Analytical performance of free testosterone calculated by direct immunoluminometric method compared with Vermeulen equation: results from a clinical series.

Cristina Dalmiglio
Università degli Studi di Siena: Universita degli Studi di Siena
Alessio Bombardieri
Università degli Studi di Siena: Universita degli Studi di Siena
Elisa Mattii
Università degli Studi di Siena: Universita degli Studi di Siena
Fausta Sestini
Siena University Hospital: Azienda Ospedaliera Universitaria Senese
Carla Fioravanti
Siena University Hospital: Azienda Ospedaliera Universitaria Senese
Maria Grazia Castagna
University of Siena: Universita degli Studi di Siena
Marcello Fiorini
Siena University Hospital: Azienda Ospedaliera Universitaria Senese
Francesco Dotta
Università degli Studi di Siena: Universita degli Studi di Siena
Silvia Cantara (cantara@unisi.it)
University of Siena: Universita degli Studi di Siena https://orcid.org/0000-0002-5741-295X

Research Article

Keywords: free testosterone, clinical series, Vermeulen formula, direct method

Posted Date: September 1st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3241874/v1

Version of Record: A version of this preprint was published at Hormones on January 5th, 2024. See the published version at https://doi.org/10.1007/s42000-023-00522-x.

Abstract

Introduction: Testosterone is a hormone crucial for primary and secondary sexual development in both males and females. Free testosterone (FT) represents the biologically active form of T, and its measurement holds significant importance in clinical practice. While equilibrium dialysis or ultrafiltration are considered the gold standard for FT assessment, these methods are expensive and not widely accessible. As an alternative, the Vermeulen formula is a commonly utilized calculated method.

Methods: This clinical study involved 190 consecutive patients, comparing FT levels obtained through direct immunoluminometric assay and the Vermeulen formula. The comparison was performed using Passing-Bablok, Deming regressions, as well as the Bland-Altman plot. Sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were assessed.

Results: The calculated method, employing the Vermeulen formula, was considered the gold standard. Passing-Bablok regression indicated a good agreement between the two methods, with slopes close to 1. The Bland-Altman demonstrated overall agreement, but a potential proportional bias was observed in females. Deming regression confirmed excellent agreement and reliable estimates. Sensitivity and specificity analysis revealed that the direct method had a sensitivity of 75.0% and specificity of 93.4% considering all patients. However, sensitivity improved to 81.0% in males and dropped to 18.2% in females, likely due to the low number of true positive cases.

Conclusion: In conclusion, the direct method exhibited comparable performance to the calculated method, but caution should be exercised when interpreting results, particularly in females. Further studies are necessary to validate its sensitivity and specificity in larger series.

Introduction

Testosterone (T) is a hormone that is primarily produced in the testes of males and in the ovaries of females (1). In men, T is responsible for primary sexual development, including testicular descent, spermatogenesis, enlargement of the penis/testes and fostering libido (2). It also drives the development of secondary male characteristics like vocal changes, male hair patterns and growth spurts in puberty (2). Although women have much lower testosterone levels than men, in female T is responsible for maintaining bone density, muscle mass, and strength. It also plays a role in sexual desire, and it can help to regulate mood and cognitive function (3). Low T levels in men can be responsible of decreased libido and sexual dysfunction (4–6); decreased muscle mass and strength (7–9); increased body fat and decreased bone density (10–11); depression, irritability, fatigue and reducedcognitive functions (i.e. memory, concentration, and mental clarity) (12–13). Also in women, low levels of testosterone can lead to decreased muscle mass, increased body fat, decreased libido, and mood disturbances (14–16). Anyhow, in clinical practice, the evaluation of testosterone levels in women is mainly aimed at excluding an excess of its production as in the case of Congenital adrenal hyperplasia (CAH), Polycystic ovary syndrome (PCOS), androgen-producing neoplasms (17). Free testosterone (FT) refers to the fraction of

