). Its levels are not affected by the action of parathyroid hormone.[14,15]

The methodologies used for the measurement of vit D and the cut-off level used to define vit D deficiency are subject to controversies. The hydrophobic and lipophilic nature of vit D leads to matrix effects; its strong binding to vit D binding protein (DBP) requires deproteinization procedures, and the structural similarities of vit D_2 and vit D_3 arise methodological problems.[16,17]

Until recently, there was no reference measurement procedure for the comparison of different vit D measurement methods. The development of the liquid chromatography tandem mass spectrometry (LCMS/MS) method, which is accepted as the reference method by the Joint Committee for Traceability in Laboratory Medicine,[18,19] has been one of the most promising advances in this field. In addition, the US National Institute of Standards and Technology (NIST) has produced two calibrators (SRM 2972) containing 25-OHD_3 and 25-OHD_2 and four serum-based standard reference materials (SRM 972). The establishment of a reference method and the use of certified reference materials have supported standardization and comparability of measurement methods.

Two major methodologies used for measurement of vit D, immunoassays (RIA and automated immunoassays) and physical detection methods [high-performance liquid chromatography (HPLC) and LCMS/MS] have advantages and disadvantages. Although immunoassays are easily automated, high-throughput, inexpensive, and user-friendly methods, they measure total vit D, are prone to matrix effects, use secondary calibrators, and cannot completely separate 25-OHD from DBP. Chromatographic methods are complicated and time-consuming procedures requiring expensive equipment and skilled personnel. They can measure vit D_2 and D_3 separately, are more precise, specific, stable and cost effective, use primary standards, and contain procedures to dissociate 25-OHD from DBP.

The difficulties associated with chromatographic methods and tremendously increased demand of vit D testing have driven laboratories to use easier and rapid immunoassay methods. Several studies on the performance of these methods have been accomplished, and a considerable discrepancy has been documented among the methods.[20-25] Despite the effort of immunoassay manufacturers to improve the performance of their assays, recent studies comparing immunoassays with the LCMS/MS reference procedure have found inconsistent results.[26-30]

The aim of this study was to compare these methods with the reference procedure, to analyze their performance according to the recommended statistical procedures, and to interpret the results from various point of views.

MATERIAL AND METHODS

STUDY DESIGN

An HPLC method and two automated immunoassays (Architect, Abbott Diagnostics and Cobas, Roche Diagnostics) were compared to LCMS/MS method using samples collected from 141 subjects with a wide range of 25-hydroxyvitamin D concentrations. All of the four measurements were performed according to the manufacturer's recommendations by different experienced biochemists. The range of the samples was between 4.44 and 116.0 ng/ml (11 to 290 nmol/L) as measured by the LCMS/MS method. Because Cobas linearity was up to 70 ng/ml, four individuals with higher values were excluded from the statistical analyses of the Cobas method. Although the linearity of the Abbott Architect system was approved up to 96 ng/ml by the FDA, we took into account the linearity limits claimed by the manufacturer. A total of 70 individuals of the total group had undetectable 25-OHD_2 levels, as determined by LCMS/MS. The 25-OHD_2 levels of the remaining 70 patients were less than 1 ng/ml (range: 0.03- 0.54 ng/ml), and only one of these patients had a 25-OHD_2 level greater than 1 ng/ml. Therefore, these levels were considered as negligible, and the patients were not excluded from the study. The values of the entire range were evaluated as a "whole group", and two different subgroups ("<50 ng/ml group" and "<20 ng/ml group") in order to determine how the performances of the methods were affected by the concentration range. These two values were chosen for clinical significance; 20 ng/ml (50 Nmol/L) as the critical point for deficiency status, and 50 ng/ml (125 Nmol/L) as one of the most frequently expected values in the reference range of healthy subjects.

The study was approved by the Ethics Review Committee of our hospital, and the patients signed their informed consents.

ASSAY METHODOLOGIES

Immunoassay (Cobas, Roche, Japan)

The Cobas Roche Vitamin D total assay (Germany) is a competitive electrochemiluminescence immunologic assay. In opposition to the old reagent of the manufacturer which was sensitive to 25-OHD_3, this new reagent launched on May 2011 can measure both 25-OHD_2 and 25-OHD_3. The method is standardized with NIST reference material. The limit of quantification (LoQ) of the method is 5.0 ng/ml. For the analyzer used in the study, the interassay coefficient variations (CV) provided by the manufacturer were less than 6.8%, and the intraassay CVs were less than 13.1% for the concentration range of 8.35 to 69.6 ng/ml. The method is linear in the concentration range of 3.0-70 ng/ml.

Immunoassay (Architect, Abbott, USA)

The Abbott Architect 25-OH Vitamin D assay (USA) is a competitive chemiluminescent microparticle immunoassay (CMIA). The method is standardized with NIST reference material. For a range of concentrations of 19.0-78.4 ng/ml, the LoQ of this method is 8.0 ng/ml, the intraassay CVs are less than 3.7%, and the interassay CVs are less than 4.6%. The method is linear up to 165.5 ng/ml.

