



Published in final edited form as:

*Anal Bioanal Chem.* 2017 October ; 409(25): 5943–5954. doi:10.1007/s00216-017-0529-x.

## Simultaneous measurement of total Estradiol and Testosterone in human serum by isotope dilution liquid chromatography tandem mass spectrometry

Hui Zhou<sup>1</sup>, Yuesong Wang<sup>1</sup>, Matthew Gatcombe<sup>1</sup>, Jacob Farris<sup>1</sup>, Julianne C. Botelho<sup>1</sup>, Samuel P. Caudill<sup>1</sup>, and Hubert W. Vesper<sup>1,\*</sup>

<sup>1</sup>Centers For Disease Control And Prevention, National Center For Environmental Health, Division of Laboratory Sciences, Clinical Chemistry Branch, 4770 Buford Hwy NE, Atlanta, Georgia, USA

### Abstract

Reliable measurement of total testosterone and estradiol is critical for their use as biomarkers of hormone related disorders in patient care and translation research. We developed and validated a mass spectrometry method to simultaneously quantify these analytes in human serum without chemical derivatization. Serum is equilibrated with isotopic internal standards and treated with acidic buffer to release hormones from their binding proteins. Lipids are isolated and polar impurities are removed by two serial liquid-liquid extraction steps. Total testosterone and estradiol are measured using liquid chromatography tandem mass spectrometry (LC-MS/MS) in combination of positive and negative electrospray ionization modes. The method shows broad analytical measurement range for both testosterone 0.03–48.5 nM (0.75–1400 ng/dL) and estradiol 11.0–5138 pM (2.99–1400 pg/mL) and excellent agreement with certified reference materials (mean bias less than 2.1% to SRM 971, BCR 576, 577, and 578) and a high order reference method (mean bias 1.25% for testosterone and –0.84% for estradiol). The high accuracy of the method was monitored and certified by CDC Hormone Standardization (HoSt) Program for two years with mean bias –0.7% (95% CI: –1.6% to 0.2%) for testosterone and 0.1% (95% CI: –2.2% to 2.3%) for estradiol. The method precision over a 2-year period for Quality Control pools at low, medium and high concentrations was 2.7–2.9% for testosterone and 3.3–5.3% for estradiol. With the consistently excellent accuracy and precision, this method is readily applicable for high-throughput clinical and epidemiological studies.

### Keywords

CDC HoSt; Estradiol; Testosterone; Hormone; Serum; LC-MS/MS

\*Send correspondence to this author: Hubert W Vesper, Centers for Disease Control and Prevention, 4770 Buford HWY NE, MS-F25, Atlanta, Georgia 30341, Fax (770) 488-7030, HVesper@cdc.gov.

Disclaimer: The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention

### Conflict of interest

The authors declare that they have no conflict of interest.

## 1. Introduction

17 $\beta$ -Estradiol (E2) and testosterone (TT) are sex hormones responsible for the development and maintenance of secondary sex characteristics and reproductive functions (1). They influence many physiological processes, such as growth, glucose homeostasis, bone and lipid metabolism, and cardiovascular function (2–5). Measurement of total circulating E2 or TT in conjunction with other biomarkers and clinical assessments is widely used for evaluation of reproductive status and function, for diagnosis of disorders such as infertility, delayed or precocious puberty, and other diseases related to altered steroid hormone metabolism. TT and E2 are also critical in monitoring therapeutic interventions such as in-vitro fertilization and antiestrogen therapy (6–11). Reliable and accurate measurements of these biomarkers are critical for correct diagnosis, treatment and prevention of diseases, clinical research, and public health activities.

Different technologies for measuring circulating E2 and TT are available such as radioimmunoassays (RIA), enzyme-linked immunoassays (EIA), and mass spectrometry methods (12, 13). Most of the analytical methods used in patient care and public health activities only measure either testosterone or estradiol (14–17), thus requiring separate assays for each analyte. Some of these assays were found to have a high level of inaccuracy, lack of sensitivity, and problems with reproducibility. These problems were reported especially at the low concentrations commonly observed for estradiol in men and postmenopausal women (approximately below 147 pM, 40 pg/ml), and testosterone in women, children and hypogonadal men (approximately less than 3.47 nM, 100 ng/dL) (18–20). Thus, new analytical methods are needed to accurately and reliably measure E2 and TT in serum.

Here we report an accurate and sensitive LC-MS/MS assay for simultaneous measurement of E2 and TT in serum without chemical derivatization. The sample preparation is automated using a liquid handling system with 96-well plates, enabling the maximal capacity of processing 146 patient samples within an 8-hour workday. Our validated method has been demonstrated to be applicable to measurements in the general population.

## 2. Material and Methods

### 2.1. Materials and Chemicals

Certified reference materials in acetonitrile 1 mg/ml 17 $\beta$ -Estradiol (purity 98.4%, National Metrology Institute of Japan, NMIJ CRM 6004-a, Japan) and testosterone (purity 99.4%, Australian National Measurement Institute, ANMI M914, Australian) were used as calibrators, and [2,3,4-<sup>13</sup>C<sub>3</sub>]-17 $\beta$ -Estradiol (<sup>13</sup>C<sub>3</sub>-E2, 98%) and [2,3,4-<sup>13</sup>C<sub>3</sub>]-Testosterone (<sup>13</sup>C<sub>3</sub>-TT, 98%) obtained from IsoSciences (King of Prussia, PA) were used as internal standards (IS). Ammonium acetate, ammonium fluoride, ethyl acetate, glacial acetic acid, hexane, methanol, sodium chloride and water were purchased from Fisher Scientific (Suwanee, GA) and ammonium bicarbonate, ammonium hydroxide and ethanol from Sigma-Aldrich (St. Louis, MO). All solvents were HPLC grade and chemicals were reagent grade. Steroids used for interference testing (Table 1) were obtained from Steraloids (Newport, RI), Sigma-Aldrich, and Cerrillant (Round Rock, TX). They were prepared in

20% methanol in water at concentrations of 200 ng/dL per steroid for assessing interferences with TT and 200 pg/mL per steroid for assessing interferences with E2.

Three levels of serum-based reference materials for E2 (BCR 576, BCR 577, and BCR 578 at 114, 690, 1340 pM (31.1, 188, and 365 pg/mL), respectively) were purchased from the Institute for Reference Materials and Measurements (IRMM, Geel, Belgium). Two levels of serum-based reference materials for TT (SRM 971 at 22.3 nM and 0.96 nM (643 ng/dL and 27.7 ng/dL) were obtained from the National Institute for Standards and technology (NIST, Gaithersburg, MD). The certified reference materials were used to assess the trueness of the method. Quality control pools at three different E2 and TT levels were created by combining human sera and were used for assessing precision and extraction efficiency of the method. The concentrations for testosterone were 1.27, 6.07, and 30.2 nM (36.6, 175, and 870 ng/dL), and for estradiol 103, 448, and 2540 pM (28.0, 122, and 692 pg/mL), for the low, medium and high QC level, respectively. Synthetic serum for matrix effect evaluations were procured by UTAK Laboratories (Valencia, CA). Serum samples from individual donors were used to evaluate the analytical method. Fresh frozen, single donor human serum samples were purchased from Bioreclamation (Westbury, NY) and Solomon Park Research Laboratories (Kirkland, WA). The company had institutional review board (IRB) approval to collect blood and obtained informed content from donors. CDC's use of the blood was consistent with the IRB approval and donor consent. No identifiers were provided to CDC.

