Peer Review File

Article information: http://dx.doi.org/ 10.21037/atm-19-4498

Reviewer A:

The authors have published more than 10 papers in different journals with the same methodology, and on the same cohort.
This time, it turns to be the biological variation of PTH, which is obviously of clinical interest.

Comment 1: I have nothing to add to the robust methodology, but it turns, its description should definitely be reduced (lines 50-85 and 30 lines also in the data analysis). The authors can refer to one of their many papers.
Reply 1: we thank the reviewer for the suggestion; we have reduced the methodology description both in the Background and in the Data Analysis.

Comment 2: Line 51, I believe that the correct name is "The International Federation of Clinical Chemistry and Laboratory Medicine Committee for Bone Metabolism (C-BM).
Reply 2: we thank the reviewer for pointing this out; The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) working group for PTH has been corrected to "The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Committee for Bone Metabolism (C-BM)".

Comment 3: Do any patients take bisphosphonates and/or calcium supplements? Do you have the 25-OHD of the patients/do patients take supplements?
Reply 3: we thank the reviewer for the pertinent observation; we have verified and in the EuBIVAS population only 5 subjects declared to take vitamin supplements: 4 from Italy and 1 from the Netherland. Of these, only one declared to take vitamin D, the other 4 subjects declared to take general vitamin supplements. In any case, the biological variation data of the 5 subjects declared to take vitamins were compared to the rest of the population, and no difference has been found.

Comment 4: According to the upper limit of normality of the assay (35 pg/mL), many patients suffer from secondary hyperparathyroidism (74/90 cross the line, 20 are clearly above the ULN). Questions: can they still be considered in "good-health" and is there a difference between BV of patients with secondary hyperpara compared to those who do not present biological signs of secondary hyperpara?

Reply 4: According to the manufacturer, the reference interval of the assay is 14.9 pg/mL – 56.9 pg/mL (2.5° and 97.5° percentile) with 31.3 pg/mL as the mean value. Considering this, only 7 subjects showed PTH mean values higher than 56.9 pg/mL but these cannot be considered as outliers for the Dixon q-test. In addition, we have evaluated the CV_I (95% CI) for these 7 subjects: 14.6 % (12.2-17.6), which does not differ from the CV_I (95% CI) estimate for the other 84 subjects: 14.7% (13.9-15.5).

Comment 5: Smoking is known to affect PTH. Is there a difference in BV between subjects who smoke and those who don't?
Reply 5: we thank the reviewer for the pertinent observation; we have verified that the BV of subjects who smoked does not differ from that estimated in subjects who were non-smokers; in smokers (20 subjects) the CV_I (95% CI) = 14.8% (13.4-16.7); while in non-smokers (71 subjects) CV_I (95% CI) = 14.6% (13.8-15.5).

Comment 6: Calcium is the main driver of PTH. Authors have published about biological variation of calcium in the same subjects. Can they speculate on how much biological variation of PTH is explained by biological variation of calcium? Especially by differentiating subjects with higher PTH values compared to those with lower PTH values?
Reply 6: we thank the reviewer for the suggestion. We have examined the relation between the Calcium and PTH 1-84 CV_I estimates and found a weak correlation, most evident for subjects with higher concentration of PTH 1-84. We have added a figure displaying this to the supplementary data.

Reviewer B:

Comment 1: Page 4, Introduction, line 31-37:
The description of PTH pathophysiology is rather simple and should be shortened or replaced by a better description of the pathophysiology.
Reply 1: we thank the reviewer for the suggestion; the description of PTH pathophysiology has been replaced with a better description.

Comment 2: Page 4, line 41: Ref. No. 2 2009 KDIGO guideline was updated in 2017 (Kidney International Supplements (2017) 7, 1–59) and this reference should be mentioned.
Reply 2: we thank the reviewer for the suggestion; the reference N° 2 has been replaced with the following: KDIGO 2017 Clinical Practice Guideline Update for the Diagnosis, Evaluation, Prevention, and Treatment of Chronic Kidney Disease–Mineral and Bone Disorder (CKD-MBD). Kidney Int Suppl 2017; 7(1):S1-59.

Comment 3: Page 7, line 101:
Serum tubes were used in the experiment; however, recommended sample for Cobas Elecsys PTH (1-84) is plasma because of the better stability of the sample. Authors should comment on this either in the Discussion section or Limitations.
Reply 3: we thank the reviewer for the pertinent observation; we have added the following sentence to the limitation section: "In addition, the analyses have been performed on serum kept at room temperature for a maximum of two hours before centrifugation, this procedure may have introduced some small degradation of PTH that reasonably can be considered acceptable (as reported by Hanon et al. (42) and Dupuy et al. (43)). In any case, considering that the possible PTH degradation would be proportional in each sample, it is improbable that it may have increased the within-subject biological variation."
According to Hanon et al. (DOI: 10.1515/cclm-2013-0315) and Dupuy et al. (DOI: 10.1515/cclm-2017-0292), after blood collection, the stability of serum PTH at room temperature is considered acceptable until 2 h and 6 h, respectively, of delay before centrifugation. This pre-analytical requirement is satisfied by the EuBIVAS protocol of serum handling (Carobene et al., 2016, DOI: 10.1515/cclm-2016-0035).

Comment 4: Page 7, line 103:
Similarly, a better description on the stability of blood is needed ("… up to 2 h at room temperature and then centrifuged…") considering the manufacturer´s recommendation, it is recommended that when serum is needed the blood is centrifuged immediately. Again, this should be commented or mentioned in the

Limitations.