HPLC (Shimadzu, Japan)

The HPLC measurement was performed using the Shimadzu HPLC system. The commercial reagents kit provided by Immuchrom (Heppenheim, Germany) was used. The method uses an acetonitrile precipitation and extraction step before the sample is injected into the system. The HPLC separation utilizes an isocratic method at 30°C with a 'reversed phase' (C18) column using a flow rate of 1.0 ml/min, a 50-µl sample volume, and a 15-minute running time. The chromatograms are detected by a UV detector at 264 nm. The results are calculated by the 'internal standard method' through the integration of the peak areas. The method is standardized with NIST reference material. The limit of detection is 2.3 ng/ml. The intraassay CVs claimed by the manufacturer are 2.6% and 1.5% for 22.6 ng/ml and 41.92 ng/ml, respectively, and the

interassay CVs are 4.0% and 3.6% for 21.64 ng/ml and 42.16 ng/ml, respectively. The method is linear up to a concentration of 500 ng/ml.

LCMS/MS (AB SCİEX, USA)

The vitamin D levels in the serum samples were measured using a 3200 Q TRAP model HPLC/MS/MS analyzer obtained from AB SCİEX. The reagent used was MassChrom 25-OH-Vitamin D3/D2, which was obtained from the Serum/Plasma kits provided by Chromsystems (Germany). These reagents have been validated for AB SCİEX HPLC/MS/MS analyzers. A trap column concentrates the analytes and separates the interfering substances. A two-position six-port-valve connects the trap column to an HPLC column, where the chromatographic separation takes place. Atmospheric pressure chemical ionization and a deuterated internal standard are used to ensure precision and robustness, and to minimize the ion suppression effects. The system is calibrated with Chromsystems 3PLUS calibrators (3PLUS1 Multilevel Serum Calibrator Set 25-OH-Vitamin D3/D2, Reference number: 62028). The NIST 972 reference material was used to set the values for the calibrators. The vitamin D concentration is calculated according to the calibration curve and the internal standard area ratios. The LoQ for vitamin D_3 is 3.0 ng/ml. The precision values claimed by the manufacturer are as the following: the intraassay CVs are 2.7% for low concentrations and 4.2% for high concentrations, and the interassay CVs are 3.9% for low concentrations and 4.0% for high concentrations. The method is linear in the concentration range of 2-250 ng/ml.

ANALYTICAL STUDIES

Two levels of serum pools were used for the quality control study. The proper number and volume of the portions were prepared. The aliquots were stored at -80°C, and freshly thawed for each run. Both pools were studied in duplicate at three different times a day, and on five consecutive days by each system. The concentrations of the low and the high pools were respectively 11.45 ng/ml and 32.25 ng/ml, as measured by LCMS-MS.

As HPLC method was the routine method of our laboratory, we made a recovery study at two decision points: 24.9 ng/ml and 54.6 ng/ml. The recovery rates were 105% and 96%, respectively.

We also calculated the bias of our home method, HPLC, using five external quality control results of our laboratory, and the laboratory bias was found as 3.6%.

STATISTICS

The results from two automated immunoassays and HPLC were compared using the LCMS/MS results as the reference. The performances of the groups were analyzed using Passing-Bablok regression, concordance correlation coefficient, inter-rater agreement, Bland-Altman plots and total error calculation. The statistical analyses were performed with Medcalc 12 statistical software.

The Bland-Altman analysis visualizes the difference between the measurements compared with the arithmetic mean of two measurements or compared with the results of the reference method. It is often recommended to plot the % difference (difference/mean x 100) or the ratio of the two measurements, especially if there is a trend towards any of the values. The mean difference and the ±1.96 SD limits, which are the so-called limits of agreement, are displayed.

The inter-rater agreement (Kappa statistics) measures the agreement between two or more observers (tests identifying diagnosis in this case). We found the agreement of the three methods with LCMS/MS, and identified the deficiency status as less than 20 ng/ml according to the Endocrine Society Clinical Practice Guidelines.[31] The interpretation of Kappa is as follows: <0: less than chance; 0.01-0.20: slight; 0.21-0.40: fair; 0.41-0.60: moderate; 0.61-0.80: substantial; and 0.81-0.99: almost perfect agreement.

The concordance correlation coefficient (CCC) is a popular index used for the assessment of agreement. This coefficient gives a measure of the agreement using both precision (Pearson correlation coefficient) and accuracy (bias correction factor). A CCC value greater than 0.99 is interpreted as ex-

cellent, whereas CCC values in the ranges of 0.99 to 0.95, 0.90 to 0.94, and less than 0.90 indicate good, moderate, and poor agreement, respectively.

The Passing-Bablok Regression calculates a regression equation between two methods (y = a + bx), and gives confidence intervals for the constant (a) and the proportional bias (b). For significant agreement, the confidence intervals should be zero for a and one for b.

