Clin Chem Lab Med 2019; 57(10): 1546–1555 DE GRUYTER

Etienne Cavalier*, Richard Eastell, Niklas Rye Jørgensen, Konstantinos Makris, Symeon Tournis, Samuel Vasikaran, John A. Kanis, Cyrus Cooper, Hans Pottel and Howard A. Morris, on behalf of the IFCC-IOF Joint Committee for Bone Metabolism (C-BM)

A multicenter study to evaluate harmonization of assays for N-terminal propeptide of type I procollagen (PINP): a report from the IFCC-IOF Joint Committee for Bone Metabolism

https://doi.org/10.1515/cclm-2019-0174
Received February 13, 2019; accepted April 8, 2019; previously published online May 14, 2019

Abstract

Background: Biochemical bone turnover markers (BTM) are useful tools to assess bone remodeling at the cellular level. N-terminal propeptide of type I procollagen (PINP) has been recommended as a reference marker for bone formation in research studies.
Methods: We describe the results of a multicenter study for routine clinical laboratory assays for PINP in serum and plasma. Four centers (Athens, Greece [GR], Copenhagen, Denmark [DK], Liege, Belgium [BE] and Sheffield, United Kingdom [UK]) collected serum and plasma (EDTA) samples from 796 patients presenting to osteoporosis clinics. Specimens were analyzed in duplicate with each of the available routine clinical laboratory methods according to the manufacturers' instructions. Passing-Bablok regressions, Bland-Altman plots, V-shape evaluation method and the concordance correlation coefficient for PINP values between serum and plasma specimens and between methods were used to determine the agreement between results. A generalized linear model was employed to identify possible variables that affected the relationship between the methods.
Results: We showed that both EDTA plasma and serum were suitable for PINP determination. We observed a significant proportional bias between Orion radioimmunoassay and the automated methods for PINP (Roche Cobas and IDS iSYS), which both gave very similar results. The multivariate model did not improve the excellent correlation that was observed between the methods.
Conclusions: Harmonization of PINP assays is possible by applying a correction factor or correctly assigning the values of the calibrators. This work will benefit from further collaboration between assays manufacturers and clinical laboratory professionals.

Keywords: bone marker; bone turnover; bone turnover markers; harmonization; N-terminal propeptide of type I procollagene; propeptide of type I procollagen (PINP).

***Corresponding author: Etienne Cavalier,** Department of Clinical Chemistry, University of Liège, CHU Sart-Tilman, Domaine du Sart-Tilman, B-4000 Liège, Belgium, E-mail: etienne.cavalier@chu.ulg.ac.be
Richard Eastell: Mellanby Centre for Bone Research, University of Sheffield, Sheffield, UK
Niklas Rye Jørgensen: Department of Clinical Biochemistry, Rigshospitalet, Glostrup, Denmark; and OPEN, Odense Patient Data Explorative Network, Odense University Hospital/Institute of Clinical Research, University of Southern Denmark, Odense, Denmark
Konstantinos Makris: Clinical Biochemistry Department, KAT General Hospital, Athens, Greece; and Laboratory for Research of the Musculoskeletal System "Th. Garofalidis", Medical School, University of Athens, Athens, Greece
Symeon Tournis: Laboratory for Research of the Musculoskeletal System "Th. Garofalidis", Medical School, University of Athens, Athens, Greece
Samuel Vasikaran: PathWest Laboratory Medicine, Fiona Stanley Hospital, Murdoch, WA, Australia
John A. Kanis: Centre for Metabolic Bone Diseases, University of Sheffield Medical School, Sheffield, UK
Cyrus Cooper: The MRC Epidemiology Resource Centre, Southampton General Hospital, University of Southampton, Southampton, UK
Hans Pottel: Department of Public Health and Primary Care, KU Leuven Campus Kulak Kortrijk, Kortrijk, Belgium
Howard A. Morris: School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, SA, Australia

Introduction

Osteoporosis is a major disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to an increased risk of fracture, morbidity and mortality. Its prevalence is increasing with aging of the population and the global burden of this disease and its complications in the European Union has been estimated at €37 billion per year [1]. Clinical and pharmacological management of osteoporosis consists in the prevention or reduction of fracture risk and monitoring the response to therapy. For that purpose, measurement of bone turnover biochemical markers could be of great interest. A consensus paper published in 2010 by the International Osteoporosis Foundation (IOF) and the International Federation of

Clinical Chemistry and Laboratory Medicine (IFCC) Joint Working Group on Bone Marker Assays (WG-BMA) recommended one bone formation marker, namely the procollagen type I N-propeptide (PINP) and one bone resorption marker, the C-terminal telopeptide of type I collagen (CTX) be used as reference markers for clinical research studies [2]. A few years later, the IOF and European Calcified Tissue Society (ECTS) proposed that PINP and CTX be measured at baseline and after 3 and 12 months of treatment with oral bisphosphonates to assess adherence to therapy [3].

PINP is a trimer constituted by three non-covalently linked type I collagen subunit chains (two pro-α1 and one pro-α2 chains) presenting a molecular mass of 35 kDa [4]. Type I collagen is synthesized by osteoblasts and deposited in the resorption cavity during bone formation phase. The aminoterminal propeptide is cleaved enzymatically during the last phase of maturation and is then released in the circulation where it is rapidly bound and internalized by the endothelial cells of the liver. In human serum, PINP is present in two major forms, a trimeric and a monomeric form, this latter being elevated in patients suffering from chronic renal failure. The serum concentration of PINP shows little diurnal or seasonal variation and is little influenced by food intake, which is an important advantage over CTX (an extensive review on factors influencing bone turnover markers can be found in [5]).