## 2.2. Calibrator Preparation

The primary stock solutions for testosterone and estradiol were prepared respectively from certified, commercial solutions with an assigned concentration of 1 mg/mL by diluting them with anhydrous ethanol (TT Stock Solution A, E2 Stock Solution A) to a concentration of 1 µg/mL. Stock B solution was created by diluting TT Stock A solution and E2 Stock A solution with 20% ethanol to a concentration of 69.3 nM (20 ng/mL) of TT and 7.34 nM (2 ng/mL) of E2. To prepare 11-point calibrator working solutions, 11 different volumes of Stock B solutions were added in 200 mL 20% ethanol to achieve the target concentrations at 0.1, 1, 4, 10, 25, 50, 100, 250, 500, 750, and 1000 for TT (ng/dL, 0.003–48.5 nM) and E2 (pg/mL, 0.37–3670 pM). The working solutions were aliquoted in cryovials and stored at -70°C. One set of calibrators was used with serum samples and processed in the same manner. A constant concentration of internal standard mixture of  $^{13}\text{C}_3$ -TT and  $^{13}\text{C}_3$ -E2 containing 3.43 nM (100 ng/dL)  $^{13}\text{C}_3$ -TT and 182 pM (50 pg/mL)  $^{13}\text{C}_3$ -E2 was added to each standard solution to establish the calibration curves during sample preparation.

## 2.3. Sample preparation

Sample preparation was conducted using an automated Hamilton Microlab STARLet Liquid Handler (Reno, NV) in 2 mL 96-well plates. Each sample batch contained three levels of quality control (QC) pools, calibrator solutions (0.003 to 34.7 nM (0.1 to 1000 ng/dL) for testosterone and 0.37 to 3670 pM (0.1 to 1000 pg/mL) for estradiol), and a reagent blank. The sample preparation procedure was based on a previously described method (21) and was modified for simultaneous measurement of TT and E2. In brief, 100 µL of IS solution (0.25% ethanol in water containing 182 pM (50 pg/mL)  $^{13}\text{C}_3$ -E2 and 3.43 nM (100 ng/dL)  $^{13}\text{C}_3$ -TT) was added to 200 µL of sample and mixed for 45 min at room temperature.

100  $\mu\text{L}$  of 0.5 M ammonium acetate (pH 5.5) was added and mixed for 30 min to dissociate the steroid hormones from their binding proteins. Steroids were extracted with 600  $\mu\text{L}$  of extraction solution (ethyl acetate/hexane, 40/60 v/v). After centrifugation, the top 500  $\mu\text{L}$  of organic layer was transferred into a well plate containing 200  $\mu\text{L}$  of buffer (0.2 M ammonium bicarbonate, pH 8.0) and mixed. The organic layer was transferred into a third well plate and the previous extraction and deprotonation steps were repeated a second time. The combined organic layers were dried under vacuum. The sample extracts were then reconstituted in 135  $\mu\text{L}$  of methanol/water (20/80 v/v) containing 0.2 mM ammonium fluoride and analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS).

#### 2.4. LC-MS/MS

Chromatographic separation was performed using an UPLC system (Shimadzu LC-10 AD VP HPLC system, Columbia, MD) equipped with a Phenyl/Hexyl Column (2.6  $\mu\text{m}$ , 150  $\times$  3.0 mm, Accucore Thermo Scientific, Waltham, MA) at 40°C. The eluents were 20% methanol in water with 0.2 mM ammonium fluoride (eluent A) and methanol (eluent B). 50  $\mu\text{L}$  of sample solution was injected and the analytes were eluted using the gradient from 40% to 72.5% of eluent B at the flow rate of 450  $\mu\text{L}/\text{min}$  with a total of run time of 10 minutes. Mass spectrometric analysis was performed using a triple quadrupole tandem mass spectrometer (AB Sciex 5500, Foster City, CA) with electrospray ionization switching in negative ion mode for E2 and positive mode for TT. Selected reaction monitoring (SRM) was used with quantitation ion transitions (QI) of  $m/z$  271  $\rightarrow$  145 for E2, 274  $\rightarrow$  148 for  $^{13}\text{C}_3$ -E2, 289  $\rightarrow$  97 for TT, and 292  $\rightarrow$  100 for  $^{13}\text{C}_3$ -TT. For monitoring confirmation ion transitions (CI),  $m/z$  271  $\rightarrow$  183 for E2, 274  $\rightarrow$  186 for  $^{13}\text{C}_3$ -E2, 289  $\rightarrow$  109 for TT, and 292  $\rightarrow$  112 for  $^{13}\text{C}_3$ -TT were used. The declustering potential (DP), entrance potential (EP) and collision cell exit potential (CXP) were -140 V, -12 V, -17 V for E2 and 240 V, 8 V, 12 V for TT, respectively. The collision energy (CE) was -51 eV for all E2 and  $^{13}\text{C}_3$ -E2 transitions and 27 eV for all TT and  $^{13}\text{C}_3$ -TT transitions.

#### 2.5. Data Analysis

Chromatographic peaks for QI and CI transitions were integrated using Analyst software version 1.6.2 (AB Sciex). To establish the best fit for the calibration curve, the peak area ratios of analytes and ISs were obtained from six sets of calibration runs and plotted with target concentrations using linear and polynomial models with no weighting, weights of  $1/X$ ,  $1/X^2$ , or  $1/(\text{Variance of } Y)$ . The average sum of squared residuals (ASSR) and the average relative sum of squared residuals (RASSR) were estimated by comparing the calculated concentrations to the target concentrations. A weight of  $1/X$  was selected for the lowest ASSR and second lowest RASSR among all models (21). The analyte concentration in serum was calculated using the peak area ratio for the unknown sample and the regression parameters of the established  $1/X$  weighted calibration curve. The QI/CI ratios from calibrators were calculated to determine the target QI/CI ratio and limits of acceptability in samples. SAS version 9.2 (SAS Institute, Inc, Cary, NC) was used to define quality control limits and to evaluate analytical runs against these limits using a multiple-rule quality control approach (22).