The total error (TE) was also calculated through the total CV of the methods, and the total bias derived from the regression equations. The CVs of each assay were calculated according to the Clinical and Laboratory Standards Institute (CLSI) EP15 protocol, with two different serum pools of critical levels. Each mean value was placed in the related regression equation, and the corresponding biases were calculated. Finally, the TEs were calculated by the formula TE=1.65xCV+bias. To evaluate the performances of the assays, the total allowable error (TEa) was determined as 25%, with a CV ≤ 10% and a bias ≤ 5% according to the recommendations of Stöckl et al.[32]

The data were categorized into three groups according to the LCMS/MS results: <20 ng/ml (50 nmol/L), <50 ng/ml (125 nmol/L), and total group. All of the statistical analyses were applied to all of the three groups.

RESULTS

The precision performance of the assays is shown in Table 1. The precision of all of the methods were below the allowable CV value of 10%. The best overall precision for both levels was achieved by LCMS/MS (2.51% and 0.88%) and HPLC (1.89% and 1.48%) for the low- and the high-concentration serum pools, respectively.

The total group, which included 141 individuals, consisted of 114 (80.9%) women and 27 (19.1%) men. The mean and SD values obtained by four methods for three groups are presented in Table 2.

The Passing-Bablok analysis revealed significant deviations from linearity with the Architect system in the total group (p=0.03) and with the Cobas system in <50 ng/ml group (p=0.05). There was no significant deviation from linearity for the rest of the groups. The Passing-Bablok regression analysis also revealed variable amounts of constant biases for all of the methods in each group (Table 2, Figure 1). The lowest biases were detected with the HPLC method. The intercept values for the HPLC method were -0.61, -1.24, and -1.72 for the total, <50 ng/ml, and <20 ng/ml groups, respectively. The other two methods exhibited worse constant biases for all three groups, with the exception of the intercept value of -0.28 for the Cobas system obtained for <20 ng/ml group (Table 2). A degree of proportional biases was also found for all of the methods, but those observed in the Architect and Cobas methods for <20 ng/ml group were markedly high (the slopes were 0.60 for Architect and 0.71 for Cobas). The line of best fit crossed the line of identity in the Architect and the HPLC methods for <20 ng/ml group.

The Bland-Altman plots obtained for three groups with three methods compared with LCMS/MS are shown in Figure 2. The % biases were lowest with the Architect method in the total and <50 ng/ml groups: 7.3% and 5.2%, respectively.

TABLE 1: The precision performance of the methods.

	Low Level			High Level		
	Mean (ng/mL) ± SD	Within run CV %	Total CV %	Mean ± SD	Within run CV %	Total CV %
LCMS/MS	11.45 ± 0.29	2.65	2.51	32.25 ± 0.29	0.92	0.88
Architect	12.36 ± 0.60	2.55	5.16	34.57 ± 1.56	2.08	4.80
HPLC	10.14 ± 0.18	1.29	1.89	31.21 ± 0.45	1.10	1.48
Cobas	9.26 ± 0.61	6.49	5.88	23.07 ± 1.66	4.90	7.51

SD: Standard deviation; CV: Coefficient variation (1 ng/mL=2.5 nmol/L).

TABLE 2: The means, standard deviations and Passing Bablok, Concordance Correlation, Interrater agreement analysis results of the methods.

				Passing-Bablok regression analysis				Concordance correlation analysis		Interrater agreement
	n	Mean (ng/mL)	SD	Intercept	95% CI	Slope	95% CI	CCC	95% CI	Kappa
Total group										
LCMS/MS	141	29.02	20.81							
Architect	141	30.69	27.65	1.61	0.00 to 3.03	0.92	0.85 to 1.00	0.89	0.86 to 0.92	0.89
HPLC	141	31.26	22.35	-0.61	-1.75 to 1.10	1.06	1.01 to 1.13	0.94	0.92 to 0.96	0.88
Cobas	137	22.98	16.14	-3.82	-5.66 to -2.07	0.99	0.90 to 1.08	0.90	0.87 to 0.93	0.70
< 50 ng/mL										
LCMS/MS	125	22.93	11.10							
Architect	125	22.30	10.14	2.95	1.97 to 4.13	0.82	0.77 to 0.88	0.87	0.83 to 0.91	
HPLC	125	25.13	13.14	-1.24	-3.43 to 0.82	1.11	1.02 to 1.21	0.85	0.80 to 0.89	
Cobas	125	20.06	12.99	-5.02	-7.22 to -3.03	1.06	0.97 to 1.16	0.82	0.76 to 0.87	
< 20 ng/mL										
LCMS/MS	53	11.99	4.17							
Architect	53	13.19	3.11	5.79	4.40 to 7.22	0.60	0.48 to 0.74	0.64	0.47 to 0.76	
HPLC	53	13.71	5.99	-1.72	-6.54 to 1.32	1.18	0.94 to 1.59	0.61	0.44 to 0.73	
Cobas	53	9.00	5.18	-0.28	-2.86 to 1.66	0.71	0.54 to 0.91	0.47	0.27 to 0.62	

SD: Standard deviation, CI: Confidence interval, LCMS/MS: Liquid chromatography tandem mass spectrometry, HPLC: High-performance liquid chromatography, CCC: Concordance correlation coefficient. 1 ng/ml =1 nmol/L.