testosterone that is not bound to proteins (approximately 2% of total T), such as sex hormone-binding globulin (SHBG) and albumin. It is considered the biologically active form of testosterone because it can enter cells and bind to androgen receptors (18). Free testosterone levels are often used to assess testosterone status because they are thought to reflect the amount of testosterone that is available to exert physiological effects (18). There are several methods to calculate free testosterone levels, including direct measurement using equilibrium dialysis or ultrafiltration or calculated methods using mathematical models such as the Vermeulen Eq. (19). The accuracy and reliability of these methods can vary depending on the population being studied and the laboratory performing the test. Scientific guidelines (20–21) recommend using the direct measurement of free testosterone by equilibrium dialysis or ultrafiltration as the gold standard method for assessing testosterone status. However, these methods are not widely available and can be expensive. Therefore, the use of calculated or estimated free testosterone levels is often used as an alternative. The most accepted calculated method is the Vermeulen formula (22) developed by Dr. Alex Vermeulen in 1999. The equation considers the serum total testosterone level (expressed in nmol/L), the SHBG level (expressed in nmol/L) and the albumin level (considered as constant at 4.3 g/dL). It is well known that every laboratory test has a degree of error or variability that can influence the accuracy of the results (23). Thus, evaluating FT by combining three different biomarkers may increase the error degree and lead to misinterpretation of the results. Additionally, reference ranges for free testosterone levels can vary depending on the assay and laboratory used and should be interpreted with caution. For all these reasons, it is of paramount importance to validate other direct method for FT evaluation, more rapid and economic to reach the goal of result harmonization.

Patients and methods

Clinical series

We investigated a consecutive series of 190 blood samples (M:134) sent for measurement of FT from February to March 2023 to the Laboratory of Clinical and Translational Research-Endocrinology Section, University Hospital of Siena (Italy). The FT evaluation was part of patient diagnostic routine for which they have signed an informed consent.

Free Testosterone measurement

Each sample was divided and evaluated both with direct and calculated methods. All hormones were measured using automated immunoassay platforms. Total testosterone and SHBG were assayed with UniCel Dxi 800 (Beckman Coulter) using a competitive immunoassay. T reference range was 0.1–0.75 ng/ml for female and 2.4–7.81 ng/ml for male. T assay was designed to have within-laboratory imprecision as ≤ 0.14 ng/mL (0.49 nmol/L) SD at concentrations ≤ 1.4 ng/mL (4.9 nmol/L) and ≤ 10.0% CV at concentrations > 1.4 ng/mL (4.9 nmol/L). The limit of detection (LoD) and the limit of quantization (LoQ) were ≤ 1.39 nmol/L. SHBG reference intervals were 18–135 nmol/L for female and 13–90 nmol/L for male. SHBG LoD was 0.33 nmol/L and the assay exhibits total imprecision of < 7% at concentrations

greater than 2 nmol/L. To apply the Vermeulen formula, T values were first converted from ng/ml to nmol/L using an online conversion tool (https://unitslab.com/node/136). Then, the algorithm was applied considering albumin a constant value of 4.3 g/dL (https://www.omnicalculator.com/health/free-testosterone). For calculated FT, normal ranges were considered > 0.225 nmol/L (> 64.9 pg/ml) for males and < 0.039 nmol/L (< 11.25 pg/ml) for females. Free testosterone was also evaluated by a competitive immunoluminometric assay (Maglumi, Snibe) that shows an intra- and inter-assay coefficients of variation (CV) of 5.7% and 6.4%, respectively. For direct method, reference range were: males 15–50 pg/ml and females < 9.0 pg/ml.