## 2.6 Method Validation

This method was validated as outlined in CLSI document C62-A (23). Measurement accuracy was assessed by analyzing serum-based certified reference materials from NIST and IRMM in replicate ( $n = 5$  per level). Agreement of the measured value with the assigned target value of the reference materials was assessed as described in NIST Special Publication 829 (24). In addition, measurement bias was evaluated following CLSI document EP9-A2 (25) using 40 single-donor serum materials with reference values assigned by higher-order reference methods listed in the Joint Committee for Traceability in Laboratory Medicine (JCTLM) database (NIST (26), University of Ghent (27, 28), and CDC Reference Laboratory (29, 30). The accuracy of the method was monitored over 2 years by measuring 10 blinded serum samples quarterly for each analyte with reference values assigned by higher-order reference methods. The measurement accuracy was compared against suggested performance criteria for TT (mean bias for TT: 6.4% (31), and for E2: 12.5% at concentrations  $> 73.4$  pM (20 pg/mL), and  $\pm 9.2$  pM (2.5 pg/mL) at concentrations  $< 73.4$  pM (20 pg/mL) (32). Deming regression analysis and difference plots were performed using Analyze-It Software, Ltd Version 4.65.2 (Leeds, UK).

Method intermediate precision was determined following CLSI document EP5-A3 (33) by analysis of three levels of QC pools in 68 different days (68 days  $\times$  2 replicates per run  $\times$  2 two runs per day) over a period of 2 years using multiple operators. The following three imprecision parameters were determined: the within-run imprecision (repeatability), the within-day and among-day imprecision, and the total imprecision, which is the within-laboratory imprecision. The data was analyzed using SAS/STAT software, version 9.2 (SAS Institute, Inc, Cary, NC). The precision was expressed as percent coefficient of variation (%CV).

Analytical specificity was assessed by testing 32 structural steroid analogs and other steroid hormones with relative molecular masses similar to TT, E2 or ISs (Table 1) for potential chromatographic co-elution with TT, E2, and corresponding ISs. The absence of a peak with mass transitions (289 $\rightarrow$ 97 for TT and 271 $\rightarrow$ 145 for E2; 292 $\rightarrow$ 100 for  $^{13}\text{C}_3$ -TT and 274 $\rightarrow$ 148 for  $^{13}\text{C}_3$ -E2) at the respective retention times for TT and its IS (7.93 min) and E2 and its IS (7.28 min) confirmed absence of interference with the quantification of total testosterone and estradiol in serum. Additionally, specificity was assessed by monitoring the QI/CI ratios in regular serum samples. A difference of more than  $\pm 20\%$  from the established target ratio was used to indicate the presence of an interfering compound (21, 23, 29, 34).

The limit of detection (LOD) was determined by analyzing 5 samples with concentrations close to the estimated LOD for each analyte for 60 days. This approach is equivalent with CLSI EP 17-A2 and includes Type II error (false negative) in estimation of LOD (35, 36).

The linear range of the method was determined using a 13-point calibration curve (E2 0.37–5138 pM (0.1–1400 pg/mL), and TT 0.003–48.5 nM (0.1–1400 ng/dL)) measured in six replicates. The applicability of sample dilution was assessed for samples with analyte concentrations above the upper limit of the linear range. Sample dilution experiments were carried out using high QC pool materials that were diluted with saline using dilution factors ranging from 1 to 10 (total of 9-level dilutions). Duplicate preparations of each diluted

sample were made. Dilution factors were applied to calculate the final concentrations and the calculated results were compared to the values obtained from undiluted samples.

Sample matrix effects (ME) were assessed on four matrices (0.9% saline solution, synthetic serum, male serum, and female serum) and one set of neat samples in ethanol (matrix-free) using previously described procedures (29). An 8-point calibration curve ranging from 33.0–3670 pM (9–1000 pg/mL) for E2 and 0.31–34.7 nM (9–1000 ng/dL) for TT were prepared in each matrix. The calibrators in the matrices were subjected to the sample preparation described above. The area count ratios of analyte over IS were compared in all four matrices to those analyzed in matrix-free condition after blank subtraction. The sample ME was calculated with the following equation:  $ME \% = B/A \times 100$ , where B is the area count ratios of analyte to IS obtained from samples in matrix, and A is the area count ratios in matrix-free samples.

The extraction efficiency was evaluated using low, medium and high QC pools (n = 4). In one set of QC pools, the IS solution was added at the beginning of sample preparation (A), and in a second set of QC pools it was added at the end of the sample preparation (B). The efficiency was calculated with the following equation: measured concentration (B)/measured concentration (A) x 100.

To assess the applicability of this method to samples in the general population, a set of 250 individual sera from female and male subjects were analyzed using the described method.

### 3.0 Results and discussion

The described method is highly accurate. The mean biases for TT to SRM 971 were –1.5% and –2.1%; for E2 to BCR 576, 577 and 578 were –1.3%, –1.2%, and –1.7%, respectively. None of these biases were statistically significant as determined according to NIST Special Publication 829 (Table 2). Comparison of our method to the established reference methods with 40 patient samples covering a concentration range from 0.20 to 28.0 nM (5.70 to 808 ng/dL) for TT and 24.2 to 1046 pM (6.60 to 285 pg/mL) for E2 showed no significant difference using Deming regression analysis (Fig 1a, TT: slope 1.013 [95%CI: 1.000–1.026], intercept: –0.0009 [95%CI: –0.013–0.011], E2: slope 1.011, 95%CI: 0.966–1.054, intercept: –1.692, 95%CI: –6.324–2.940) as well as bias plot analysis (Fig 1b, mean bias 1.25% [95%CI: 0.36–2.14%] for TT and –0.84% [95%CI: –2.67–0.99%] for E2). The good agreement with the reference materials and methods is also reflected in the narrow confidence intervals observed in each experiment. Regression analysis of the mean bias from 10 samples for each analyte measured over 8 quarters (total of 80 samples) showed a slope not significantly different from zero, indicating that the method is highly accurate over 2 years (Fig 1c). The mean bias across all eight quarters was –0.7% (95%CI: –1.6% to 0.2%) for TT and 0.1% (95%CI: –2.2% to 2.3%) for E2. The mean bias observed with our method is well below the suggested bias limits for TT (31), and for E2 at concentrations above 20 pg/mL (32). Individual measurements (n=4) bias of all 80 samples for each analyte ranged between 6.6–9.1% (95%CI) for TT and 15.0–21.6% (95%CI) for E2, and met the total error criteria for TT of 16.7% and for E2 of 27.0% (37).

The developed analytical method is highly specific. Using the described chromatographic conditions, we observed that TT and E2 are well separated from potentially interfering compounds such as those listed in Table 1 (Fig 2). In addition, no isotopic interferences were observed when monitoring the IS and analyte transitions. Other potential interferences can be detected using the QI/CI ratio. The mean QI/CI ratios calculated using data obtained from calibrators measured in 2 sample batches (n=20) were 1.41 (95%CI: 1.36–1.46) for TT and 1.12 (95%CI: 1.07–1.16) for E2. The mean QI/CI in 134 serum samples for TT was 1.42 (range: 1.41–1.44) and for E2 was 1.12 (range: 1.08–1.16). The QI/CI ratios of all 134 serum samples were almost identical to those measured in the calibrators, and were well below the recommended maximum difference of  $\pm 20\%$  (23), suggesting that no interfering compounds are present in the 134 serum samples measured.