In <20 ng/ml group, the best result was obtained with the HPLC method: 18.4%. The highest biases were obtained with the Cobas assay: 14.8%, 15.4%, and 23.1% for total, <50 ng/ml, and <20 ng/ml groups, respectively (Figure 2).

The values of CCC for the total group obtained with the HPLC and Cobas methods were greater than 0.90, and the CCC value for the Architect method was very close (0.89), as observed in Table 2. The CCC values for the other two groups decreased gradually. In <20 ng/ml group, the CCC values reached markedly low levels (0.64 for Architect, 0.61 for HPLC, and 0.47 for Cobas).

The calculated TE values for all three groups are shown in Table 3. The TE results of the HPLC and Architect systems for each group were below the TEa, which was determined to be 25%. In contrast, the TE values obtained for the Cobas system were greater than 25% for all of the groups except <50 ng/ml group at the high concentration range.

The Kappa values used to measure the agreement were 0.89, 0.88, and 0.70 for the Architect, HPLC, and Cobas systems, respectively (Table 3). The p values for all three methods were <0.001 indicating that their measures of agreement between LCMS/MS were significant.

Table 4 displays the overall performance characteristics of three methods. The acceptability criteria were set as follows: The confidence intervals should contain zero for a and one for b for significant agreement in the Passing-Bablok regression. The values of the limits of agreement should be less than 5% in the Bland-Altman plots. Additionally, the CCC value should be greater than 0.90, the TEa should be greater than 25%, and the value of Kappa should be greater than 0.61 to be considered acceptable.

DISCUSSION

The performance of the methods used for the measurement of vitamin D has been studied for several years. Despite huge number of studies, there is not a clear agreement on this subject due to several reasons: the study range and the number of participants, the cross-reactivity of heterophilic antibodies, the 25-OHD_2- and 25-OHD_3-epimer interferences with the methods used in the studies, and the influence of the concentration of vitamin D binding protein.[33,34] In addition to these factors, we hypothesized that the statistical procedures used to