Statistical analysis

Statistical analysis and graphs were performed using the MedCalc® program. Different approaches were used to calculate the concordance between the two methods. The Passing-Bablok regression was utilized to estimate the slope and intercept of the linear relationship between the values obtained from both methods. In this regression, the slope represents the level of agreement between the two methods, ideally close to 1, indicating perfect agreement. The intercept should be close to 0, indicating no systematic difference between the methods. Concordance within the calculated and direct FT measurements was also assessed using the Bland-Altman plot, which displays the differences between the measurements on the y-axis and the average of measurements on the x-axis. The Deming regression was applied to analyze series with a proportional bias. Additionally, the mountain plot, which shows the differences between the methods on the y-axis and the average of the two methods on the x-axis, was used. Cohen's kappa index was evaluated starting from a 2X2 contingency table. In summary, Cohen's kappa measures the standardized difference between the observed rate of agreement and the rate of agreement that would be expected by chance alone. Cohen's kappa values range from ≤ 0, indicating no agreement, to 0.01–0.20, indicating poor agreement, 0.21–0.40 for fair agreement, 0.41–0.60 for moderate correlation, 0.61–0.80 for good agreement, and 0.81–1.00 for very good correlation (24).

Results

Analytical performance

We analyzed 190 consecutive patients (M = 134) for FT with both direct and calculated method. The last was considered the gold standard. Mean values and range for each method are reported in Table 1. Calculated FT with Vermulen formula is usually expressed in nmol/L. To compare the two methods and apply the Passing-Bablok regression, we transformed the data in pg/ml considering the equation for T: 1 pmol/litre = 0.2885 pg/mL. We estimated the Passing-Bablok for the whole series and for males/females alone (Fig. 1). Considering all patients together we obtained an intercept of 0.19 (CI 95% -0.1722 to 0.4984) and a slope of 1.023 (CI 95% 0.962 to 1.0854). For males, the slope was 1.2150 (CI 95% 1.0978–1.3620); for females, the slope was 0.7389 (CI 95% 0.5434–0.9486). For all comparisons the slopes around 1, or close, suggested that the two methods agreed well, and the intercepts were concordant with the absence of systematic differences between the two methods. Moreover, for both intercept and slope

the CI 95% intervals were narrow enough to indicate that the estimates were reliable. The Bland-Altam plot confirmed for the whole and male series an overall agreement of the two methods as the plots showed a random scatter of points around the mean difference line, with no systematic pattern or trend (Fig. 2). For females, the points were more concentrated around the same position, possibly indicating the presence of a proportional bias. In addition, significance levels for both slope and intercept ($p < 0.001$ and $p < 0.0001$, respectively) suggested that the difference between the two methods was not likely to be due to chance. For all series the mean difference was closed to 0, indicating that the two methods agreed well on average (Fig. 2). From the moment that for female series, the Bland-Altam was suggestive of a possible proportional bias, the Deming regression was performed. This method estimates both the slope and intercept of the line that best describes the relationship between the two methods, considering the measurement error of both methods. By that, we obtained a slope of 1.88 (indicating perfect agreement between the two methods), an intercept of -1.19 and a confidence interval of 0.8262–0.9372 which suggested the reliability of the estimates (Fig. 2B). For all series the Mountain graphs showed a symmetric distribution centred around 0 and narrow, peaked plots were congruent with a good agreement between methods (Fig. 3). A Cohen's k coefficient of 0.61 was found again indicating a good correlation.

Table 1
Characteristics of study population of 190 patients

Males (n = 134)			
	Mean	**Range**	**Reference Intervals**
T (ng/ml)	3.2 ± 1.43	< 0.1–7.52	2.4–7.81
SHBG (nmol/L)	37.44 ± 16.06	6.5–87.9	13–90
Calculated FT (nmol/L)	0.217 ± 0.11	0.006–0.93	> 0.225
Calculated FT (pg/ml)	62.7 ± 31.9	0.87-268.31	> 64.9
Direct FT (pg/ml)	15.55 ± 8.39	0.913–39.9	15–50
Females (n = 56)			
T (ng/ml)	0.51 ± 0.4	< 0.1–2.93	0.1–0.75
SHBG (nmol/L)	53.5 ± 24.28	9.3-102.8	18–135
Calculated FT (nmol/L)	0.028 ± 0.031	0.0035-0.22	< 0.039
Calculated FT (pg/ml)	8.11 ± 9.03	1.154–63.75	< 11.25
Direct FT (pg/ml)	2.63 ± 4.09	1.11-32	< 9.0