Our method is highly precise over long time periods. The within-run (repeatability), among-day (68 days), and total within-laboratory CV ranged between 2.4–5.0%, 1.2–1.8%, and 2.7–5.3%, respectively (Table 3). The precision observed over 2 years is well below the recommended maximum imprecision for total testosterone measurements of 5.3% CV (31) and estradiol measurements of 11.3% CV (38), and smaller than the precision reported by other mass spectrometry-based methods (32, 39).

The method is very sensitive with a wide linearity and analytical measurement range. The LOD of the method is 11.0 pM (2.99 pg/mL) and 0.03 nM (0.75 ng/dL), for E2 and TT respectively. These LODs are similar to those reported for methods using derivatization procedures (40, 41), and lower than some reported for methods without derivatization using higher sample volume (42, 43). The method was found to be linear within the ranges of 11.0–5138 pM (2.99–1400 pg/mL) for E2 ( $r^2 > 0.999$ ) and 0.03–48.5 nM (0.75–1400 ng/dL) for TT ( $r^2 > 0.999$ ). No significant polynomial term was detected. The weighted regression parameters from six replicate calibration curves are very consistent (regression parameters for E2: mean slope 0.009, 95%CI: 0.009–0.010, mean intercept: 0.004, 95%CI: 0.003–0.005; regression parameters for TT: mean slope 0.031, 95%CI: 0.030–0.031, mean intercept: 0.010, 95%CI: 0.007–0.013). Dilutions of samples up to 10-fold with saline were found to result in accurate measurements (accuracy of diluted samples: 100% (95%CI: 98.0–101.7%) for TT and 98.9% (95%CI: 96.4–101.7%) for E2, Table 4) allowing for an extended analytical measurement range of 11.0–51380 pM (2.99–14,000 pg/mL) for E2, and 0.03–485.4 nM (0.75–14,000 ng/dL) for TT. This allows for measurements of TT and E2 in both pre- and post-menopausal women, the measurement of the mid-cycle peak in women without sample dilution, and in pregnant women and women on ovarian stimulation treatment with dilution.

The method is minimally affected by different matrices. Consistent with our observations on diluted samples, the mean ME% determined in 5 different matrices is 98.4% (95%CI: 96.2–100.5%) for TT and 98.5% (95%CI: 96.8–100.1%) for E2. The matrices studied included ethanol, saline, synthetic serum, male serum and female serum, in which all studied curves gave the identical slopes.

The high accuracy, precision, specificity and matrix-independence of measurement results are achieved through the use of stable isotope-labeled standards in combination with special

sample handling procedures and good chromatographic separation.  $^{13}\text{C}$ -labeled internal standards showed the same chromatographic properties as the corresponding analytes (Fig 2). The IS compensate for losses during sample preparation. Several steps in the sample handling process contributed to the observed excellent analytical performance. Protein precipitation, commonly performed in other methods, with incomplete equilibrium of the IS with the binding proteins such as sex hormone binding globulin, and without complete dissociation of analytes from binding protein may result in incomplete or inconsistent analyte recovery (30, 44). The IS solution used in our method contains minimum amounts of ethanol, in order to avoid unnecessary protein precipitation that can lead to loss of protein-bound analytes prior to the equilibration with the IS. Mixing of the serum sample with the IS solution to allow for adequate equilibration and avoiding protein precipitation step may contribute further to the excellent within-run and between-run precision (44).

Introducing the second LLE markedly improved precision, especially at low E2 concentrations, consistency of chromatographic separation, and prolonged column lifetime. During the sample preparation, the lipid extract from the first liquid-liquid extraction (LLE) contains many nonpolar compounds, such as fatty acids and phospholipids. These lipids can accumulate on the analytical columns, deteriorating the separation and providing a source of ion suppression and matrix effects in LC-MS/MS analyses (21). Performing a second LLE using a basic aqueous solution (0.2 M ammonium bicarbonate buffer, pH 8.0) facilitates removal of polar compounds. Though a higher pH would increase the removal of these polar compounds (21), the pH of 8.0 was selected to be 2 pH units below the pK<sub>A</sub> of E2 (10.5). This avoids deprotonation and subsequent decrease in recovery of estradiol (29, 45). Repeating the entire two-step LLE procedure yielded in a higher extraction efficiency of 70% for E2 (95% CI: 67–72%) and 72% for TT (95% CI: 70–75%), compared to 43% for E2 (95% CI: 40–46%) and 44% for TT (95% CI: 43–46%) with only one extraction.

The use of a Phenyl-Hexyl column, which offers unique selectivity for aromatic analytes, instead of a C18 column, contributed to the high specificity of the chromatographic separation. This observation is consistent with another report (29).

The liquid chromatography conditions and MS/MS detection parameters were optimized to achieve the highest sensitivity for estradiol. Because of its non-polar structure and low proton affinity of estradiol, it is difficult to achieve the required level of sensitivity in positive electrospray ionization (ESI). For better yield, negative ESI was chosen for estradiol ion transitions and positive ionization for testosterone. This negative/positive polarity ESI switching enables the highly sensitive analysis of both analytes within a single MS assay. In addition, a mobile phase additive of ammonium fluoride was used to further improve E2 negative ionization efficiency as described previously [24, 46]. Several LC-MS/MS methods for E2 or steroid profiles use dansyl chloride as a derivatizing agent to enhance the ionization and thus to achieve higher sensitivity (46–49). However, the dansyl fragment ion is used to quantitate E2 with these methods, therefore limiting the specificity of the method. This limitation was overcome in our method by measuring E2 without derivatization.

In a single specimen, the concentration of testosterone could be ten to hundred-fold higher than that of estradiol. The challenge of quantifying both compounds in a single assay is to



avoid detector saturation without loss of sensitivity. TT signals were detuned by using an offset DP value of 240 eV instead of the optimal DP value of 90 eV, which was intended to minimize ionization and detector saturation.

Our method can measure E2 and TT in samples from children, women and men. We measured the total testosterone and estradiol in 250 individual sera from male and female donors (Table 5). In adult males age 18 years and older, serum TT concentrations ranged from 0.62 to 32.5 nM (18.0 ng/dL to 938 ng/dL), and E2 concentrations ranged from 17.3 to 167 pM (4.72–45.6 pg/mL). In women age 18 to 59 years, TT concentrations ranged from 0.28 to 2.55 nM (8.03 to 73.6 ng/dL), and E2 concentrations ranged from 11.4 to 12357 pM (3.11 to 3367 pg/mL). Among women age 60 years and older, TT concentrations ranged from 0.26 to 4.13 nM (7.54 to 119 ng/dL), and E2 levels ranged from 9.25 to 1152 pM (2.52 to 314 pg/mL). In children at 6–11 years of age, 90% of boys had E2 levels below 11 pM (3 pg/mL), while in girls the E2 concentrations ranged from 2.39 to 177 pM (0.65–48.3 pg/mL). TT levels were detectable in all 100 samples from children, with boys had much broader range (0.03–12.6 nM (0.77–363 ng/dL)) comparing to girls (0.04–0.85 nM (1.07–24.6 ng/dL)).