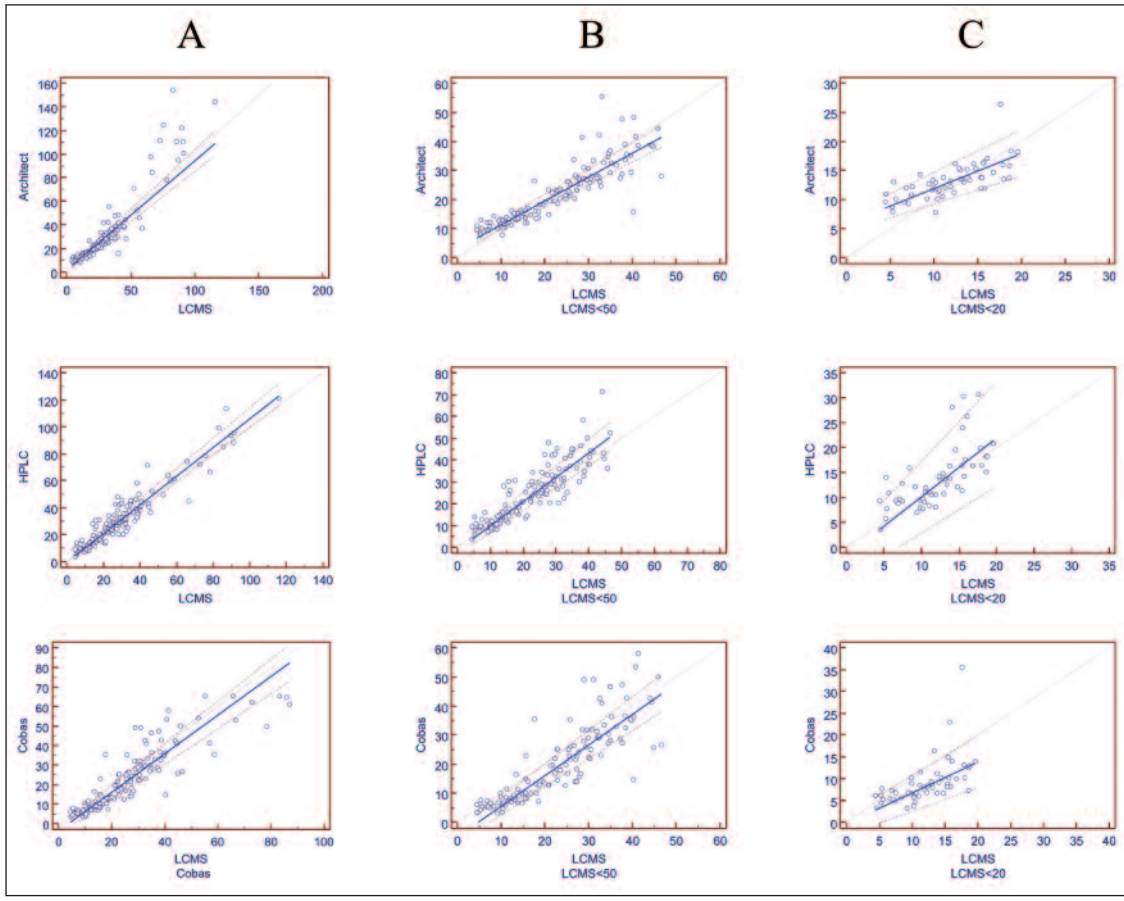

FIGURE 1: Passing Bablok regression analysis graphs of the total (column a), <50 ng/mL (column b), and <20 ng/mL groups (column c).

evaluate the performance of the methods can also contribute to the inconsistency observed between different studies. Therefore, in the present study, we attempted to investigate how the performances of three methods changed according to the study range chosen, and the statistical method used. For these purposes, the results were evaluated in total group and two different subgroups of <50 ng/ml and <20 ng/ml, and different statistical procedures were applied to determine how the performances of the methods varied.

In the present study, all of the statistical analyses were affected by the measurement ranges. As observed in Figure 1, the Passing-Bablok regression analysis revealed that, for the full cohort, almost all of the intercept and slope values of all three methods were acceptable. These findings obtained for the whole group showed variable concordance with other studies and were compatible with recent studies on the performance of the Architect and Cobas systems.[35,36] These regression analysis results were also comparable with the results reported by Ong et al. In another comparison study conducted by Farrel et al., both the Architect and the Cobas systems exhibited dissimilar results.[37,38] Heijboer et al. evaluated the influence of DBP on the accuracy of the assays and found better results for the Cobas system and worse results for the Architect system using healthy individuals.[34] Nevertheless, the regression analysis results of all the methods worsened in <50 ng/ml, and especially in <20 ng/ml groups. This finding indicated that the performances of all of the methods decreased at a lower concentration range.

Similar results were obtained with the CCC analysis. The total group results exhibited moderate agreement with CCC values greater than 90 for the HPLC and the Cobas systems, but poor agreement for the Architect method (CCC= 89). However, for <50 ng/ml group, the results of all of the methods