Sensitivity and specificity of direct method compared with Vermeulen formula

For direct method, we calculated statistical parameters commonly used to assess the validity of a laboratory test. Again, the calculated method was considered the gold standard. For both males and females, we considered as true positive (TP) FT pathological levels (lower for males and higher in females) being, our, a clinical series. Evaluating all patients together, for the direct method, we obtained a sensitivity of 75.0%, a specificity of 93.4% with a PPV of 89.8%, a NPV of 81.0% and an accuracy of 84.2% (Table 2). From this analysis and being the specificity the ability of a laboratory test to correctly identify individuals who do not have the condition, direct method appears more likely to properly identify true negative patients resulting in fewer false positives. Sensitivity was increased to 81.0% considering male only, while dropped to 18.2% in female probably since in our series we observed only 2/56 (3.6%) true positive patients and 9/56 (16%) false negative patients compared to the gold standard. In male, the specificity was 90.0% with a PPV of 90.9%, a NPV of 79.4% and an accuracy of 85% (Table 2). In female the specificity was 97.8% with a PPV of 66.7%, NPV of 83% and accuracy 82.2%.

Table 2
Statistical parameters of direct method compared with Vermeulen formula (gold standard)

	Whole series (n = 190)	**Male (n = 134)**	**Female (n = 56)**
Sensitivity (%)	75.0	81.0	18.2
Specificity (%)	93.4	90.0	97.8
Positive predictive value (PPV, %)	89.8	90.9	66.7
Negative predictive value (NPV, %)	81.0	79.4	83.0
Accuracy (%)	84.2	85	82.2

Discussion

In clinical practice, measuring testosterone (T) levels in adult men primarily serves the purpose of ruling out hypogonadism, a condition characterized by low T levels and symptoms of androgen deficiency, such as erectile dysfunction, low libido, and reduced nocturnal erections. Over the past two decades, there has been increasing awareness regarding male hypogonadism, a condition that is often underdiagnosed and undertreated due to the similarity of its symptoms with those of other common disorders (2, 21).

Testosterone also plays a crucial role in women, exerting physiological effects directly or through conversion to estradiol in reproductive and non-reproductive tissues. However, in clinical practice, assessing T levels in female patients mainly aims to detect excessive production of testosterone, which can occur in conditions such as congenital adrenal hyperplasia (CAH) and hirsutism/polycystic ovary

syndrome (PCOS) (14, 17). The actions of testosterone are mediated through its unbound form called free testosterone (FT).

According to scientific guidelines (20–21), equilibrium dialysis (ED) is considered the most accurate method to measure FT. In a study by Fiers et al. (25), ED results were compared with the Vermeulen formula, empirical equations according to Ly, and a proposed calculation based on a dynamic, multistep, allosteric model by Zakharov. The study found that ED is more accurate and reliable for measuring FT (25). However, ED is costly, time-consuming, and impractical for routine clinical use.

In order to find other methods to measure FT, aside from ED, radioimmunoassay (RIA) and calculated FT versus equilibrium dialysis were considered by Kacker et al. in a clinical series (26). They found a strong correlation for RIA and ED ($r = 0.966$) and for calculated FT and ED ($r = 0.986$). However, RIA is expensive and not feasible in all laboratory facilities. Generally, calculated methods are the easiest to implement in clinical practice because they can be performed in all laboratories. However, calculated FT values are based on the assumption of a normal steady-state protein binding for testosterone and depend on the reliability and calibration of the testosterone and SHBG assays (27–29). These calculated methods may fail in the presence of high concentrations of competing steroids, significant deviations from normal protein concentrations, or in several clinical conditions that can influence SHBG concentrations (such as drugs, some endocrine disorders, cirrhosis and hepatitis, aging, etc). Among the calculated methods, the Vermeulen formula (19) was found to be the most accurate, despite slightly overestimating FT values as it appears to be influenced by SHBG, albumin, and testosterone levels to a lesser extent (25).

