Circulating Sex Steroid Measurements of Men by Mass Spectrometry Are Highly Reproducible after Prolonged Frozen Storage.

Revised manuscript

Short title: Steroids in Long-Term Frozen Storage

DJ Handelsman, R Desai, MJ Seibel, DG Le Couteur, RG Cumming

Andrology Laboratory, ANZAC Research Institute, University of Sydney

Correspondence:
Professor DJ Handelsman
Orchid identifier 0000-0002-4200-7476
ANZAC Research Institute
Sydney NSW 2139
Australia
T: +61-2-9767 9100
E: djh @anzac.edu.au

Key words: testosterone, dihydrotestosterone, estradiol, estrone, long-term storage, steroid mass spectrometry, reproducibility

Acknowledgements: Nil

Disclaimer/Conflict of interest: There was no external funding for this study. DJH has received institutional grants without personal income for investigator-initiated androgen pharmacology studies and provides expert testimony to anti-doping and professional standards tribunals and testosterone litigation. No other authors have any relevant declarations.

Words: 1481 References: 18 Tables: 2 Figures: 4

Abstract

Long-term studies investigating hormone-dependent cancers and reproductive health often require prolonged frozen storage of serum which assumes that the steroid molecules and measurements are stable over that time. Previous studies of reproducibility of circulating steroids have relied upon flawed historical rather than contemporaneous controls. We measured serum testosterone (T), dihydrotestosterone (DHT), estradiol (E2) and estrone (E1) in 150 randomly selected serum samples by liquid chromatography-mass spectrometry (LC-MS) from men 70 years or older (mean age 77 years) in the CHAMP study. The original measurements in 2009 were repeated 10 years later using the identical serum aliquot (having undergone 2-4 freeze-thaw cycles in the interim) in 2019 together with another never-thawed aliquot of the same serum sample. The results of all three sets of measurements were evaluated by Passing-Bablok regression and Bland-Altman difference analysis. Serum androgens (T, DHT) and estrogens (E2, E1) measured by LC-MS display excellent reproducibility when stored for 10 years at -80 C without thawing. Serum T and DHT displayed high level of reproducibility across all three sets of measurements. Multiple freeze-thaw cycles over those storage conditions do not significantly affect serum T, DHT and E1 concentrations but produce a modest increase (21%) in serum E2 measurements.

Highlights

- Reproducibility of steroid measurement in long-term frozen storage is crucial for long-term epidemiological research studies
- Using liquid chromatography mass spectrometry serum testosterone, dihydrotestosterone estradiol and estrone were highly reproducible after 10 years unthawed frozen storage as well as after multiple freeze-thaw cycles, apart from small increases in serum estradiol

Introduction

Measuring circulating steroids has pivotal importance to clinical and epidemiological research in hormone-dependent cancers and reproductive health. This requires long-term frozen storage of serum (or plasma) samples and analysis that assumes the stability of those analytes and measurements. The ideal test of stability of steroid measurements is to measure the same serum sample before and then at various times afterwards in prolonged frozen storage using the same measurement method; however, that is a challenging objective to conduct prospectively requiring safe storage and consistent measurement methodology that may change over time. Consequently, previous studies investigating the stability of steroids in long-term frozen storage have relied upon suboptimal design. Rather than using contemporaneous controls, they have relied on surrogate, historical controls such as samples obtained at different times from other participants in the cohort *(1-6)* or reference ranges from comparable studies in other centres and laboratories *(7)*. Most studies have focused on female samples *(2-4, 6, 7)* and used steroid immunoassays which are subject to non-specificity from structurally related cross-reacting precursors and/or metabolites as well as matrix effects in non-extraction direct assays *(1-4, 6, 7)*. Only two studies investigated male sample *(1, 5)* with only one using liquid chromatography-mass spectrometry (LC-MS) methods *(5)*. In this study we aimed to repeat serum steroid measurements by the same LC-MS method using different aliquots of the same serum samples kept in frozen storage at -80 C for 10 years. We measured serum androgens, testosterone (T) and dihydrotestosterone (DHT), and estrogens, estradiol (E2) and estrone (E1), in the same, multi-thawed aliquot together with a never-thawed aliquot of the original serum sample.