Reply 4: we thank the reviewer for the pertinent observation on PTH stability, even if the literature indicates a reasonable stability of PTH in serum samples (Hanon et al. (DOI: 10.1515/cclm-2013-0315) and Dupuy et al. (DOI: 10.1515/cclm-2017-0292), after blood collection, the stability of serum PTH at room temperature is considered acceptable until 2 h and 6 h, respectively, of delay before centrifugation), we decided to follow the reviewer's suggestion adding to Limitation the following sentence: "The analyses have been performed on serum kept at room temperature for a maximum of two hours before centrifugation, this procedure may have introduced some small degradation of PTH that reasonably can be considered acceptable (as reported by Hanon et al. (42) and Dupuy et al. (43)). In any case, considering that the possible PTH degradation would be proportional in each sample, it is improbable that it may have increased the within-subject biological variation."

Comment 5: Page 7, line 104:
The serum samples were stored at -80 degrees Celsius before shipping to the coordinating centre. It should be clearly stated that the serum samples were frozen also during transport.
Reply 5: we agree with the reviewer suggestion; it has been stated in the text that the serum samples were frozen also during transport.

Comment 6: Page 9, line 161:
A new sentence should be added - authors should clearly declare whether the study was fully compliant with the BIVAC check-list (14 quality items of grade A or not?).
Reply 6: after the sentence at page 6, line 81-83 " …fully compliant with the BIVAC" we added its explanation in bracket "(all 14 quality items were scored as A)". Thank you for asking a clarification.

Comment 7: Page 12, line 220:
Are the authors able to predict biological variation results of their cohort after application of the Bayesian model?
Reply 7: In our opinion, the sentence in the paper at page 12, line 214-219 "Standard statistical approaches used for estimating BV, such as the CV-ANOVA applied in our study, are sensitive to "noisy data" and assume homogeneity of the within-participant CV. Thus, the EuBIVAS data set has been trimmed to achieve this for PTH, requiring 3-4% of data to be excluded. A recently published Bayesian model for estimating BV is robust to "noisy data" and is an

approach for delivering BV estimates without *the need for data trimming (39)"* is sufficiently exhaustive. Running a Bayesian analysis is outside the scope of the paper.
In any case we can state that performing the analysis on the full data set (ie. prior to excluding any outliers and other data points) gave similar estimates as on the "cleaned" data set. So, we can guess that the Bayesan approach for this data set would not provide different estimates.

Comment 8: Page 15, line 280.
The last sentence of the Study limitations is redundant, as only reagents of one manufacturer´s were used and the results are therefore valid only for Cobas PTH (1-84).
Reply 8: we do not completely agree with the reviewer observation, the BV estimates are specific for a component, different methods can give different CV_A, and therefore different RCVs, but not different CV_I. An example to support this can be found in Carobene et al., 2017 (DOI: 10.1373/clinchem.2017.275115), the EuBIVAs BV estimates for creatinine, obtained using two different methods, were similar.
We have clarified the sentence in the study limitations.

Comment 9: Page 24, line 415:
The use of the Pearson correlation is questionable.
Reply 9: we agree with the reviewer suggestion, the Pearson correlation coefficient has been removed from the description of Figure 3.

Comment 10: Page 27, Table 2.
Authors should comment on the number of samples per individual (mean less than 10) and on the number of replicates per sample (mean less than 2, e.g. 1,87 in postmenopausal women). It means that duplicate analysis was not performed in all samples (line 432 and line 436).
Reply 10: we thank the reviewer for the observation.
It is correct that few samples were not analyzed in duplicate. In the EuBIVAS, analysis for around 20 different measurands were performed simultaneously on the same aliquot, thus, there may be a lack of sample for the second replicate depending on the initial volume in the aliquot, and random sampling errors. Moreover, the number of samples per individual and the replicates for samples are the values obtained after the elimination of outliers following variance homogeneity assessment. For all these reasons, the median value of 1.87 replicates for sample is reasonable. In addition, for 13 subjects the number of samples was less than 10, in fact 77 participants completed all 10 collections, 10

completed 9, 2 completed 8, and 2 completed 7. We have added this last information to the method session.

Comment 11: Supplementary data, Table 2. According to Braga (Crit Rev Clin Lab Sci, 2016; 53(5): 313–325), smoking can be used as an exclusion criterion. There were 3 heavy smokers and 17 smokers (1 – 10 cigarettes per day). Additionally, 11 men and 11 women declared alcohol consumption 1 – 2 U/day, i.e. 8 – 16 g of pure alcohol (exclusion criterion by Braga is above 10 g of ethanol/day). Therefore, the tested population is far from "healthy" population. This should be at least discussed or a subanalysis after application of these exclusion criteria should be done.

Reply 11: We have verified, the BV of subjects who smoke do not differ from subjects who don't smoke. In particular, smokers (20 subjects) CV_I (95% CI) = 14.8% (13.4-16.7) while the non-smokers (71 subjects) CV_I (95% CI) = 14.6% (13.8-15.5).

We have also verified, the BV of subjects who have an alcohol intake of 1-2U/day do not differ from subjects who don't have. In particular, the CV_I (95% CI) of the 22 subjects who have an alcohol intake of 1-2U/day is 14.6% (13.1-16.3); while for the 69 subjects who don't have an alcohol intake of 1-2U/day is 14.6% (13.8-15.5).

The EuBIVAS population was selected a priori as healthy (see Carobene et al., 2016, DOI: 10.1515/cclm-2016-0035) because it was expected for a large number of measurands. So, for each measurand, we have checked with statistical analysis whether or not to delete a specific subject.