Currently there are two commercially available automated immunoassays for the detection of PINP in serum and/or plasma, from Immunodiagnostic Systems plc on the iSYS automated analyzer (IDS, Boldon, UK) and from Roche Diagnostics (Mannheim, Germany) on the cobas e family instruments and one manual radioimmunoassay (RIA) from Orion Diagnostica (UniQ PINP RIA, Orion Diagnostica, Espoo, Finland). However, these kits are not equivalent, as they do not recognize to the same extent the monomeric form of the peptide: the "Total" PINP assay (Roche Elecsys) recognizes both the trimeric and the monomeric forms whereas the "Intact" PINP assays (IDS iSYS and Orion Diagnostica) recognize the trimeric form only. As the monomers accumulate in patients suffering from chronic kidney diseases (particularly when the glomerular filtration rate is lower than 30 mL/min/1.73 m^2), they will interfere with the "total" PINP assay from Roche, leading to spuriously elevated PINP results [6]. This is also the case with bedridden geriatric patients [7] and with patients suffering from breast cancer [8].

Previous studies have shown that there is some disagreement between the results generated from patient samples by the three PINP assays [8–13].

Hence, in order to establish the clinical value of PINP as a reference bone biomarker, harmonization of the results from different assays for this biomarker is necessary in people with normal renal function. In this paper, we now report the results of a study conducted by the Joint Committee IFCC-IOF for Bone Metabolism (C-BM) to compare the PINP results generated by each of the available routine clinical assays of patients presenting to osteoporosis clinics.

Materials and methods

Patients and samples

Four centers located in Athens (Greece, GR), Copenhagen (Denmark, DK), Liege (Belgium, BE) and Sheffield (United Kingdom, UK), experienced with performance of these assays, took part in the study. After a familiarization experiment, and according to the agreed protocol, each center recruited 200 patients attending a local osteoporosis clinic. Approval by a local Research Ethics Committee was obtained for each center with all participating patients signing an informed consent form.

Patient blood samples were collected in K_3-EDTA tubes for analyses on plasma and in tubes with gel for analyses on serum. Separation of plasma and serum was achieved in the hour following the sampling and aliquots were stored at −80 °C until PINP determination (maximum 2 years). Demographic data, including sex, age fasting status, current medications and areal bone mineral density (BMD) were also collected. BMD was measured at the lumbar spine (L1–L4) and at the total hip by dual-energy X-ray absorptiometry (DXA) with a Hologic instrument in three centers and with a Lunar machine in one center. The transformation of BMD into standardized BMD (sBMD) was performed according to the equations published by Genant et al. [14].

During the familiarization experiment, each center was asked to run two serum pools in five replicates over 5 consecutive days, with each method, without recalibration according to CLSI EP15-A3 [15]. This performance study was run in parallel with the measurements of the patients' samples. The two serum pools were constituted in BE using remnant human samples that had been stored at −80 °C for >5 years. These samples were mixed together according to their original value to target a final PINP concentration of ≈30 and ≈150 µg/L. After thorough homogenization during 1 h on a rotating plate, pools were centrifuged and aliquoted. The aliquots were stored at −80 °C until shipment on dry ice to each participating center. Each center thus received five aliquots (one for each day) for each instrument. Up to three to five cycles of freeze-thaw do not have a significant effect on measured PINP [12]. Evaluation of the imprecision was performed according to the biological variability concept and the goals were classically calculated as $z \times$ (coefficient of variation (CV) w where $z = 0.50$ and 0.75 for desirable and minimal imprecision and CVw is the within-individual coefficient of variation. We reviewed the literature available on PINP biological variability (Table 1) and found that the median CVw for PINP was 7.3%. Hence, the desirable and minimal CV should be 3.7% and 5.5%.

Table 1: Biological variation of PINP.

First author and reference	Population	Assay	Scheme	CVw	CVg	Desirable CV, %	Desirable bias, %
Healthy subjects							
Alvarez, 2000 [16]	12 healthy premenopausal women	Orion RIA	From two to five samples (median: 4) per person over a period of 1 year	6.2%	18.4%	3.1%	4.9%
Hannon 1998 [17]	11 postmenopausal women	Orion RIA	Four samples: at months 0, 1, 13 and 25	7.4	NA	3.7%	NA
Garnero, 2008 [13]	15 untreated postmenopausal women	Roche Elecsys	Three samples: baseline, day 30 and day 90	7.2% p25–p75 5.3–13.9	NA	3.6%	NA
Clowes 2002 [18]	10 fasting premenopausal women	Roche Elecsys	Five samples, 1 every 2 days, 10 days	10.6%	NA	5.4%	NA
Rogers 2009 [19]	51 postmenopausal women	Orion RIA	Two samples, at two time-points, 2 weeks apart	9.1%	NA	4.6%	NA
Cavalier (not published)	22 healthy individuals (11 males and 11 females)	IDS iSYS	Five samples, one sample per week	5.9%	30.8%	3.0%	7.8%
MEDIAN				7.3%	24.6%	3.7%	6.4%

Analytical methods

The IVD companies producing PINP assay kits (IDS, Roche and Orion Diagnostica) provided reagents and calibrators to the participating laboratories. All reagents were from the same lot and a single calibration was used (except for the RIA). For Roche, the cobas e411 analyzer was the instrument used in all centers. All the laboratories had previous experience in running these methods and all assays were run in all laboratories, except Orion Diagnostica RIA which was only run in BE and DK.