Materials and Methods

Samples

CHAMP is a longitudinal epidemiological cohort study of the health outcomes of 1705 community-dwelling men 70 years and older *(8)*. Serum samples (n=150) from the baseline CHAMP survey were drawn randomly from aliquots stored at -80 C in freezers subject to

continuously monitored by an automated building monitoring system without any freezer failure for over 10 years. The serum samples were originally obtained with ethical approval from the CHAMP study from 2008-9 and first analysed in 2009. The same aliquots used in 2009 (which in the interim underwent between 2-4 freeze-thaw cycles) were re-measured in March 2019. Finally, a fresh never-thawed aliquot of the same serum sample measured on the previous two runs (2009, March 2019) was then re-measured in June 2019.

Serum steroid measurement

Serum T, DHT, E2 and E1 were measured in a single run using a non-derivatization LC-MS method described and validated in detail elsewhere *(9)*. Minor changes between the methods used in 2009 and 2019 comprised a change in the liquid-liquid extraction method (hexane/ethyl acetate to methyl tert-butyl ether) and introduction of ultra-pressure replacing high pressure liquid chromatography with corresponding smaller injection volume. These changes were validated according to FDA criteria *(10)* as reported previously (see supplementary methods in *(11)*) to provide the same results. Overlapping multi-level quality control samples over the years confirm no drift in these analytes over the years of storage *(12)*.

Data analysis

Comparisons between measurements for the same serum sample at the three different times were analysed pairwise by Passing-Bablok regression and Bland-Altman difference analysis using NCSS (NCSS 2019, Kaysville, Utah, USA) and MedCalc (version 19.0.3, Ostend, Belgium) software. The non-parametric Passing-Bablok regression provided a slope and intercept with 95% confidence limits and a Spearman rank correlation coefficient (NCSS). Perpendicular residual standard deviation around the regression (MedCalc) were scaled by the global mean for the analyte to facilitate dimensionless comparison of variability between analytes and samples. The Bland-Altman plot provides a mean difference and its 95% confidence limits.

Results

Descriptive statistics of the three sets of measurements of aliquots of the same serum samples is provided in table 1. The reproducibility of individual serum samples measured on three occasions is shown in figures 1 (T), figure 2 (DHT), figure 3 (E2) and figure 4 (E1). The metrics from the Passing-Bablok regression and Bland-Altman difference analysis are listed in table 2.

The reproducibility of the androgens, T and DHT, were excellent with minimal, non-significant differences in the descriptive statistics and with high rank correlation. In the regression analyses the slopes did not deviate significantly from 1.0. A single regression intercept for DHT with a borderline (p=0.05) significant deviation from the expected value of 0 (table 2) is likely attributable to multiple comparisons. The latter is unlikely to be of importance as among the 30 statistical comparisons for androgens in table 2, one significant difference does not exceed the expectation of a chance finding arising from the multiple comparisons based on the conventional $p<0.05$ level of significance.

The reproducibility of estrogens was less satisfactory although not very evident in the between-person descriptive statistics (table 1) which showed that the multi-thawed aliquots displayed a 21% higher serum E2 and 1.4% lower serum E1 compared with the never-thawed aliquots before or after 10 year frozen storage at -80 C. Congruently in the regression analysis, serum E2 showed significantly higher concentrations in the multi-thawed aliquots compared with both the 2009 and 2019 never-thawed aliquots of the same serum samples whereas the never=thawed samples showed high reproducibility. For E1, the pattern was similar to that for serum E2 but the deviations of the multi-thawed aliquot were less marked.

Discussion

The present findings indicate that LC-MS measurement of bioactive androgens and estrogens are stable and reproducible for a period of at least 10 years in frozen storage at -80 C without thawing. These findings are as well controlled as it is feasible over prolonged periods. Crucially we used different aliquots of the same serum samples, including the same aliquot measured in 2009 after 2-4 freeze-thaw cycles as well as another never-thawed aliquot, measured at an interval of 10 years in the same laboratory using the same LC-MS method. The slight differences in LC-MS method between 2009 and 2019 representing contemporary technological improvements was rigorously validated to demonstrate that, within experimental error, similar results were produced without any long-term drift. This approach provides better control for the original measurements than previously used historical controls drawn at different times from within the same cohort or from external studies in other centres.