In clinical biochemistry, the error of a measure refers to the degree of inaccuracy or imprecision in the results obtained from laboratory tests or assays. No measurement is entirely free from error; that means that calculated FT may increase the probability of having a non-reliable value due to errors in T and SHBG measurements. On the contrary, an immunoluminometric assay using specific antibodies to target the analyte of interest ensures high specificity, minimizes the risk of cross-reactivity with other substances, and can detect very low analyte concentrations. While calculated methods may be more prone to variability, immunoluminometric assays exhibit excellent reproducibility and can be easily implemented in all automated laboratory platforms with lower costs.

In this article, we compared FT levels obtained by a direct immunoluminometric method or with the Vermeulen formula in a clinical series. The assessment of the clinical performance of the two methods indicates approximately the same variability of the dosed FT levels compared to the calculated one. Passing-Bablok regression analysis showed a good agreement between the two methods, with slopes close to 1 and narrow confidence intervals. The Bland-Altman plot demonstrated overall agreement, but a possible proportional bias was observed in females. Thus, Deming regression was applied, and it confirmed excellent agreement with a slope of 1.88 and reliable estimates. Sensitivity and specificity analysis showed that the direct method had a sensitivity of 75.0% and specificity of 93.4% when all patients were considered. Sensitivity improved to 81.0% in males but dropped to 18.2% in females, likely due to the low number of true positive cases.

In conclusion, the direct method exhibited comparable performance to the Vermeulen equation, but it doesn't take into account SHBG measurement, which could be important for a comprehensive assessment of hormonal status and patient management, although susceptible to variation. Further studies are needed, in larger series, aimed to compare the immunoluminometric assay with ED, the actual gold standard for FT evaluation, to validate sensitivity, specificity, and clinical utility of the direct method.

Declarations

Conflict of Interest: The authors declare no conflict of interest

Ethical Approval: The study was conducted in accordance with the Declaration of Helsinki

Informed Consent: The FT evaluation was part of patient diagnostic routine for which patients have signed an informed consent

References

1. Kelly DM, Jones TH (2013) Testosterone: a metabolic hormone in health and disease. J Endocrinol 217(3):R25–45
2. Rastrelli G, Corona G, Maggi M (2018) Testosterone and sexual function in men. Maturitas 112:46–52
3. McHenry J, Carrier N, Hull E, Kabbaj M (2014) Sex differences in anxiety and depression: role of testosterone. Front Neuroendocrinol 35(1):42–57
4. Sharma A, Minhas S, Dhillo WS, Jayasena CN (2021) Male infertility due to testicular disorders. J Clin Endocrinol Metab 106(2):e442–e459
5. Oduwole OO, Huhtaniemi IT, Misrahi M (2021) The Roles of Luteinizing Hormone, Follicle-Stimulating Hormone and Testosterone in Spermatogenesis and Folliculogenesis Revisited. Int J Mol Sci 22(23):12735
6. Rajfer J (2000) Relationship between testosterone and erectile dysfunction. Rev Urol 2(2):122–128
7. Kloner RA, Carson C 3rd, Dobs A, Kopecky S, Mohler ER 3 (2016) rd. Testosterone and Cardiovascular Disease. J Am Coll Cardiol 67(5):545–557
8. Bhasin S, Valderrábano RJ, Gagliano-Jucá T et al (2022) Age-Related Changes in the Male Reproductive System. In: Feingold KR, Anawalt B, Blackman MR, eds. Endotext. South Dartmouth (MA): MDText.com, Inc.; February 10,
9. Linderman JK, O'Hara R, Ordway J (2020) Effect of Special Operations Training on Testosterone, Lean Body Mass, and Strength and the Potential for Therapeutic Testosterone Replacement: A Review of the Literature. J Spec Oper Med 20(1):94–100
10. Khodamoradi K, Khosravizadeh Z, Seetharam D, Mallepalli S, Farber N, Arora H (2022) The role of leptin and low testosterone in obesity. Int J Impot Res 34(7):704–713