Roche and IDS claim that PINP can either be measured in serum and plasma whereas Orion Diagnostica only claims the use of serum. Plasma and serum samples were run in duplicate on all methods according to the manufacturers' instructions (and with the agreement of Orion to also use EDTA plasma for their method) and results were calculated if the standard curves and the manufacturers' supplied internal quality control (QC) specimens were within the specifications. All measurements took place between December 2016 and March 2017.

Statistical methods

MedCalc (Mariakerke, BE) was used to calculate the Passing-Bablok regressions between serum and plasma and between methods as well as Bland-Altman plots. The coefficients of variation were calculated on duplicates to determine the repeatability of each assay. The mean of the duplicates was used to compare the results. The Mann-Whitney test was used to compare the medians.

The familiarization panel was analyzed per level using an ANOVA model accounting for the effects of center, day and the interaction between day and center and the QC results between centers were compared with ANOVA, followed by Tukey's multiple comparison test with the 95% confidence intervals (CIs) method (Graphpad Prism 6).

If the agreement between results would indicate considerable deviation from the ideal situation (slope $\neq$ '1', intercept $\neq$ '0', $R^2 \neq 1$), we had decided to use the GLMSELECT procedure, using the backward selection option in SAS 9.4, to establish a generalized linear model (GLM) for each comparison with the aim to identify variables which affected the differences between methods and ultimately to obtain a more acceptable prediction model. GLMSELECT provides t-values for the coefficients (instead of p-values). An absolute t-value >1.96 corresponds with a p-value <0.05. The larger the absolute t-value, the more important the coefficient in the GLM.

As small deviations from the ideal situation will be statistically significant because of the (very) large sample size of this study, we reviewed the literature on PINP biological variation (Table 1) which allowed us to define specifications corresponding to a desirable bias of ±9.3% for the slope and ±5 µg/L for the intercept corresponding to the limit of quantification of the assays to build V-shape limits (defined as $y = -0.093\,x - 5$ and $y = 0.093\,x + 5$) for the regression of differences on averages. We calculated the percentages of samples comprised between the V-shape limits and considered that methods were equivalent if 90% of the samples were comprised within the limits.

Finally, we calculated the concordance correlation coefficients (CCC) factor, which evaluates the degree to which pairs of observations fall on the 45° line through the origin according to Lin et al. [20] and the strength of agreement according to McBride et al. [21] as well as the Pearson correlation coefficient (ρ), which measures how far each observation deviates from the best-fit line, and is thus a measure of precision, and a bias correction factor (C_b) that measures how far the best-fit line deviates from the 45° line through the origin, and is thus a measure of accuracy.

Results

All the calibration curves were accepted by the instruments and all the QCs were within the specifications provided by the manufacturers.

Patients

Seven hundred and ninety-six patients (692 women, 104 men) were included in the study. The mean age (±SD) was 66.1 (±11.7) years and mean BMI was 25.9 (±4.8) kg/m^2. There were mean age differences for patients recruited at the various centers: GR had the youngest patient group (mean age = 61.6 ± 8.8) and DK had the oldest patient group (mean age = 70.1 ± 11.3). All patients were in a fasting status in BE and GR whereas this was not necessarily the case in DK and UK. Regarding their treatment, 65.8% were taking calcium, 60.8% vitamin D, 11.1% active vitamin D, 25.9% bisphosphonates, 0.3% strontium ranelate, 9.0% denosumab, 2.0% teriparatide and 1.1% was treated by selective estrogen receptor modulators. The median sBMD was 814 mg/cm^2 (interquartile range [IQR]: 798–827 mg/cm^2) at the spine and 753 mg/cm^2 (IQR: 743–764 mg/cm^2) at the hip. None of the patients presented an eGFR lower than 30 mL/min/1.73 m^2.

Imprecision of the methods and performance evaluation study

The mean of pools were different according to the methods: 30.5 ± 1.6 µg/L and 156.0 ± 8.0 µg/L for Elecsys on level 1 and 2, respectively, vs. 30.8 ± 2.4 µg/L and 174.0 ± 12.0 µg/L for iSYS and 25.7 ± 2.2 and 132.0 ± 24.0 µg/L for Orion RIA. The imprecision of the methods according to the CLSI EP15-A3 guideline, is presented in Table 2. As expected, the manual RIA method presented higher CVs than the automated ones, and this was especially true for pool 2. The ANOVA results showed that for iSYS, the center was the major source of variability whereas for the two other methods, both center and day significantly influenced the results.

For iSYS, the CV ranged between desirable (3.7%) and minimal (5.5%) CV for all centers, except in GR which presented CVs higher than the minimal CV. There are no clear explanations why GR presented higher CVs than the other three centers: all the QC were in the manufacturer's range and the operating conditions were totally controlled. A lack of homogenization of the frozen samples might be the explanation. For Elecsys, the CV ranged between the desirable and minimal (6.2%) CV for all centers and for the RIA, all CVs were higher than the minimal CV (Table 2).