A strength of this study is the use of contemporaneous controls by measuring sex steroids in aliquots of the same serum samples stored over a 10-year interval rather than relying on historical controls. The reliance on historical controls can be misleading. For example, one study reporting that serum T measured by LC-MS may have increased over prolonged storage (22 years) in different cohorts of Norwegian men *(5)*. This interpretation assumes that the different cohorts recruited over the 22 years of the study were essentially identical. However, this interpretation is confounded by the population trends in Western countries of decreasing serum T reported in Scandinavia *(13)* and the USA *(14)* over those years. Hence, the downward trend in serum T in that Norwegian study over the period of the study probably reflects biological changes in the population sampled rather than in analytical variability of frozen stored samples.

The unexpected finding was that E2 and, to a lesser extent, E1 showed some deviations attributable to the multi-thawing of samples over prolonged frozen storage. These effects were relatively subtle and resulted surprisingly in small but systematic increases in serum E2. This difference was hard to discern in the cross-sectional statistics of between-person descriptive statistics such as relied upon by previous studies using historical controls. The reasons for the

deviations in estradiol in the multi-thawed aliquots are not known although the relative RSD (table 2) shows that even for the other three steroids, the perpendicular scatter of data around regressions between unthawed samples were somewhat tighter than those involving thawed aliquots. Previous studies of the impact of repeated freeze-thaw cycles on circulating steroids mostly show minimal effects over multiple freeze-thaw cycles but only using steroid immunoassays and over frozen storage for up to 3 months *(15-17)*. In these studies, serum T in the female range was stable over 10 freeze-thaw cycles *(17)*, serum T, DHT and E2 were stable over 12 freeze-thaw cycles *(15)* and serum T and DHT in a pool of male or of female sera were stable over 3 freeze-thaw cycles *(16)*. Hence it is not clear why the circulating estrogens are modified by freeze-thaw cycles during long-term storage. Nevertheless a chemical interaction between frozen-thawed steroids and the matrix over prolonged frozen storage is possible. For example, the impact of freeze-thaw cycles was congruent with the presence of 2 (E2), 1 (E1) and no (T , DHT) hydroxyl groups in the steroid; however, we did not measure estriol (E3) in the male samples which, if present, would have arisen by this mechanism. So, whether non-enzymatic hydroxylation or alternatively deconjugation of sulphated or glucuronide conjugates provide a chemical basis for the changes remains speculative. Another possibility of in vitro conversion of other steroids to E2, achieved in Nature solely by a single, highly conserved enzyme aromatase *(18)*, seems remote due to the complex chemical nature of aromatisation. The clinical significance of such measurement artefacts will depend on the nature of the study but could tend to nullify genuine estrogen effects.

This study provides both reassurance and caution for long-term clinical studies of reproductive health and hormone-dependent cancer. On the one hand the stability of androgens over at least a decade, and presumably indefinitely if the samples are stored without thawing, is reassuring for the validity of androgen measurement over long-term sample frozen storage. The changes in estrogens are modest in magnitude and tend to produce a small increase in serum E2 measurements while mostly preserving their between-person rank order of the results. If an internal correction is required, the results for any measured steroid including those not included in this reproducibility study, could be compared with those for serum T as a benchmark for stability.