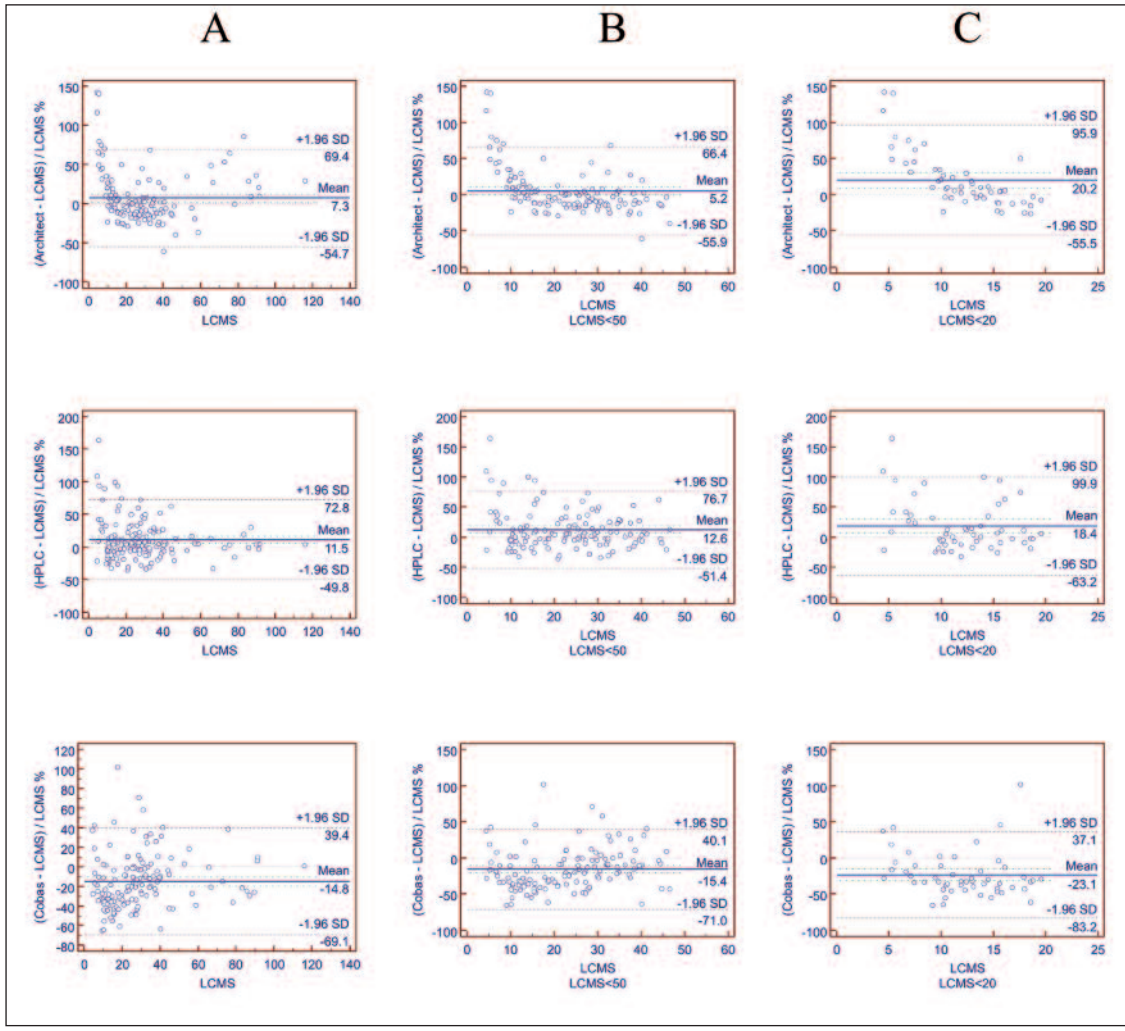

FIGURE 2: Bland-Altman plots of the total (column a), <50 ng/ml (column b), and <20 ng/ml groups (column c).

worsened, and, for <20 ng/ml group, all of the methods were unacceptable.

The total error analysis revealed that the Cobas system results were higher than the TEa, which was determined to be 25% for all three groups (only in <50 ng/ml group at the high concentration range, the TE of the Cobas system was less than 25% at 22.37%). The Architect and HPLC methods exhibited TE values lower than the TEa for all three groups, and worse performances at the lowest concentrations (Table 2).

The Bland-Altman plots for all three methods revealed that the mean % bias obtained with 20 ng/ml group was markedly higher than other groups (Figure 2). The bias was greater for values <10 ng/ml. Farrel et al. also found that most immunoassays exhibited poorer performance at lower concentrations.[38] These researchers chose <8 ng/ml as their lowest concentration, which is lower than the concentration of 20 ng/ml chosen in our study. Farrel et al. observed extremely high % mean differences for all immunoassays compared with LCMS/MS (104.5% for Architect and 35.2% for Cobas). In our study, although the Bland-Altman plots revealed worse % differences for <20 ng/ml group compared to the other groups, the magnitudes of these differences were not as high (20.2% for Architect and -23.1% for Cobas). Farrel et al. concluded that, according to the manufacturer's claims, results that were less than 8 ng/ml are not reported in the Architect system, and that this poor performance does not contribute to the clinical judgement.

TABLE 3: Total error results for all three groups.

Concentration range	n	Regression equation	Low concentration: 11.45 ng/mL			High concentration: 32.25 ng/mL		
			Bias %	CV %	TE %	Bias %	CV %	TE %
Total								
Architect	141	y = 0.9235x +1.6126	6.43	5.16	14.94	2.65	4.80	10.57
HPLC	141	y = 1.0634x - 0.6064	1.04	1.89	4.16	4.46	1.48	6.90
Cobas	136	y = 0,9921x -3,8185	34.14	5.88	43.84	12.63	7.51	25.02
< 50 ng/mL								
Architect	125	y = 0.8250x + 2.9550	8.31	5.16	16.82	8.34	4.80	16.26
HPLC	125	y = 1.1114x - 1.2368	0.34	1.89	3.46	7.30	1.48	9.74
Cobas	125	y = 1.0559x - 5.0206	38.25	5.88	47.95	9.98	7.51	22.37
< 20 ng/mL								
Architect	53	y = 0.6040x + 5.7864	10.93	5.16	19.44			
HPLC	53	y = 1.1817x -1.7188	3.16	1.89	6.28			
Cobas	53	y = 0.7055x - 0.2844	31.93	5.88	41.63			

HPLC: High-performance liquid chromatography, TE: Total error, CV: Coefficient variation. 1 ng/ml =1 nmol/L.

All of the data generated from the various statistical analyses revealed that all three methods investigated in this study performed worse for lower concentrations. This finding indicates that a method that performed well in the whole-concentration range can have poor analytic performance for lower concentration range, e.g., a range that includes the clinical decision limit. In our study, the worst performances were obtained for <20 ng/ml group. This is the cut-off value for insufficiency, and "any method's performance under this level is of less importance" may be true in case of vit D. However, from a technical point of view, this has to be considered as a part of the total performance of the method. Consequently, the outcomes of the method performance studies are concentration-dependent, and this fact should be taken into account.

Another result derived from this study was that each method showed variable performance according to the statistical analysis applied. The performance data for all of the methods in every statistical procedure are summarized in Table 4. The HPLC method, which met all of the performance goals for the total group and most of the goals for the other two groups, exhibited the best consistency with LCMS/MS (intercept, slope, CCC, TEa, and Bland-Altman). However, it failed to achieve the CCC goals in <50 ng/ml and <20 ng/ml groups.

The Architect system met most of the performance goals for the total group (it only failed in the CCC analysis), but could not meet the same goals in the other two groups (intercept, slope, CCC, and Bland-Altman).