11. Corona G, Vena W, Pizzocaro A, Giagulli VA, Francomano D, Rastrelli G, Mazziotti G, Aversa A, Isidori AM, Pivonello R, Vignozzi L, Mannucci E, Maggi M, Ferlin A (2022) Testosterone supplementation and bone parameters: a systematic review and meta-analysis study. J Endocrinol Invest 45(5):911–926
12. Corona G, Guaraldi F, Rastrelli G, Sforza A, Maggi M (2021) Testosterone Deficiency and Risk of Cognitive Disorders in Aging Males. World J Mens Health 39(1):9–18
13. Pintana H, Chattipakorn N, Chattipakorn S (2015) Testosterone deficiency, insulin-resistant obesity and cognitive function. Metab Brain Dis 30(4):853–876
14. Davis SR, Wahlin-Jacobsen S (2015) Testosterone in women–the clinical significance. Lancet Diabetes Endocrinol 3(12):980–992
15. Smith T, Batur P (2021) Prescribing testosterone and DHEA: The role of androgens in women. Cleve Clin J Med 88(1):35–43
16. Le J, Thomas N, Gurvich C, Cognition (2020) The Menstrual Cycle, and Premenstrual Disorders: A Review. Brain Sci 10(4):198
17. Stanczyk FZ (2006) Diagnosis of hyperandrogenism: biochemical criteria. Best Pract Res Clin Endocrinol Metab 20(2):177–191
18. Goldman AL, Bhasin S, Wu FCW, Krishna M, Matsumoto AM, Jasuja R (2017) A Reappraisal of Testosterone's Binding in Circulation: Physiological and Clinical Implications. Endocr Rev 38(4):302–324
19. Vermeulen A (1973) Testosterone in plasma. A physiopathological study. Verhandelingen – Koninklijke Academie voor Geneeskunde van Belgie 35:95–180
20. Bhasin S, Brito JP, Cunningham GR, Hayes FJ, Hodis HN, Matsumoto AM, Snyder PJ, Swerdloff RS, Wu FC, Yialamas MA (2018) Testosterone Therapy in Men With Hypogonadism: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 103(5):1715–1744
21. Isidori AM, Aversa A, Calogero A, Ferlin A, Francavilla S, Lanfranco F, Pivonello R, Rochira V, Corona G, Maggi M (2022) Adult- and late-onset male hypogonadism: the clinical practice guidelines of the Italian Society of Andrology and Sexual Medicine (SIAMS) and the Italian Society of Endocrinology (SIE). J Endocrinol Invest 45(12):2385–2403
22. Vermeulen A (2005) Hormonal cut-offs of partial androgen deficiency: a survey of androgen assays. J Endocrinol Invest 28(3 Suppl):28–31
23. Plebani M (2006) Errors in clinical laboratories or errors in laboratory medicine? Clin Chem Lab Med 44(6):750–759
24. Casagrande A, Fabris F, Girometti R (2020) Beyond kappa: an informational index for diagnostic agreement in dichotomous and multivalue ordered-categorical ratings. Med Biol Eng Comput 58(12):3089–3099
25. Fiers T, Wu F, Moghetti P, Vanderschueren D, Lapauw B, Kaufman JM (2018) Reassessing Free-Testosterone Calculation by Liquid Chromatography-Tandem Mass Spectrometry Direct Equilibrium Dialysis. J Clin Endocrinol Metab 103(6):2167–2174

26. Kacker R, Hornstein A, Morgentaler A (2013) Free testosterone by direct and calculated measurement versus equilibrium dialysis in a clinical population. Aging Male 16(4):164–168
27. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H (2007) Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. J Clin Endocrinol Metab 92(2):405–413
28. Matsumoto AM, Bremner WJ (2004) Serum testosterone assays–accuracy matters. J Clin Endocrinol Metab 89(2):520–524
29. Tosi F, Fiers T, Kaufman JM, Dall'Alda M, Moretta R, Giagulli VA, Bonora E, Moghetti P (2016) Implications of Androgen Assay Accuracy in the Phenotyping of Women With Polycystic Ovary Syndrome. J Clin Endocrinol Metab 101(2):610–618