190

191

References

1. Kley HK, Schlaghecke R, Kruskemper HL. Stability of steroids in plasma over a 10-year period. J Clin Chem Clin Biochem 1985;23:875-8.
2. Bolelli G, Muti P, Micheli A, Sciajno R, Franceschetti F, Krogh V, et al. Validity for epidemiological studies of long-term cryoconservation of steroid and protein hormones in serum and plasma. Cancer Epidemiol Biomarkers Prev 1995;4:509-13.
3. Missmer SA, Spiegelman D, Bertone-Johnson ER, Barbieri RL, Pollak MN, Hankinson SE. Reproducibility of plasma steroid hormones, prolactin, and insulin-like growth factor levels among premenopausal women over a 2- to 3-year period. Cancer Epidemiol Biomarkers Prev 2006;15:972-8.
4. Holl K, Lundin E, Kaasila M, Grankvist K, Afanasyeva Y, Hallmans G, et al. Effect of long-term storage on hormone measurements in samples from pregnant women: The experience of the finnish maternity cohort. Acta oncologica (Stockholm, Sweden) 2008;47:406-12.
5. Gislefoss RE, Grimsrud TK, Hoie K, Morkrid L. Stability of testosterone measured in male archival serum samples by two different methods. Scand J Clin Lab Invest 2012;72:555-62.
6. Jones ME, Schoemaker MJ, Rae M, Folkerd EJ, Dowsett M, Ashworth A, Swerdlow AJ. Reproducibility of estradiol and testosterone levels in postmenopausal women over 5 years: Results from the breakthrough generations study. Am J Epidemiol 2014;179:1128-33.
7. Stroud LR, Solomon C, Shenassa E, Papandonatos G, Niaura R, Lipsitt LP, et al. Long-term stability of maternal prenatal steroid hormones from the national collaborative perinatal project: Still valid after all these years. Psychoneuroendocrinology 2007;32:140-50.
8. Cumming RG, Handelsman D, Seibel MJ, Creasey H, Sambrook P, Waite L, et al. Cohort profile: The concord health and ageing in men project (champ). International Journal of Epidemiology 2009;38:374-8.
9. Harwood DT, Handelsman DJ. Development and validation of a sensitive liquid chromatography-tandem mass spectrometry assay to simultaneously measure androgens and estrogens in serum without derivatization. Clin Chim Acta 2009;409:78-84.
10. FDA CDER. Bioanalytical method validation. Vol., 2018.
11. Keski-Rahkonen P, Desai R, Jimenez M, Harwood DT, Handelsman DJ. Measurement of estradiol in human serum by lc-ms/ms using a novel estrogen-specific derivatization reagent. Anal Chem 2015;87:7180-6.

12. Hsu B, Cumming RG, Hirani V, Blyth FM, Naganathan V, Le Couteur DG, et al. Temporal trend in androgen status and androgen-sensitive outcomes in older men. J Clin Endocrinol Metab 2016;101:1836-46.
13. Andersson AM, Jensen TK, Juul A, Petersen JH, Jorgensen T, Skakkebaek NE. Secular decline in male testosterone and sex hormone binding globulin serum levels in danish population surveys. J Clin Endocrinol Metab 2007;92:4696-705.
14. Travison TG, Araujo AB, O'Donnell AB, Kupelian V, McKinlay JB. A population-level decline in serum testosterone levels in american men. J Clin Endocrinol Metab 2007;92:196-202.
15. Bauman JE. Stability of radioimmunoassayable steroid and protein hormones after repeated freeze-thaw cycles. Clin Chem 1982;28:2336-7.
16. Hsing AW, Comstock GW, Polk BF. Effect of repeated freezing and thawing on vitamins and hormones in serum. Clin Chem 1989;35:2145.
17. Gislefoss RE, Lauritzen M, Langseth H, Morkrid L. Effect of multiple freeze-thaw cycles on selected biochemical serum components. Clin Chem Lab Med 2017;55:967-73.
18. Di Nardo G, Gilardi G. Human aromatase: Perspectives in biochemistry and biotechnology. Biotechnol Appl Biochem 2013;60:92-101.