Comparison plasma vs. serum

All Passing-Bablok regressions for method comparisons of the same assay in serum and plasma are presented in Table 3. All the slopes were comprised between 0.94 and 1.05. The V-shape model (Figure 1) shows that more than 95% of the observations fitted within the limits with very little center disparities showing that the test gave equivalent results on both serum and plasma.

Comparison of methods in plasma and serum

Roche Elecsys vs. IDS iSYS

In plasma, the Passing-Bablok regression on the relationship between Elecsys and iSYS was: Roche Elecsys = 0.86 × IDS iSYS + 2.9 and in serum, it was: Roche Elecsys = 0.91 × IDS iSYS + 2.6 (Table 4). The Bland-Altman plots and the V-shaped models (Figure 2A and B) show that, overall, 87.4% and 86.1% of the values were within the limits for serum and plasma, respectively (chi-square [χ^2] = NS). There was some disparity between the centers, with a percentage of agreement ranging from 81.8% in BE to 93.2% in DK for serum and from 78.0 in the UK to 92.2% in GR for plasma.

Roche Elecsys vs. Orion RIA

In plasma, the Passing-Bablok regression on the relationship between Elecsys and Orion RIA was: Elecsys = 1.24 × Orion RIA-0.2 and in serum, it was: Elecsys = 1.22 × Orion

Table 2: Imprecision (CV%) of Roche Elecsys, IDS iSYS and Orion Diagnostica RIA of PINP assays according to the CLSI EP15-A3 guideline.

	Roche Elecsys		IDS iSYS		Orion RIA	
	Pool 1	Pool 2	Pool 1	Pool 2	Pool 1	Pool 2
Mean ± SD, µg/L	30.5 ± 1.6	156.0 ± 8.0	30.8 ± 2.4	174.0 ± 12.0	25.7 ± 2.2	132.0 ± 24.0
ALL	*5.1*	*5.2*	***7.7***	***6.7***	***8.6***	***18.1***
BE	3.1	3.4	3.2	3.0	***7.1***	***20.5***
DK	*4.5*	*4.6*	*3.9*	3.7	***9.9***	***11.3***
GR	*4.5*	3.1	***8.4***	***7.6***	NP	NP
UK	*5.0*	*5.2*	3.2	2.3	NP	NP

The values in italics are those higher than the desirable CV (3.7%) and those in bold higher than the minimal CV (5.5%), based on intra-individual variation of the biomarker. NP, not performed.

Table 3: Passing-Bablok correlation between PINP run in plasma (y) and serum (x).

	Roche Elecsys	IDS iSYS	Orion Diagnostica
ALL			
Slope (95% CI)	0.98 (0.97; 0.98)	1.03 (1.02; 1.04)	0.98 (0.96; 1.00)
Intercept (95% CI)	−0.1 (−0.4; 0.1)	−0.4 (−0.7; −0.1)	−0.1 (−0.5; 0.3)
n	796	794	383
BE			
Slope (95% CI)	1.01 (0.99; 1.03)	1.05 (1.04; 1.07)	1.00 (0.98; 1.03)
Intercept (95% CI)	−0.3 (−0.9; 0.4)	2.7 (2.0; 3.5)	−0.4 (−1.3; 0.3)
n	200	198	200
DK			
Slope (95% CI)	0.95 (0.94; 0.97)	1.01 (0.99; 1.02)	0.94 (0.92; 0.96)
Intercept (95% CI)	−0.0 (−0.3; 0.4)	−0.30 (−0.56; −0.01)	0.4 (−0.1; 0.9)
n	192	192	183
GR			
Slope (95% CI)	0.97 (0.96; 0.99)	0.99 (0.97; 1.01)	NP
Intercept (95% CI)	−0.1 (−0.9; 0.6)	−0.1 (−1.2; 0.8)	
n	204	204	
UK			
Slope (95% CI)	0.96 (0.95; 0.97)	1.05 (1.03; 1.06)	NP
Intercept (95% CI)	0.2 (−0.3; 0.7)	−0.2 (−0.9; 0.5)	
n	200	200	

NP, not performed.