Figures

Fig. 1

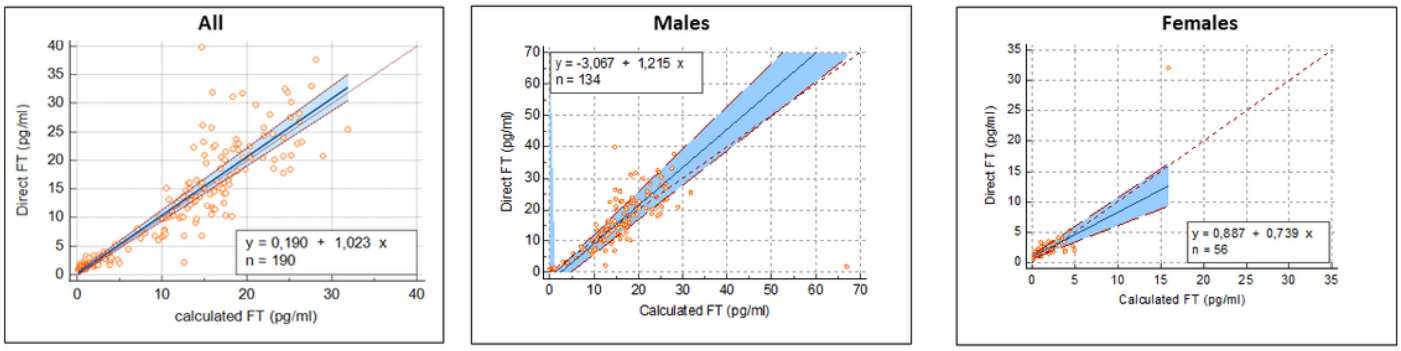

Figure 1

Legend not included with this version.