Table 1 – Descriptive statistics of steroid measurement of three sets of aliquots of the same serum samples stored frozen at -80 C for 10 years

	N	Mean	SE	Lo	Q1	Q2	Q3	Hi
Testosterone (ng/ml)								
2009 (never thawed)	150	4.11	0.16	0.01	3.13	4.11	5.20	11.3
3/2019 (multi-thawed)	150	4.17	0.17	0.01	3.09	4.16	5.38	12.1
6/2019 (never thawed)	150	4.22	0.16	0.03	3.19	4.14	5.30	10.8
DHT (ng/ml)								
2009 (never thawed)	135	0.39	0.02	0.11	0.24	0.35	0.49	1.05
3/2019 (multi-thawed)	140	0.41	0.02	0.05	0.27	0.37	0.52	1.27
6/2019 (never thawed)	147	0.36	0.02	0.05	0.22	0.33	0.47	1.19
Estradiol (pg/ml)								
2009 (never thawed)	140	24.0	0.68	5.1	18.0	23.2	29.3	50
3/2019 (multi-thawed)	142	29.1	1.17	6.0	18.8	29.0	37.0	87
6/2019 (never thawed)	144	23.8	0.73	5.0	18.0	23.0	30.1	51
Estrone (pg/ml)								
2009 (never thawed)	149	37.9	1.24	7.0	28.1	36.8	45.5	95
3/2019 (multi-thawed)	150	36.2	1.20	4.7	26.5	34.1	44.8	84
6/2019 (never thawed)	150	37.5	1.27	4.0	27.7	36.0	46.3	93

Lo refers to the minimum and Hi to the maximum concentrations. Q1, Q2 (median) and Q3 refer to quartiles of the distribution. 2009 refers to the original never-thawed serum aliquot measured in 2009, 3/19 refers to the multi-thawed aliquot of the same serum sample measured in March 2019 and 6/19 refers to the never-thawed aliquot of the same serum sample measured in 2019. Numbers less than 150 reflect samples with undetectable concentrations for that analyte except for one extreme outlier in serum E2 aliquot which was excluded from analysis as its asymmetrical peak shape indicated an unknown interference in the measurement.

Table 2 – Method Comparison Statistics from Passing-Bablok Regression and Bland-Altman Difference Analysis

	Slope	Intercept	R	RSD	Mean difference
Testosterone					
Original vs 3/19	1.02 [1.00,1.05]	-0.13 [-0.10,0.05]	0.98	0.30 [0.07]	-0.06 [-0.90, 0.78]
Original vs 6/19	1.02 [1.00,1.04]	0.02 [-0.04,0.10]	0.99	0.18 [0.04]	-0.11 [-0.66, 0.38]
3/19 vs 6/19	1.00 [0.96,1.03]	0.16 [-0.03,0.17]	0.97	0.33 0.08]	-0.06 [-0.98, 0.87]
DHT					
Original vs 3/19	0.98 [0.90,1.07]	**0.05 [0.02,0.08]**	0.89	0.065 [0.16]	0.04 [-0.14, 0.22]
Original vs 6/19	1.03 [0.97,1.09]	0.00 [-0.03,0.18]	0.93	0.056 [0.14]	0.00 [-0.15, 0.15]
3/19 vs 6/19	1.02 [0.92,1.18]	-0.04 [-0.08,0.00]	0.85	0.081 [0.20]	-0.04 [-0.26, 0.19]
Estradiol					
Original vs 3/19	**1.57 [1.39,1.85]**	**-10.3 [-16.2, -6.35]**	0.69	5.56 [0.21]	-4.93 [-23.8, 13.9]
Original vs 6/19	1.06 [1.00,1.14]	**-1.53 [-3.24, -0.25]**	0.94	2.11 [0.09]	-0.12 [-6.0, 5.8]
3/19 vs 6/19	**0.65 [0.55,0.75]**	**6.11 [3.50,8.46]**	0.69	5.85 [0.22]	5.15 [-14.5, 24.8]
Estrone					
Original vs 3/19	**0.94 [0.89,0.99]**	0.52 [-1.54,2.48]	0.93	3.94 [0.11]	-1.51 [-12.4, 9.43]
Original vs 6/19	1.00 [0.96,1.05]	-0.60 [-1.62,0.94]	0.97	2.64 [0.02]	-0.14 [-7.42, 7.14]
3/19 vs 6/19	1.07 [1.00, 1.15]	-1.45 [-4.16,1.20]	0.92	4.24 [0.11]	1.36 {-10.5, 13.2]

Original refers to the never-thawed serum aliquot originally measured in 2009, 3/19 refers to the multi-thawed aliquot of the same serum sample measured in March 2019 and 6/19 refers to the never-thawed aliquot of the same serum sample measured in 2019.