The Cobas system exhibited variable performances in three groups. In the analysis of the total group, the Cobas method had poor performance in the TE and the intercept results. In <50 ng/ml group, it was poor in the CCC analysis, and, in <20 ng/ml group, it was poor in all of statistical methods, with the exception of the intercept value, which exhibited considerably poor performance in the other two groups.

In contrast, the inter-rater agreement analysis showed that all three methods performed very well; the HPLC and Architect systems exhibited almost perfect agreement (Kappa values were 0.88 and 0.89, respectively), whereas the Cobas method exhibited substantial agreement (Kappa value was 0.70). This result is consistent with the findings reported by Ajuria-Morentin et al., who had chosen 30 ng/ml as the decision limit.[30] This finding indicates that a method that could not meet some performance goals could perfectly differentiate patients from healthy individuals.

TABLE 4: Overall performance evaluation of the methods.

	Intercept	Slope	CCC	Total error		Bland- Altman	Interrater agreement
				Low	High		
Total group							
Architect	A/N	A	N	A	A	A	A
HPLC	A	A	A	A	A	A	A
Cobas	N	A	A	N	N	A	A
< 50 ng/mL							
Architect	N	N	N	A	A	N	
HPLC	A	A	N	A	A	A	
Cobas	N	A	N	N	A	A	
< 20 ng/mL							
Architect	N	N	N	A		N	
HPLC	A	A	N	A		A	
Cobas	A	N	N	N		N	

HPLC: High-performance liquid chromatography, A: acceptable, N: not acceptable, CCC: Concordance correlation analysis.

Consequently, the data derived from this study suggest that, when a method is compared with a reference method, a good correlation obtained through Passing-Bablok analysis or a low bias in the Bland-Altman plots does not necessarily ensure an acceptable total error. In addition, a method can differentiate disease from non-disease states, but may have a TE > 25%, as in the case of the Cobas system. Thus, a result that is correctly classified as diseased or healthy can still be at least 25% different from the true value. Therefore, one should keep in mind these statistical variabilities when evaluating a method.

Because the amounts of 25-(OH)D_2 measured by LCMS/MS were negligible, we suppose that the assay variabilities in the detection of 25-(OH)D_2 hardly contributed to the inconsistency of the methods with LCMS/MS. In contrast, the LCMS/MS system is sensitive to the 3-epi-25(OH)D_3. This epimer could not be analyzed, and its influence on the results could not be investigated in the study. Another limitation was the number of samples. The results could be more significant if the subgroups containing more samples.

CONCLUSION

Method comparison statistics are mathematical procedures applied to the results obtained by different methods. Each of the statistics reveals a different performance characteristic of a method. From the technical point of view, an assay's performance should meet the acceptability criteria for the total analytical range. However, this sometimes may not be achieved, potentially due to the nature of the analyte itself. In such cases, laboratory professionals should determine the range that is important from the clinical point of view, and evaluate the results according to the total allowable error for that range. Briefly, method comparison studies should be based on clinical requirements.

REFERENCES

1. Cranney A, Horsley T, O'Donnell S, Weiler H, Puil L, Ooi D, et al. Effectiveness and safety of vitamin D in relation to bone health. Evid Rep Technol Assess (Full Rep) 2007;158:1-235.
2. Bouvard B, Annweiler C, Sallé A, Beauchet O, Chappard D, Audran M, et al. Extraskeletal effects of vitamin D: facts, uncertainties, and controversies. Joint Bone Spine 2011;78(1): 10-6.
3. Bischoff-Ferrari HA, Giovannucci E, Willett WC, Dietrich T, Dawson-Hughes B. Estimation of optimal serum concentrations of 25-hydroxyvitamin D for multiple health outcomes. Am J Clin Nutr 2006;84(1):18-28.
4. Holick MF. Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. Am J Clin Nutr 2004;80(6 Suppl):1678S-88S.