In conclusion, a highly reliable LC-MS/MS method with high accuracy, precision, sensitivity and specificity for measurement of total estradiol and testosterone in human serum has been developed. The method showed high level of accuracy and precision over two years. This method can measure TT over a wide concentration range that covers the low concentrations observed in children, women and men. It is also capable of measuring E2 concentrations ranging from very low concentrations typically seen in men and postmenopausal women to the very high concentrations observed in women at an active reproductive stage. It is therefore applicable to long term large population studies such as the National Health and Nutrition Examination Survey and other epidemiology studies.

## Acknowledgments

We would like to acknowledge CDC Hormone Laboratory members Lumi Duke, Paul Kim, Gabrielle D. Gay, Krista Poynter and Otoe Sugahara, for their support in the laboratory measurements. We thank Brandon Laughlin and Dr. Uliana Danilenko for their edits and review (all with the National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia).

## References

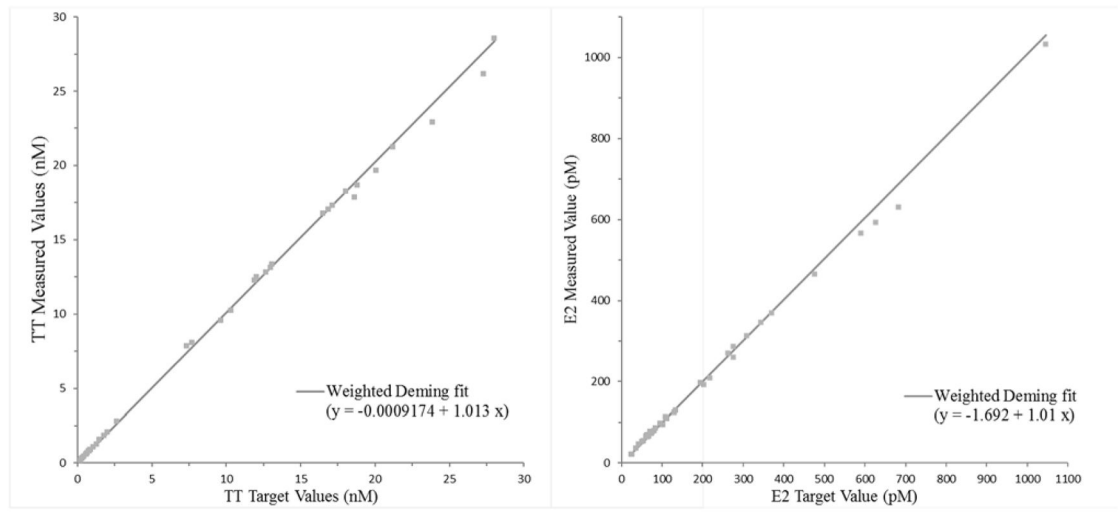
1. Sizonenko PC. Normal sexual maturation. *Pediatrician*. 1987; 14(4):191–201.
2. Grossmann M. Testosterone and glucose metabolism in men: current concepts and controversies. *J Endocrinol*. 2014; 220(3):R37–55. [PubMed: 24353306]
3. Khera M. Male hormones and men's quality of life. *Curr Opin Urol*. 2016; 26(2):152–7. [PubMed: 26765046]
4. Cauley JA. Estrogen and bone health in men and women. *Steroids*. 2015; 99(Pt A):11–5. [PubMed: 25555470]
5. Mongraw-Chaffin ML, Anderson CA, Allison MA, Ouyang P, Szklo M, Vaidya D, et al. Association between sex hormones and adiposity: qualitative differences in women and men in the multi-ethnic study of atherosclerosis. *J Clin Endocrinol Metab*. 2015; 100(4):E596–600. [PubMed: 25636047]
6. Loomba-Albrecht LA, Styne DM. The physiology of puberty and its disorders. *Pediatr Ann*. 2012; 41(4):e1–9.

7. Rahhal SN, Fuqua JS, Lee PA. The impact of assay sensitivity in the assessment of diseases and disorders in children. *Steroids*. 2008; 73(13):1322–7. [PubMed: 18565561]
8. Khera M, Adaikan G, Buvat J, Carrier S, El-Meliegy A, Hatzimouratidis K, et al. Diagnosis and Treatment of Testosterone Deficiency: Recommendations From the Fourth International Consultation for Sexual Medicine (ICSM 2015). *J Sex Med*. 2016; 13(12):1787–804. [PubMed: 27914560]
9. Schulman CC, Irani J, Morote J, Schalken JA, Montorsi F, Chlosta PL, et al. Testosterone measurement in patients with prostate cancer. *Eur Urol*. 2010; 58(1):65–74. [PubMed: 20434831]
10. Peavey M, Akbas N, Gibbons W, Zarutskie P, Devaraj S. ANNALS EXPRESS: Optimization of Estradiol Assays to Improve Utility in an In Vitro Fertilization Setting. *Ann Clin Biochem*. 2017 4563217691788.
11. Lonning PE. Estradiol measurement in translational studies of breast cancer. *Steroids*. 2015; 99(Pt A):26–31. [PubMed: 25159101]
12. Demers LM. Testosterone and estradiol assays: current and future trends. *Steroids*. 2008; 73(13): 1333–8. [PubMed: 18565562]
13. Herati AS, Cengiz C, Lamb DJ. Assays of Serum Testosterone. *Urol Clin North Am*. 2016; 43(2): 177–84. [PubMed: 27132574]
14. Santen RJ, Lee JS, Wang S, Demers LM, Mauras N, Wang H, et al. Potential role of ultra-sensitive estradiol assays in estimating the risk of breast cancer and fractures. *Steroids*. 2008; 73(13):1318–21. [PubMed: 18614192]
15. Cawood ML, Field HP, Ford CG, Gillingwater S, Kicman A, Cowan D, et al. Testosterone measurement by isotope-dilution liquid chromatography-tandem mass spectrometry: validation of a method for routine clinical practice. *Clin Chem*. 2005; 51(8):1472–9. [PubMed: 15961553]
16. Wang C, Catlin DH, Demers LM, Starcevic B, Swerdloff RS. Measurement of total serum testosterone in adult men: comparison of current laboratory methods versus liquid chromatography-tandem mass spectrometry. *J Clin Endocrinol Metab*. 2004; 89(2):534–43. [PubMed: 14764758]
17. Kushnir MM, Rockwood AL, Roberts WL, Pattison EG, Bunker AM, Fitzgerald RL, et al. Performance characteristics of a novel tandem mass spectrometry assay for serum testosterone. *Clin Chem*. 2006; 52(1):120–8. [PubMed: 16299050]
18. Rosner W, Hankinson SE, Sluss PM, Vesper HW, Wierman ME. Challenges to the measurement of estradiol: an endocrine society position statement. *J Clin Endocrinol Metab*. 2013; 98(4):1376–87. [PubMed: 23463657]
19. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*. 2007; 92(2):405–13. [PubMed: 17090633]
20. Demers LM, Hankinson SE, Haymond S, Key T, Rosner W, Santen RJ, et al. Measuring Estrogen Exposure and Metabolism: Workshop Recommendations on Clinical Issues. *J Clin Endocrinol Metab*. 2015; 100(6):2165–70. [PubMed: 25850026]
21. Wang Y, Gay GD, Botelho JC, Caudill SP, Vesper HW. Total testosterone quantitative measurement in serum by LC-MS/MS. *Clin Chim Acta*. 2014; 436:263–7. [PubMed: 24960363]
22. Caudill SP, Schleicher RL, Pirkle JL. Multi-rule quality control for the age-related eye disease study. *Stat Med*. 2008; 27(20):4094–106. [PubMed: 18344178]
23. CLSI. C62-A: Liquid chromatography-mass spectrometry methods; approved guideline. 1. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.
24. Becker, RCD., Currie, L., Diamondstone, B., Eberhardt, K., Gills, T., Hertz, H., Klouda, G., Moody, J., Parris, R., Schaffer, R., Steel, E., Taylor, J., Watters, R., Zeisler, R. Use of NIST standard reference materials for decisions on performance of analytical chemical methods and laboratories. Gaithersburg, MD: NIST Special Publication; 1992. p. 829
25. CLSI. EP9-A2: Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline. 2. Wayne, PA: Clinical and Laboratory Standards Institute; 2002.
26. Tai SS, Xu B, Welch MJ, Phinney KW. Development and evaluation of a candidate reference measurement procedure for the determination of testosterone in human serum using isotope