Statistically significant differences at 95% confidence level are indicated by bold highlighting. Slope and intercept with their 95% confidence limits in brackets are from the Passing-Bablok regression (NCSS). R is the Spearman rank correlation coefficient. RSD is standard deviation of the perpendicular residuals from the regression line of best fit (from MedCalc Passing-Bablok regression) NS is followed by a scaled RSD in brackets which is the RSD divided by the grand mean of the analytes to allow a dimensionless comparison for each analyte of the (perpendicular) scatter of deviations around the the regression line.

Figure legends

Figure 1 – **Plot of serum testosterone measurements of three sets of aliquots of the same serum samples measured stored with or without thawing and frozen at -80 C for up to 10 years**

Individual data scatter plots (filled circles) for measurements of serum testosterone the same 150 serum samples are shown in three columns. The left columns compare the original measurements in 2009 (x axis) with that of multi-thawed samples (y axis). The middle columns compare the original measurements in 2009 with a never thawed aliquot. The right columns compare the measurement of the multi-thawed samples (x axis) with that of the never thawed samples (y axis). The upper row shows the Passing-Bablok regression with the regression in a solid line and the line of identity as a dashed line. The lower row shows the Bland-Altman difference plots with the mean difference in a solid line, the line of identity as a dashed line and the 95% confidence limits of the difference in the shaded region. For further quantitative details of the analysis of the plots see table 1.

Figure 2 – **Plot of serum dihydrotestosterone measurements of three sets of aliquots of the same serum samples measured stored with or without thawing and frozen at -80 C for up to 10 years**

Individual data scatter plots (filled circles) and linear regression (solid line with line of identity in dashed line) for measurements of serum DHT the same 150 serum samples are shown in three columns. The left columns compare the original measurements in 2009 (x axis) with that of multi-thawed samples (y axis). The middle columns compare the original measurements in 2009 with a never thawed aliquot. The right columns compare the measurement of the multi-thawed samples (x axis) with that of the never thawed samples (y axis). The upper row shows the Passing-Bablok regression with the regression in a solid line and the line of identity as a dashed line. The lower row shows the Bland-Altman difference plots with the mean difference in a solid line, the line of identity as a dashed line and the 95% confidence limits of the difference in the shaded region. For further quantitative details of the analysis of the plots see table 1.

Figure 3 – **Plot of serum estradiol measurements of three sets of aliquots of the same serum samples measured stored with or without thawing and frozen at -80 C for up to 10 years**

Individual data scatter plots (filled circles) and linear regression (solid line with line of identity in dashed line) for measurements of serum estradiol the same 150 serum samples are shown in three columns. The left columns compare the original measurements in 2009 (x axis) with that of multi-thawed samples (y axis). The middle columns compare the original measurements in 2009 with a never thawed aliquot. The right columns compare the measurement of the multi-thawed samples (x axis) with that of the never thawed samples (y axis). The upper row shows the Passing-Bablok regression with the regression in a solid line and the line of identity as a dashed line. The lower row shows the Bland-Altman difference plots with the mean difference in a solid line, the line of identity as a dashed line and the 95% confidence limits of the difference in the shaded region. For further quantitative details of the analysis of the plots see table 1.

Figure 4 – **Plot of serum estrone measurements of three sets of aliquots of the same serum samples measured stored with or without thawing and frozen at -80 C for up to 10 years**

Individual data scatter plots (filled circles) and linear regression (solid line with line of identity in dashed line) for measurements of serum E1 the same 150 serum samples are shown in three columns. The left columns compare the original measurements in 2009 (x axis) with that of multi-thawed samples (y axis). The middle columns compare the original measurements in 2009 with a never thawed aliquot. The right columns compare the measurement of the multi-thawed samples (x axis) with that of the never thawed samples (y axis). The upper row shows the Passing-Bablok regression with the regression in a solid line and the line of identity as a dashed line. The lower row shows the Bland-Altman difference plots with the mean difference in a solid line, the line of identity as a dashed line and the 95% confidence limits of the difference in the shaded region. For further quantitative details of the analysis of the plots see table 1.