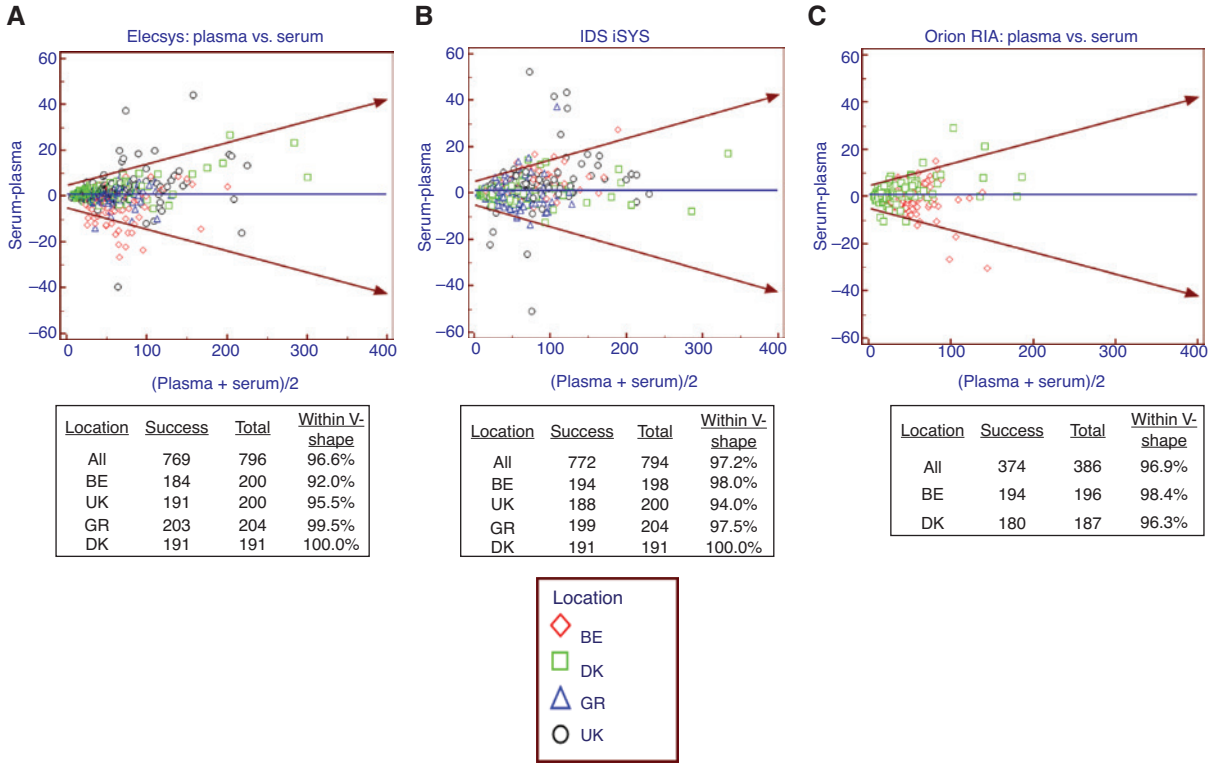

Figure 1: Bland-Altman plots and V-shape models comparing values observed in serum vs. plasma with Roche Elecsys (A), IDS iSYS (B) and Orion RIA (C).
More than 95% of the observations fit within the limits showing that the results are equivalent in serum and plasma.

Table 4: Passing-Bablok regression analyses of PINP values performed with Roche Elecsys, IDS iSYS and Orion RIA assays on serum and plasma specimens.

Plasma

								X
				IDS iSYS				**Orion RIA**
Y				**All**				**All**
Roche Elecsys								
Slope (95% CI)				0.86 (0.85; 0.87)				1.24 (1.21; 1.27)
Intercept (95% CI)				2.9 (2.5; 3.2)				−0.2 (−1.1; 0.6)
n				794				382
IDS iSYS								
Slope (95% CI)								1.45 (1.42; 1.48)
Intercept (95% CI)								−3.7 (−4.5; −2.9)
n								381
								X
				iSYS				**Orion RIA**
Y	**BE**	**UK**	**GR**	**DK**	**BE**	**UK**	**GR**	**DK**
Roche Elecsys								
Slope (95% CI)	0.85 (0.83; 0.87)	0.86 (0.84; 0.89)	0.87 (0.85; 0.90)	0.89 (0.86; 0.92)	1.26 (1.21; 1.30)	NP	NP	1.34 (1.28; 1.41)
Intercept (95% CI)	2.4 (1.7; 3.0)	1.5 (0.5; 2.2)	4.1 (3.0; 5.2)	2.8 (2.3; 3.2)	−2.0 (−3.0; −0.8)			−0.8 (−1.8; 0.7)
n	198	200	204	192	199			183
IDS iSYS								
Slope (95% CI)					1.45 (1.41; 1.49)	NP	NP	1.48 (1.44; 1.53)
Intercept (95% CI)					−4.7 (−6.1; −3.5)			−4.0 (−4.9; −2.9)
n					199			183

Serum

								X
				IDS iSYS				**Orion RIA**
Y				**All**				**All**
Roche Elecsys								
Slope (95% CI)				0.91 (0.90; 092)				1.22 (1.19; 1.26)
Intercept (95% CI)				2.6 (2.2; 3.1)				0.0 (−0.7; 1.2)
n				794				387
IDS iSYS								
Slope (95% CI)								1.35 (1.32; 1.38)
Intercept (95% CI)								−3.2 (−3.8; −2.4)
n								386
								X
				IDS iSYS				**Orion RIA**
Y	**BE**	**UK**	**GR**	**DK**	**BE**	**UK**	**GR**	**DK**
Roche Elecsys								
Slope (95% CI)	0.89 (0.87; 0.92)	0.95 (0.92; 0.98)	0.89 (0.86; 0.92)	0.95 (0.93; 0.98)	1.24 (1.19; 1.29)	NP	NP	1.31 (1.26; 1.35)
Intercept (95% CI)	2.5 (1.5; 3.5)	0.8 (−0.5; 2.1)	3.7 (2.1; 5.1)	2.5 (1.9; 3.0)	−1.9 (−3.5; −0.2)			0.6 (−0.3; 1.5)
n	198	200	204	192	199			188
IDS iSYS								
Slope (95% CI)					1.38 (1.35; 1.42)	NP	NP	1.36 (1.31; 1.40)
Intercept (95% CI)					−4.6 (−5.8; −3.5)			−1.9 (−3.2; −1.1)
n					198			188

The upper line of result corresponds to the slope (95% CI) and the second line to the intercept (95% CI). NP, not performed.