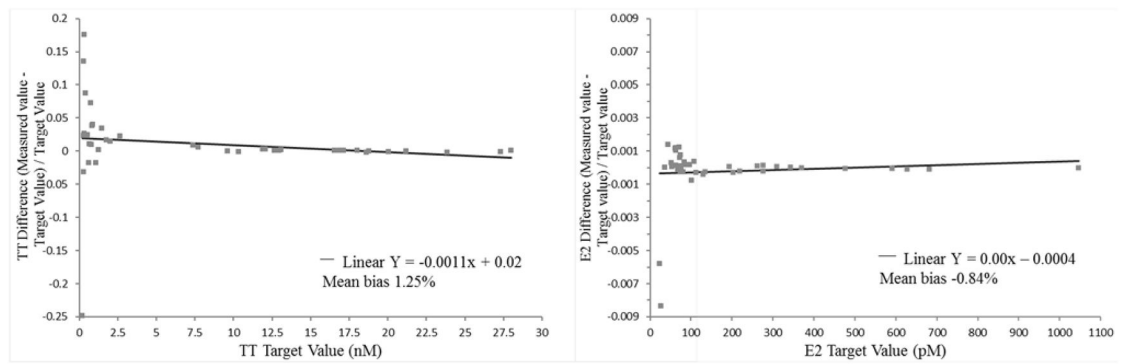
- dilution liquid chromatography/tandem mass spectrometry. *Anal Bioanal Chem.* 2007; 388(5–6): 1087–94. [PubMed: 17530229]
27. Thienpont LM, De Leenheer AP. Efforts by industry toward standardization of serum estradiol-17 beta measurements. *Clin Chem.* 1998; 44(3):671–4. [PubMed: 9510882]
  28. Thienpont LM, De Brabandere VI, Stockl D, De Leenheer AP. Use of cyclodextrins for prepurification of progesterone and testosterone from human serum prior to determination with isotope dilution gas chromatography/mass spectrometry. *Anal Chem.* 1994; 66(22):4116–9. [PubMed: 7810905]
  29. Botelho JC, Ribera A, Cooper HC, Vesper HW. Evaluation of an Isotope Dilution HPLC Tandem Mass Spectrometry Candidate Reference Measurement Procedure for Total 17-beta Estradiol in Human Serum. *Anal Chem.* 2016; 88(22):11123–9. [PubMed: 27744701]
  30. Botelho JC, Shacklady C, Cooper HC, Tai SS, Van Uytvanghe K, Thienpont LM, et al. Isotope-dilution liquid chromatography-tandem mass spectrometry candidate reference method for total testosterone in human serum. *Clin Chem.* 2013; 59(2):372–80. [PubMed: 23213081]
  31. Yun YM, Botelho JC, Chandler DW, Katayev A, Roberts WL, Stanczyk FZ, et al. Performance criteria for testosterone measurements based on biological variation in adult males: recommendations from the Partnership for the Accurate Testing of Hormones. *Clin Chem.* 2012; 58(12):1703–10. [PubMed: 23065474]
  32. Vesper HW, Botelho JC, Vidal ML, Rahmani Y, Thienpont LM, Caudill SP. High variability in serum estradiol measurements in men and women. *Steroids.* 2014; 82:7–13. [PubMed: 24407040]
  33. CLSI. EP5-A3: Evaluation of Precision of Quantitative Measurement Procedures; approved guideline. 3. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.
  34. Rodriguez M, Orescan DB. Confirmation and quantitation of selected sulfonylurea, imidazolinone, and sulfonamide herbicides in surface water using electrospray LC/MS. *Anal Chem.* 1998; 70(13): 2710–7. [PubMed: 9666735]
  35. CLSI. EP17-A2: Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline. 2. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.
  36. Taylor, JK. Quality assurance of chemical measurements. Boca Raton, FL: CRC-Press; 1987.
  37. Fraser, CG. Biological variation: from principles to practice. Washington, DC: American Association for Clinical Chemistry, Inc; 2001.
  38. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, et al. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest.* 1999; 59(7):491–500. [PubMed: 10667686]
  39. Vesper HW, Bhasin S, Wang C, Tai SS, Dodge LA, Singh RJ, et al. Interlaboratory comparison study of serum total testosterone [corrected] measurements performed by mass spectrometry methods. *Steroids.* 2009; 74(6):498–503. [PubMed: 19428438]
  40. Nelson RE, Grebe SK, DJOK, Singh RJ. Liquid chromatography-tandem mass spectrometry assay for simultaneous measurement of estradiol and estrone in human plasma. *Clin Chem.* 2004; 50(2): 373–84. [PubMed: 14656902]
  41. Kushnir MM, Rockwood AL, Bergquist J, Varshavsky M, Roberts WL, Yue B, et al. High-sensitivity tandem mass spectrometry assay for serum estrone and estradiol. *Am J Clin Pathol.* 2008; 129(4):530–9. [PubMed: 18343779]
  42. Schofield RC, Mendu DR, Ramanathan LV, Pessin MS, Carlow DC. Sensitive simultaneous quantitation of testosterone and estradiol in serum by LC–MS/MS without derivatization and comparison with the CDC HoSt program. *Journal of Chromatography B.* 2017; 1048:70–6.
  43. Zhang Q, Han L, Wang J, Lin H, Ke P, Zhuang J, et al. Simultaneous quantitation of endogenous estrone, 17beta-estradiol, and estriol in human serum by isotope-dilution liquid chromatography-tandem mass spectrometry for clinical laboratory applications. *Anal Bioanal Chem.* 2017; 409(10): 2627–38. [PubMed: 28224243]
  44. Vesper HW, Botelho JC. Standardization of testosterone measurements in humans. *J Steroid Biochem Mol Biol.* 2010; 121(3–5):513–9. [PubMed: 20302935]
  45. Hurwitz AR, Liu ST. Determination of aqueous solubility and pKa values of estrogens. *J Pharm Sci.* 1977; 66(5):624–7. [PubMed: 17710]