Figure 1

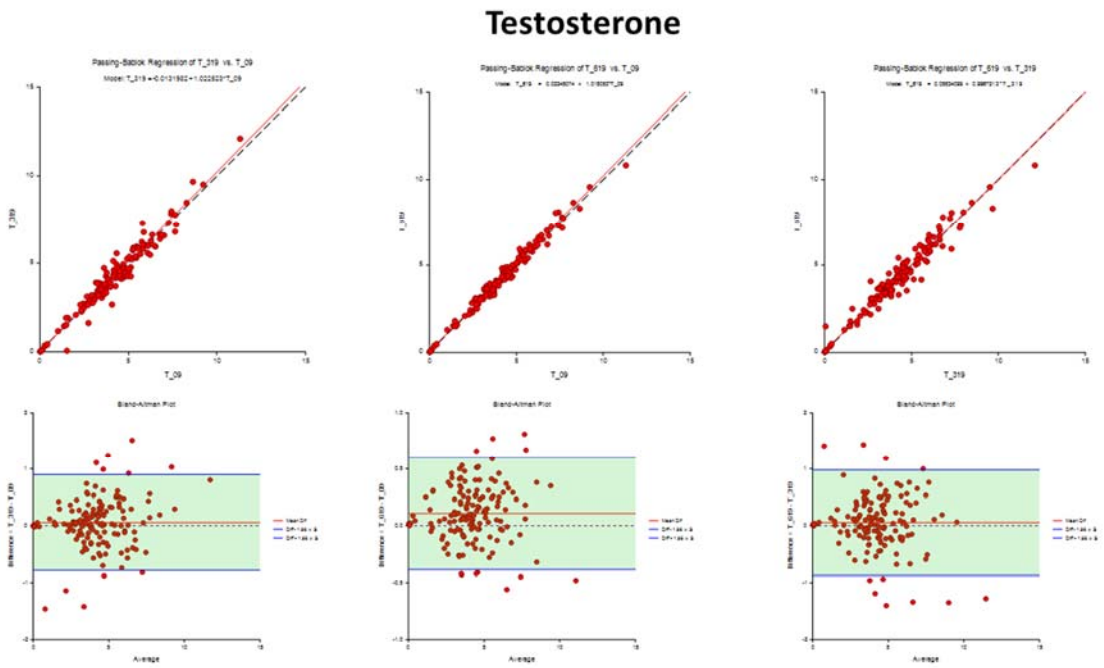

Figure 2