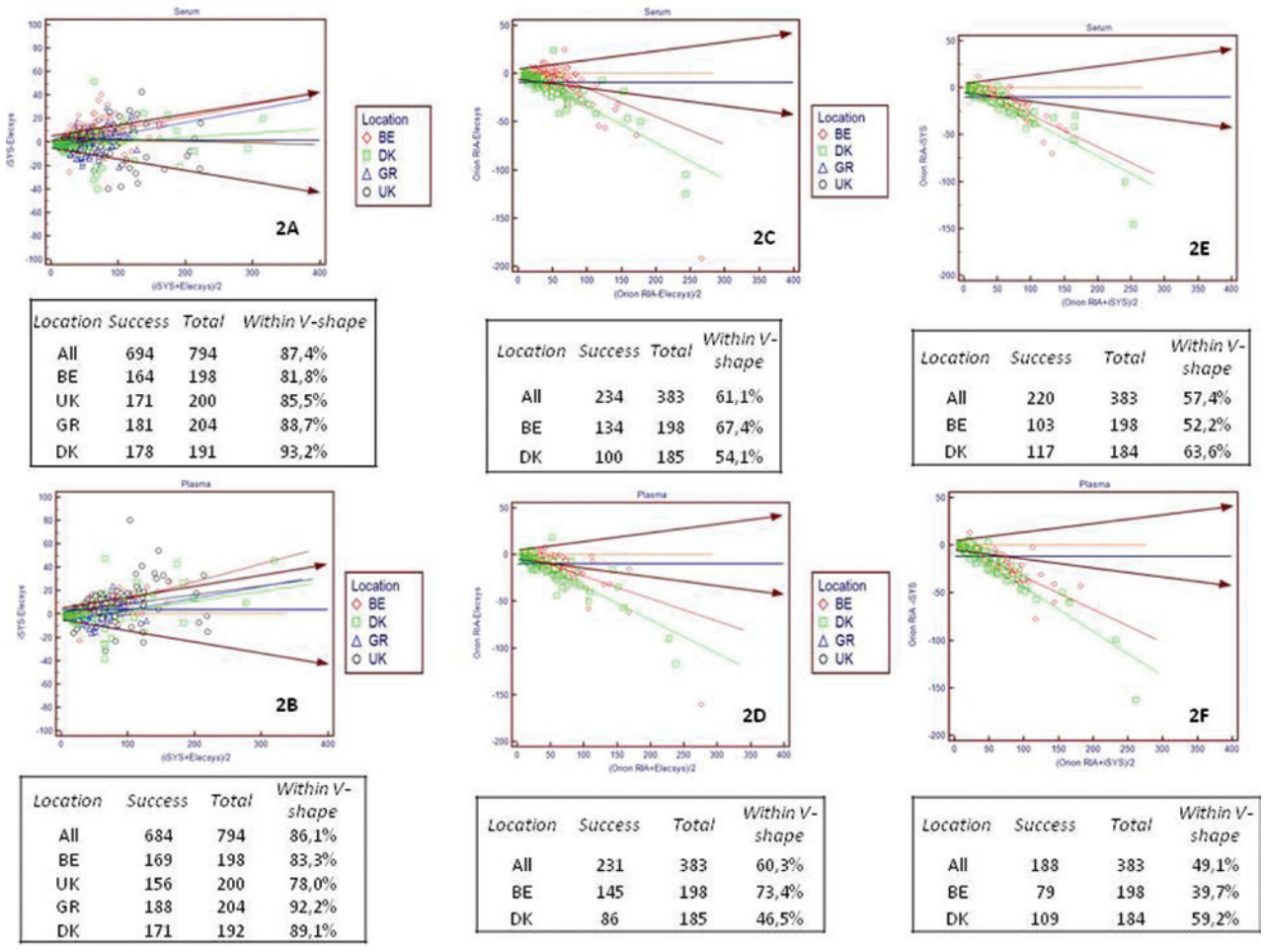

Figure 2: Bland-Altman and V-shape models comparing values observed in serum vs. plasma on iSYS vs. Elecsys (A and B), Elecsys vs. Orion RIA (C and D) and iSYS vs. Orion RIA (E and F).
The Bland-Altman plots show a significant proportional difference between Orion RIA and both automated methods from Roche and IDS, which show a rather good agreement.

RIA-0.0 (Table 3). The Bland-Altman plots (Figure 2C and D) shows an important proportional bias and the V-shape limits show an overall agreement of 61.1% for serum vs. 60.3% for plasma (χ^2 = NS). The percentage of agreement was significantly higher in BE than in DK.

IDS iSYS vs. Orion RIA

In plasma, the Passing-Bablok regressions on the relationship between IDS iSYS and Orion RIA was IDS iSYS = 1.45 × Orion RIA-3.7 and in serum, the relation was IDS iSYS = 1.35 × Orion RIA-3.2 (Table 3). The Bland-Altman plots also show an important proportional bias (Figure 2E and F) and the V-shape limits showing percentages of samples between the limits of 57.4% in serum vs. 49.1% in plasma (χ^2 = 0.09).