46. Boggs AS, Bowden JA, Galligan TM, Guillette LJ Jr, Kucklick JR. Development of a multi-class steroid hormone screening method using Liquid Chromatography/Tandem Mass Spectrometry (LC-MS/MS). *Anal Bioanal Chem.* 2016; 408(15):4179–90. [PubMed: 27039201]
47. Ray JA, Kushnir MM, Rockwood AL, Meikle AW. Direct Measurement of Free Estradiol in Human Serum and Plasma by Equilibrium Dialysis-Liquid Chromatography-Tandem Mass Spectrometry. *Methods Mol Biol.* 2016; 1378:99–108. [PubMed: 26602122]
48. Vitku J, Chlupacova T, Sosvorova L, Hampl R, Hill M, Heracek J, et al. Development and validation of LC-MS/MS method for quantification of bisphenol A and estrogens in human plasma and seminal fluid. *Talanta.* 2015; 140:62–7. [PubMed: 26048824]
49. 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(14):7180–6. [PubMed: 26090565]

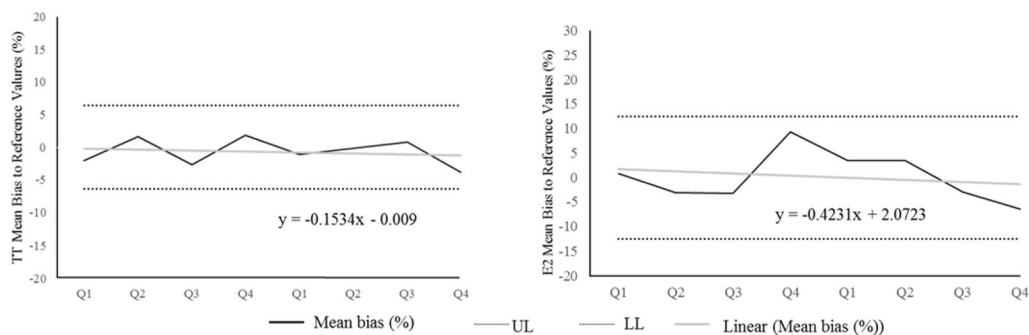
(a) Scatter Plot with Deming-fit



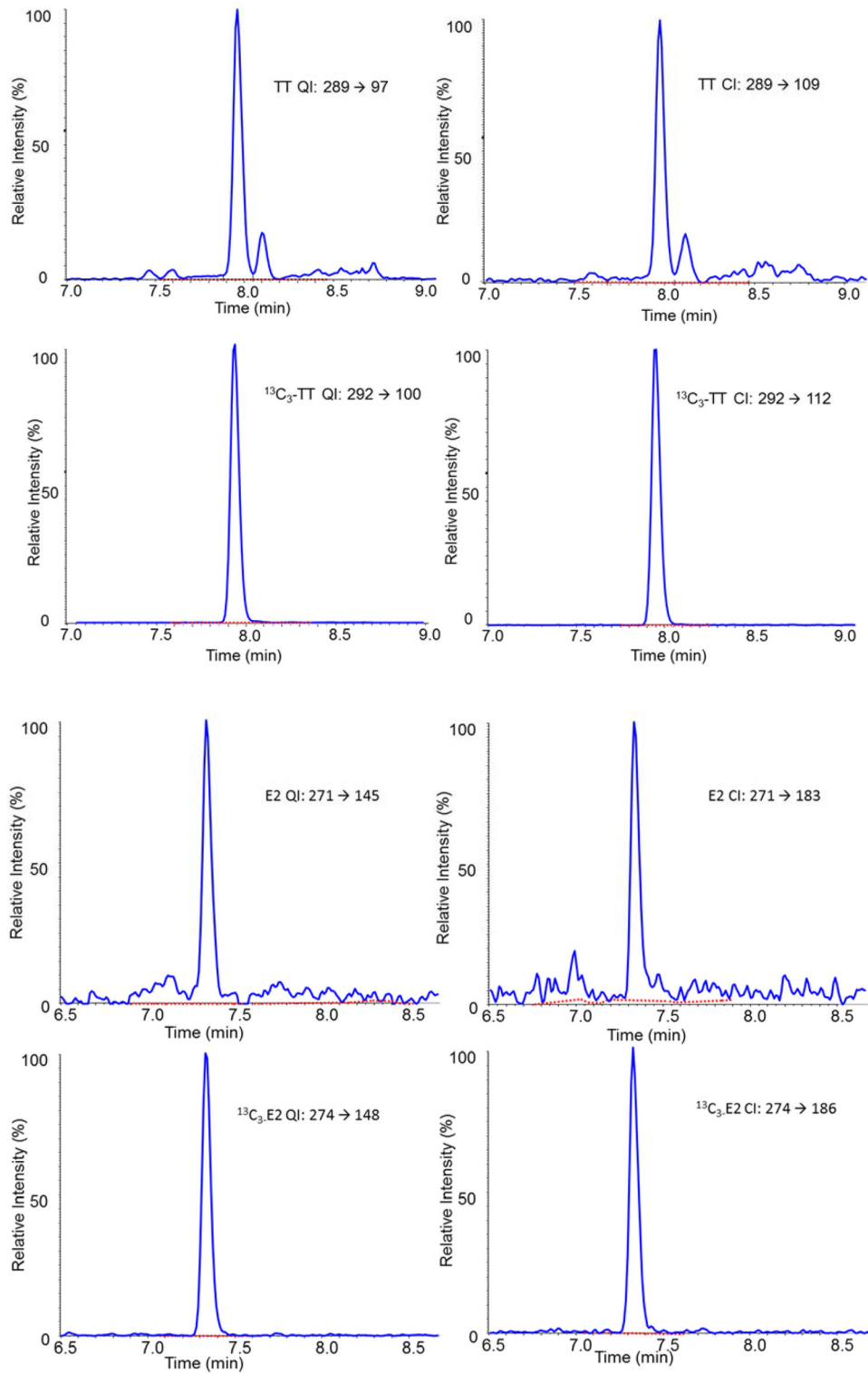
(b) Difference plot



(c) Mean bias by quarter in CDC HoSt program



**Fig. 1.** Deming regression (a) and bias plot (b) between described method and metrological reference methods (26–30) using 40 individual patient serum samples. (c) TT and E2 measurement performance (mean bias (%)) using CDC HoSt reference materials that included 10 serum samples for each analyte per quarter, total of 80 samples for each analyte. The performance is shown over 2-year period.