5. Melamed ML, Michos ED, Post W, Astor B. 25-hydroxyvitamin D levels and the risk of mortality in the general population. Arch Intern Med 2008;168(15):1629-37.
6. Ng K, Wolpin BM, Meyerhardt JA, Wu K, Chan AT, Hollis BW, et al. Prospective study of predictors of vitamin D status and survival in patients with colorectal cancer. Br J Cancer 2009;101(6):916-23.
7. Giovannucci E, Liu Y, Rimm EB, Hollis BW, Fuchs CS, Stampfer MJ, et al. Prospective study of predictors of vitamin D status and cancer incidence and mortality in men. J Natl Cancer Inst 2006;98(7):451-9.
8. Lappe JM, Travers-Gustafson D, Davies KM, Recker RR, Heaney RP. Vitamin D and calcium supplementation reduces cancer risk: results of a randomized trial. Am J Clin Nutr 2007;85(6):1586-91.
9. Chiu KC, Chu A, Go VL, Saad MF. Hypovitaminosis D is associated with insulin resistance and beta cell dysfunction. Am J Clin Nutr 2004;79(5):820-5.
10. Pittas AG, Lau J, Hu FB, Dawson-Hughes B. The role of vitamin D and calcium in type 2 diabetes. A systematic review and meta-analysis. J Clin Endocrinol Metab 2007; 92(6):2017-29.
11. Gatenby P, Lucas R, Swaminathan A. Vitamin D deficiency and risk for rheumatic diseases: an update. Curr Opin Rheumatol 2013;25(2): 184-91.
12. Pierrot-Deseilligny C, Souberbielle JC. Contribution of vitamin D insufficiency to the pathogenesis of multiple sclerosis. Ther Adv Neurol Disord 2013;6(2):81-116.
13. Dusso AS, Brown AJ, Slatopolsky E. Vitamin D. Am J Physiol Renal Physiol 2005;289(1): F8-28.
14. Holick MF. Vitamin D status: measurement, interpretation, and clinical application. Ann Epidemiol 2009;19(2):73-8.
15. Holick MF. The cutaneous photosynthesis of previtamin D3: a unique photoendocrine system. J Invest Dermatol 1981;77(1):51-8.
16. Carter GD. Accuracy of 25-hydroxyvitamin D assays: confronting the issues. Curr Drug Targets 2011;12(1):19-28.
17. Carter GD. 25-hydroxyvitamin D: a difficult analyte. Clin Chem 2012;58(3):486-8.
18. 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(5):1942-8.
19. 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(3): 441-8.
20. Binkley N, Krueger D, Cowgill CS, Plum L, Lake E, Hansen KE, et al. Assay variation confounds the diagnosis of hypovitaminosis D: a call for standardization. J Clin Endocrinol Metab 2004;89(7):3152-7.
21. Binkley N, Krueger D, Gemar D, Drezner MK. Correlation among 25-hydroxy-vitamin D assays. J Clin Endocrinol Metab 2008;93(5): 1804-8.
22. 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(11):2195-7.
23. Roth HJ, Schmidt-Gayk H, Weber H, Niederau C. Accuracy and clinical implications of seven 25-hydroxyvitamin D methods compared with liquid chromatography-tandem mass spectrometry as a reference. Ann Clin Biochem 2008;45(Pt 2):153-9.
24. Wagner D, Hanwell HE, Vieth R. An evaluation of automated methods for measurement of serum 25-hydroxyvitamin D. Clin Biochem 2009;42(15):1549-56.
25. 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(7):477-88.
26. van den Ouweland JM, Beijers AM, Demacker PN, van Daal H. Measurement of 25-OH-vitamin D in human serum using liquid chromatography tandem-mass spectrometry with comparison to radioimmunoassay and automated immunoassay. J Chromatogr B Analyt Technol Biomed Life Sci 2010;878(15-16):1163-8.
27. 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(23-24):1976-82.
28. Janssen MJ, Wielders JP, Bekker CC, Boesten LS, Buijs MM, Heijboer AC, et al. Multicenter comparison study of current methods to measure 25-hydroxyvitamin D in serum. Steroids 2012;77(13):1366-72.
29. Depreter B, Heijboer AC, Langlois MR. Accuracy of three automated 25-hydroxyvitamin D assays in hemodialysis patients. Clin Chim Acta 2013;415:255-60.
30. Ajuria-Morentin I, Mar-Medina C, Bereciartua-Urbieta E, Izquierdo-Quirce F, Valladares-Gómez C, Crespo-Picot E, et al. Lack of transferability between different immunoassays and LC-MS/MS for total 25-hydroxyvitamin D measurement and disagreement defining deficiency. Scand J Clin Lab Invest 2013;73(1):82-6.
31. Holick MF, Binkley NC, Bischoff-Ferrari HA, Gordon CM, Hanley DA, Heaney RP, et al.; Endocrine Society. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2011;96(7):1911-30.
32. Stöckl D, Sluss PM, Thienpont LM. Specifications for trueness and precision of a reference measurement system for serum/plasma 25-hydroxyvitamin D analysis. Clin Chim Acta 2009;408(1-2):8-13.
33. Herrmann M. The measurement of 25-hydroxy vitamin D - an analytical challenge. Clin Chem Lab Med 2012;50(11):1873-5.
34. 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(3):543-8.
35. Cavalier E, Carlisi A, Bekaert AC, Rousselle O, Chapelle JP, Souberbielle JC. Analytical evaluation of the new Abbott Architect 25-OH vitamin D assay. Clin Biochem 2012;45(6): 505-8.
36. Emmen JM, Wielders JP, Boer AK, van den Ouweland JM, Vader HL. The new Roche Vitamin D Total assay: fit for its purpose? Clin Chem Lab Med 2012;50(11):1969-72.
37. Ong L, Saw S, Sahabdeen NB, Tey KT, Ho CS, Sethi SK. Current 25-hydroxyvitamin D assays: do they pass the test? Clin Chim Acta 2012;413(13-14):1127-34.
38. Farrell CJ, 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(3):531-42.