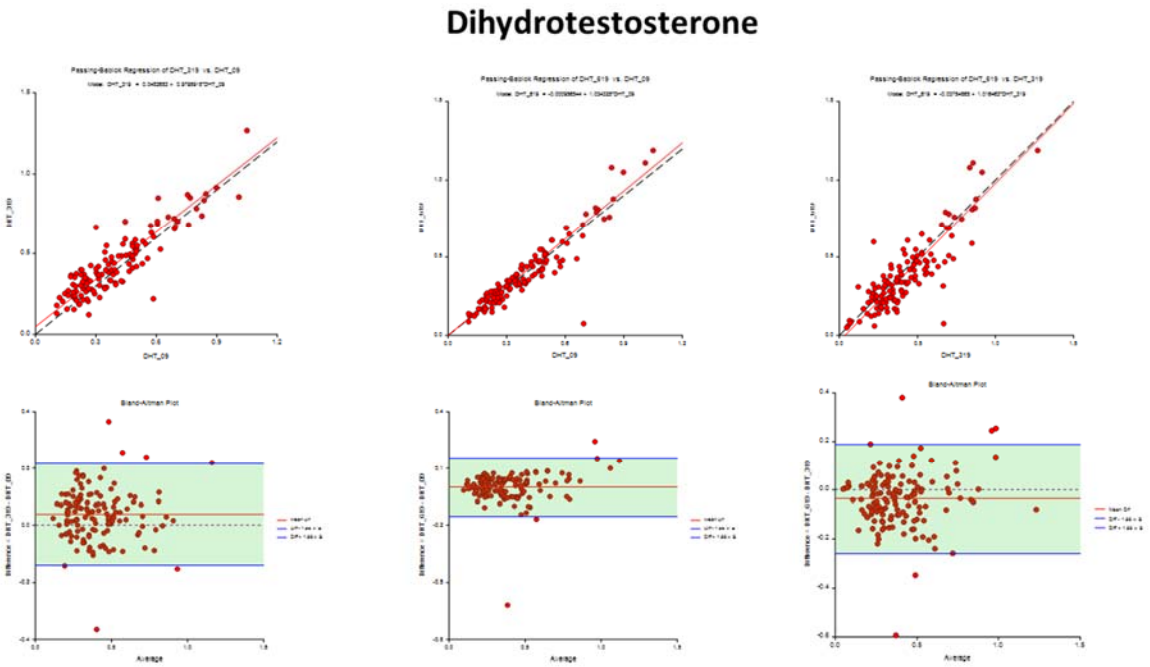

Figure 3

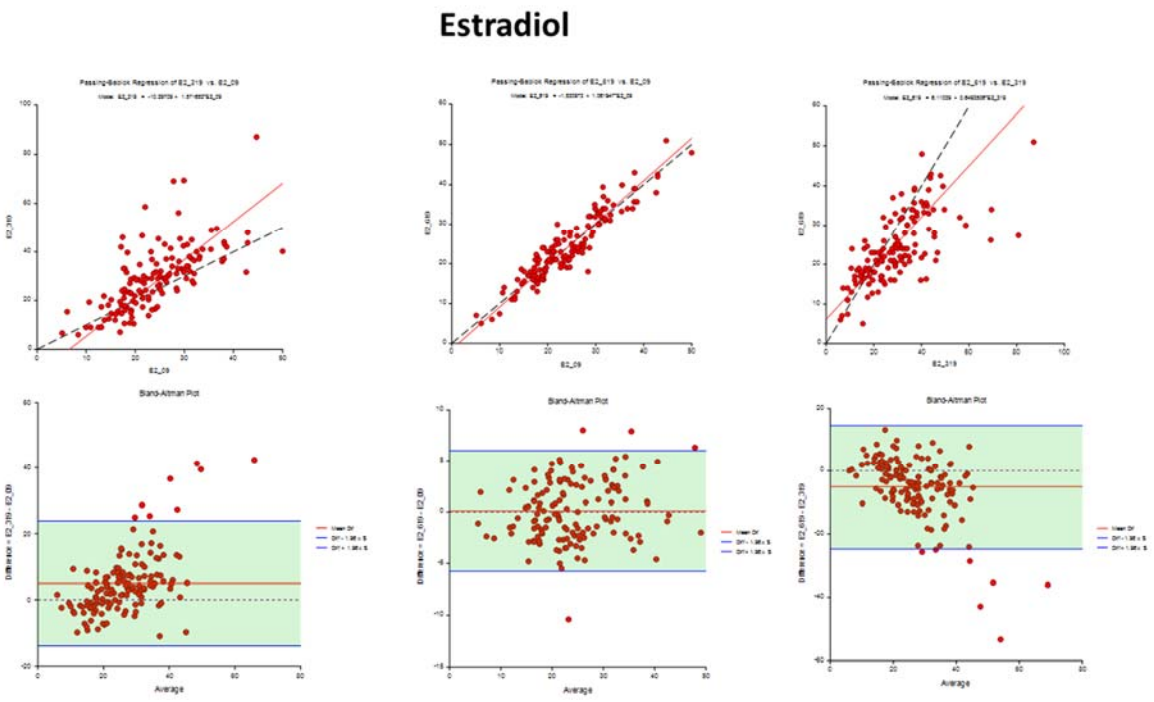

Figure 4