To identify factors contributing to the variation between the PINP values generated by the three assays, we calculated multivariable models with age, sex, center, BMI, weight, height, BMD, fasting status and medication as independent covariates (data not shown). Due to the excellent correlation between the measurements (univariate $R^2 \geq 0.97$), the multivariate models only marginally improved the results. The concordance correlation coefficient, which takes precision and accuracy into consideration is substantial (>0.95) when we compared the two automated methods, whereas it was poor (<0.90) when they were compared with the Orion RIA (Table 5), either for serum or plasma. The explanation for the poor performance when Orion RIA was taken into consideration was not due to the precision (ρ), which was >0.97, but rather to a poor accuracy between the manual and automated methods whereas it was of 0.99 when both automated methods were compared together.

Table 5: Concordance correlation coefficients (CCC), Pearson's correlation coefficient (ρ), which measures the precision and bias correction factor (C_b) which measures the accuracy.

	IDS iSYS	Orion RIA
	All	All
Plasma		
Roche Elecsys		
CCC (95% CI)	0.9649 (0.9604–0.9689)	0.8905 (0.8746–0.9044)
ρ	0.9765	0.9713
C_b	0.9881	0.9168
Strength of agreement	Substantial	Poor
IDS iSYS		
CCC (95% CI)		0.8386 (0.8193–0.8560)
ρ		0.9784
C_b		0.8572
Strength of agreement		Poor
Serum		
Roche Elecsys		
CCC (95% CI)	0.9706 (0.9664–0.9743)	0.8883 (0.8719–0.9027)
ρ	0.9722	0.9630
C_b	0.9984	0.9224
Strength of agreement	Substantial	Poor
IDS iSYS		
CCC (95% CI)		0.8760 (0.8605–0.8898)
ρ		0.9812
C_b		0.8927
Strength of agreement		Poor

The CCC evaluates the degree to which pairs of observation fall on the 45° line through the origin and corresponds to $\rho \cdot C_b$. NP, not performed.

Discussion

Bone turnover markers are commonly used to monitor osteoporotic patients' response to pharmacological treatment and monitoring compliance [3]. They have the great advantage, over bone mineral densitometry, to change rapidly according to treatment and these changes have been shown to correlate with reduction in fracture risk [22]. Hence, the IOF and IFCC have established a Working Group for the standardization of bone marker assays, whose task was to standardize or harmonize (as technically feasible or appropriate at this time) bone markers assays available for routine and research use, in serum and EDTA plasma. The WG has decided to compare the results of BTM in a very well-defined multicenter cohort of approximately 800 patients attending osteoporosis clinics. This comparison was achieved in serum and plasma, with the different commercially available methods that are routinely used. In this study, we report the results of PINP evaluation. PINP is a bone formation marker which presents very interesting features, namely a very limited impact of food consumption [18] and circadian variation [23] on the results. PINP has thus been recommended by the IOF-IFCC WG as the reference bone formation marker. PINP is present in two major forms, the trimeric form and a monomeric one, this latter being elevated in patients suffering from chronic renal failure.

Different reports have already evaluated PINP as measured by IDS iSYS, Roche Elecsys and Orion RIA. In 72 self-reported healthy volunteers and 55 patients suffering from rheumatoid arthritis (RA), Wheater et al. found a median difference of 2.0 µg/L between Elecsys total PINP and iSYS intact PINP in serum, with Roche giving significantly higher PINP concentrations compared to IDS ($p < 0.001$) [9]. In our study, the median difference observed between the two methods was 0.25 µg/L (not significant). The Passing-Bablok regression equation in Wheater's et al. study was $\text{IDS} = 0.98 \times \text{Roche} - 1.4$; with a slope not significantly different from 1.0, the intercept being significantly different from 0.0, whereas we found that the slope and intercept of the Passing-Bablok in our study were significantly different from 1 to 0, may be due to a higher number of observations and a higher range of measurement. The Bland-Altman plot showed a bias of 2.6 (Roche-IDS) with 95% limits of agreement ranging from –13.1 to 18.2 but the authors showed positive differences between total and intact assays at higher (>80 µg/L) levels, mainly found in patients suffering from RA. This gave a funnel shape of the Bland-Altman plot and the discrepancy was explained by the presence of monomers in RA patients. We did not have the information on RA in our patients, but the Bland-Altman plot showed a bias of 1.3 with 95% limits of agreement ranging from –15.1 to 17.7. In 2008, Garnero et al. [13] found that, in 59 healthy pre- and postmenopausal women, serum Elecsys values were on average 9.8 µg/L higher than those obtained by Orion RIA and that was a trend in increased difference between the two methods with increasing PINP concentrations. This is in line with our results as we observed a median difference of 12.5 µg/L and an important proportional bias. Like us, Koivula et al. [10] observed an excellent correlation between IDS iSYS, Roche Elecsys and Orion RIA. Finally, one of us compared the Roche and IDS assays in a large population of 2308 individuals. The result showed a R^2 of 0.8573 (much lower than our results) and a significant mean difference of 3 µg/L.

These findings combined to the results we present in this study clearly show a significant proportional difference

between Orion RIA and both automated methods. This bias can certainly be due to a difference in the assignment of the calibrator's values. As there are excellent correlations observed between the methods, the good news is that a harmonization of the methods should be possible. This harmonization will however be restricted to patients presenting GFR above 30 mL/min/1.73 m^2 as below this threshold, monomers start to accumulate and interfere with the total PINP assay from Roche. The next steps should thus include the preparation of a commutable international reference material for common calibration of the different assays and the development of a reference method as needed.