Fig. 2

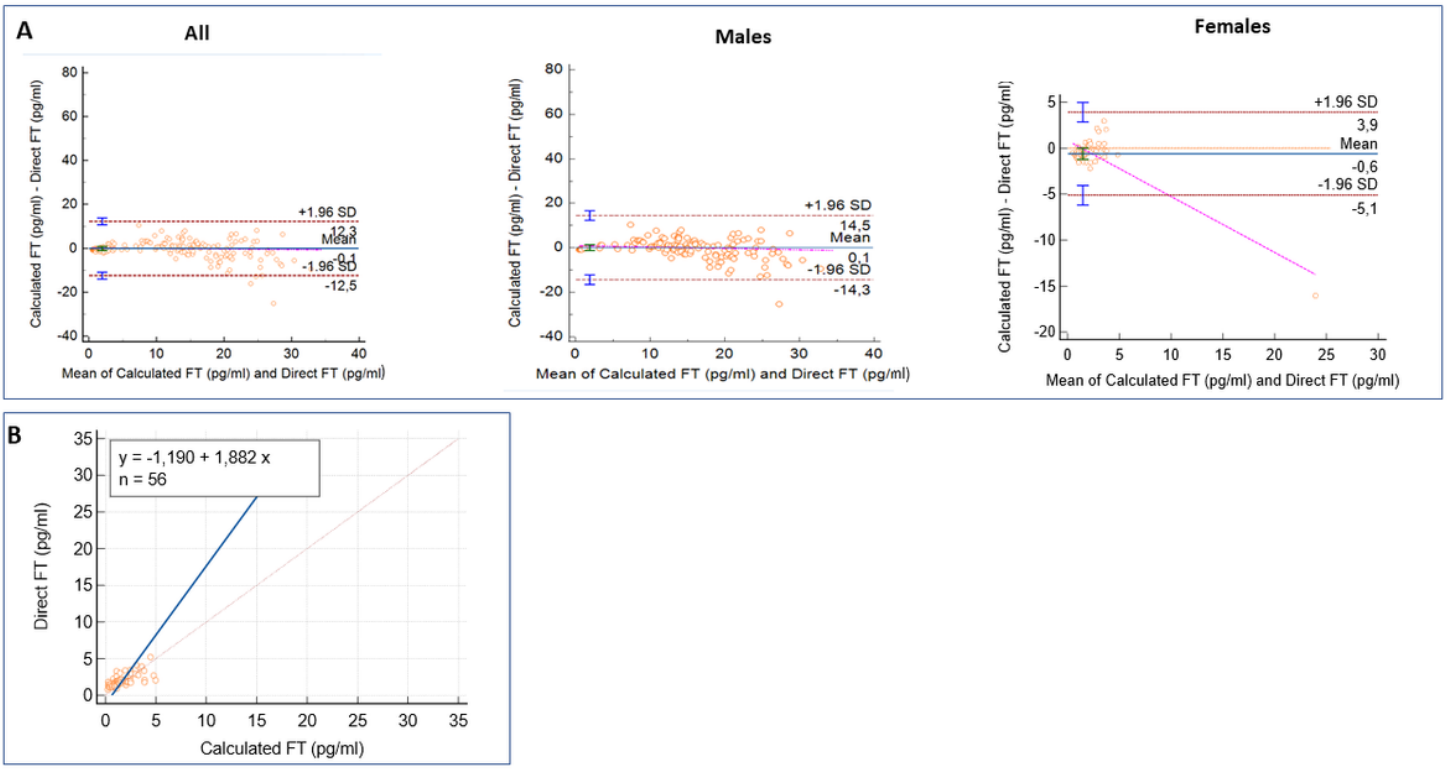

Figure 2

Legend not included with this version.

Fig. 3

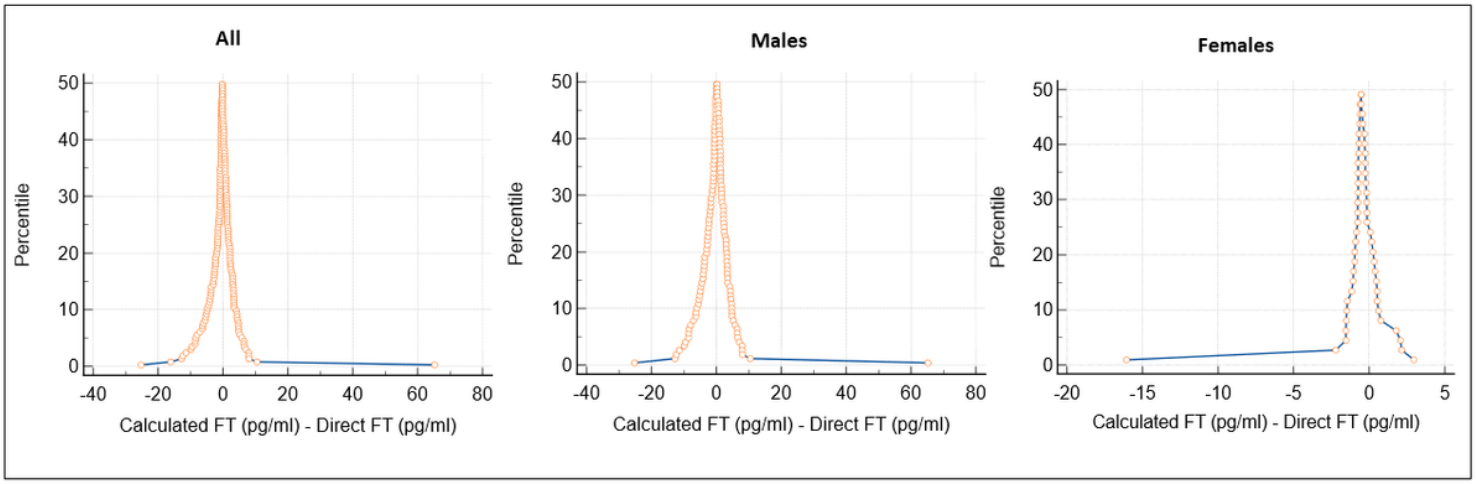

Figure 3

Legend not included with this version.