Acknowledgments: We acknowledge the support of Roche Diagnostics International Ltd, Immunodiagnostics Systems Holdings plc and Orion Diagnostica Oy for financial support and supply of reagents, calibrators and internal quality control specimens for the PINP assays.
Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Hernlund E, Svedbom A, Ivergård M, Compston J, Cooper C, Stenmark J, et al. Osteoporosis in the European Union: medical management, epidemiology and economic burden: a report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). Arch Osteoporos 2013;8:136.
2. Vasikaran S, Cooper C, Eastell R, Griesmacher A, Morris HA, Trenti T, et al. International Osteoporosis Foundation and International Federation of Clinical Chemistry and Laboratory Medicine Position on bone marker standards in osteoporosis. Clin Chem Lab Med 2011;49:1271–4.
3. Diez-Perez A, Naylor KE, Abrahamsen B, Agnusdei D, Brandi ML, Cooper C, et al. International Osteoporosis Foundation and European Calcified Tissue Society Working Group. Recommendations for the screening of adherence to oral bisphosphonates. Osteoporos Int 2017;28:767–74.
4. Koivula MK, Risteli L, Risteli J. Measurement of aminoterminal propeptide of type I procollagen (PINP) in serum. Clin Biochem 2012;45:920–7.
5. Szulc P, Naylor K, Hoyle NR, Eastell R, Leary ET. Use of CTX-I and PINP as bone turnover markers: National Bone Health Alliance recommendations to standardize sample handling and patient preparation to reduce pre-analytical variability. Osteoporos Int 2017;28:2541–56.
6. Cavalier E, Lukas P, Carlisi A, Gadisseur R, Delanaye P. Aminoterminal propeptide of type I procollagen (PINP) in chronic kidney disease patients: the assay matters. Clin Chim Acta 2013;425:117–8.
7. Koivula MK, Ruotsalainen V, Björkman M, Nurmenniemi S, Ikäheimo R, Savolainen K, et al. Difference between total and intact assays for N-terminal propeptide of type I procollagen reflects degradation of pN-collagen rather than denaturation of intact propeptide. Ann Clin Biochem 2010;47:67–71.
8. Marin L, Koivula MK, Jukkola-Vuorinen A, Leino A, Risteli J. Comparison of total and intact aminoterminal propeptide of type I procollagen assays in patients with breast cancer with or without bone metastases. Ann Clin Biochem 2011;48:447–51.
9. Wheater G, Goodrum C, Tuck SP, Datta HK, van Laar JM. Method-specific differences in β-isomerised carboxy-terminal cross-linking telopeptide of type I collagen and procollagen type I amino-terminal propeptide using two fully automated immunoassays. Clin Chem Lab Med 2014;52:135–8.
10. Koivula MK, Richardson J, Leino A, Valleala H, Griffiths K, Barnes A, et al. Validation of an automated intact N-terminal propeptide of type I procollagen (PINP) assay. Clin Biochem 2010;43:1453–7.
11. Jørgensen NR, Møllehave LT, Hansen YB, Quardon N, Lylloff L, Linneberg A. Comparison of two automated assays of BTM (CTX and P1NP) and reference intervals in a Danish population. Osteoporos Int 2017;28:2103–13.
12. Morovat A, Catchpole A, Meurisse A, Carlisi A, Bekaert A-C, Rousselle O, et al. IDS iSYS automated intact procollagen-1-N-terminus pro-peptide assay: method evaluation and reference intervals in adults and children. Clin Chem Lab Med 2013;51:2009–18.
13. Garnero P, Vergnaud P, Hoyle N. Evaluation of a fully automated serum assay for total N-terminal propeptide of type I collagen in postmenopausal osteoporosis. Clin Chem 2008;54:188–96.
14. Genant HK, Grampp S, Glüer CC, Faulkner KG, Jergas M, Engelke K, et al. Universal standardization for dual X-ray absorptiometry: patient and phantom cross-calibration results. J Bone Miner Res 1994;9:1503–14.
15. CLSI. EP15-A3 User Verification of Precision and Estimation of Bias. 2014.
16. Alvarez L, RicOs C, Peris P, GuaNabens N, Monegal A, Pons F, et al. Components of biological variation of biochemical markers of bone turnover in Paget's bone disease. Bone 2000;26:571–6.
17. Hannon R, Blumsohn A, Naylor K, Eastell R. Response of biochemical markers of bone turnover to hormone replacement therapy: impact of biological variability. J Bone Min Res 1998;13:1124–33.
18. Clowes JA, Hannon RA, Yap TS, Hoyle NR, Blumsohn A, Eastell R. Effect of feeding on bone turnover markers and its impact on biological variability of measurements. Bone 2002;30:886–90.
19. Rogers A, Glover SJ, Eastell R. A randomised, double-blinded, placebo-controlled, trial to determine the individual response

in bone turnover markers to lasofoxifene therapy. Bone 2009;45:1044–52.
20. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255–68.
21. McBride GB. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Rep 2005;HAM2005-06:14.
22. Morris HA, Eastell R, Jorgesen NR, Cavalier E, Vasikaran S, Chubb SA, et al. Clinical usefulness of bone turnover marker concentrations in osteoporosis. Clin Chim Acta 2017;467:34–41.
23. Redmond J, Fulford AJ, Jarjou L, Zhou B, Prentice A, Schoenmakers I. Diurnal rhythms of bone turnover markers in three ethnic groups. J Clin Endocrinol Metab 2016;101:3222–30.