**Fig. 2.**

LC-MS/MS selected ion chromatograms of the quantification transitions and confirmation transitions for testosterone 1.20 nM (34.5 ng/dL) and estradiol 103 pM (28.1 pg/mL) in a patient serum sample.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 1**

Steroid hormones used for interference analysis. Their Molecular weight and source of materials are provided.

<b>Steroid</b>	<b>MW (g/mol)</b>	<b>Vendor</b>
1,4-androstadien-17b-ol-3-one	286.41	Steraloids, Inc. (Newport, RI)
4, 16-Androstadien-3 $\beta$ -ol	272.43	Steraloids, Inc. (Newport, RI)
5, 16-Androstadien-3 $\beta$ -ol	272.43	Steraloids, Inc. (Newport, RI)
5-androsten-3b 17-diol	292.46	Steraloids, Inc. (Newport, RI)
4,6-androstadien-17b-ol-3-one	286.41	Steraloids, Inc. (Newport, RI)
Androstenediol	290.44	Steraloids, Inc. (Newport, RI)
Androstenedione	290.44	LGC Standards (Manchester, NH)
16, (5 $\alpha$ )-Androsten-3-one	272.43	Steraloids, Inc. (Newport, RI)
2, (5 $\alpha$ )-Androsten-17-one	272.43	Steraloids, Inc. (Newport, RI)
Androsterone	290.4	Steraloids, Inc. (Newport, RI)
Corticosterone	346.46	Cerilliant (Round Rock, TX)
Cortisone	360.44	Cerilliant (Round Rock, TX)
Trans-Dehydroandrosterone	288.42	Sigma-Aldrich (St. Louis, MO)
Dehydroepiandrosterone (DHEA)	288.42	Cerilliant (Round Rock, TX)
Dehydroepiandrosterone sulfate (DHEAS)	368.49	Cerilliant (Round Rock, TX)
11-Deoxycortisol	346.46	Cerilliant (Round Rock, TX)
5 $\alpha$ -Dihydrotestosterone (DHT)	290.44	Cerilliant (Round Rock, TX)
Epitestosterone	288.42	Sigma-Aldrich (St. Louis, MO)
Estriol	288.38	Cerilliant (Round Rock, TX)
Estrone	270.37	Cerilliant (Round Rock, TX)
17 $\alpha$ -Ethinylestradiol	296.40	Cerilliant (Round Rock, TX)
Etiocholanolone	290.45	Cerilliant (Round Rock, TX)
Hydrocortisone	362.46	Steraloids, Inc. (Newport, RI)
17 $\alpha$ -Hydroxypregnenolone	332.48	Cerilliant (Round Rock, TX) (NMI)
17 $\alpha$ -Hydroxyprogesterone	330.46	Cerilliant (Round Rock, TX)
21-Hydroxyprogesterone	330.46	Sigma-Aldrich (St. Louis, MO)
17- $\alpha$ Methyltestosterone	302.46	Cerilliant (Round Rock, TX)
19-Norethindrone	298.42	Steraloids, Inc. (Newport, RI)
D(-)-Norgestrel	312.45	Sigma-Aldrich (St. Louis, MO)
5-Pregnen-3 $\beta$ -ol-20-one	316.48	Sigma-Aldrich (St. Louis, MO)
Pregnenolone	316.48	IsoSciences (King of Prussia, PA)
Progesterone	314.46	Cerilliant (Round Rock, TX)

**Table 2**

Method accuracy, expressed as percent bias, evaluated using certified, serum-based reference materials

Samples (analyte)	Certified value (95%CI)	Measured value (n=5) (95%CI)	Bias in % (95%CI)
SRM 971 M (TT)	22.3 nM (643 ng/dL) 21.8–22.8 (629–658)	21.9 nM (632 ng/dL) .7–22.1 (627–636)	–1.5 (–2.1–0.83) <sup>a</sup>
SRM 971 F (TT)	0.96 nM (27.7 ng/dL) 0.94–0.98 (27.1–28.4)	0.94 nM (27.1 ng/dL) 0.91–0.97 (26.3–28.0)	–2.1 (–5.0–0.92) <sup>a</sup>
BCR 576 (E2)	114 pM (31.1 pg/mL) 109–119 (29.7–32.4)	113 pM (30.7 pg/mL) 107–118 (29.1–32.2)	–1.3 (–6.28–3.59) <sup>a</sup>
BCR 577 (E2)	690 pM (188 pg/mL) 650–730 (177–199)	683 pM (186 pg/mL) 664–697 (181–190)	–1.2 (–3.53–1.13) <sup>a</sup>
BCR 578 (E2)	1340 pM (365 pg/mL) 1270–1410 (346–384)	1318 pM (359 pg/mL) 1296–1336 (353–364)	–1.7 (–3.25–0.22) <sup>a</sup>

<sup>a</sup>No significant bias, as determined following NIST Special Publication 829 (24).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Assay precision for testosterone (TT) and estradiol (E2) in serum at three concentration levels determined in duplicates per run, with two runs per day over 68 days (2-year period)

**Table 3**

Sample Description	Mean value (n = 272)		Within-Run precision		Between-Run precision		Between-Day precision		Within-Laboratory precision	
	TT nM (ng/dL)	E2 pM (pg/mL)	TT %CV	E2 %CV	TT %CV	E2 %CV	TT %CV	E2 %CV	TT %CV	E2 %CV
QC High	30.2 (870)	2540 (692)	2.43	2.99	2.43	3.13	1.23	1.46	2.72	3.33
QC Medium	6.07 (175)	448 (122)	2.4	3.3	2.4	3.3	1.56	1.23	2.87	3.53
QC Low	1.27 (36.6)	103 (28.0)	2.65	5.02	2.65	5.02	1.26	1.84	2.93	5.34

Comparison of testosterone (TT) and estradiol (E2) concentrations obtained from diluted samples and undiluted serum samples

**Table 4**

	TT nM(ng/dL)			E2 pM (pg/mL)			TT Recovery (%)			E2 Recovery (%)		
	Mean	95%CI		Mean	95%CI		Mean	95%CI		Mean	95%CI	
Undiluted	29.6 (853)	29.1–30.1 (838–868)	3070 (836)	3005–3134 (819–854)	100.0	98.0–101.7	98.9	96.4–101.4				
Diluted	29.6 (853)	29.4–29.9 (848–864)	3035 (827)	2958–3191 (806–870)								

**Table 5**  
Serum E2 and TT concentrations in individuals at different age groups and gender

Age group /Gender	n	E2 Median pM (pg/mL)	E2 (Min-max pM (pg/mL)	TT MediannM (ng/dL)	TT (Min-max) nM (ng/dL)
6-11 years male	50	< 11.0 (2.99)	0.00-39.6 (0.00-10.8)	0.11 (3.08)	0.03-12.6 (0.77-363)
6-11 years female	50	13.0 (3.54)	2.39-177 (0.65-48.3)	0.23 (6.53)	0.04-0.85 (1.07-24.6)
18 years and older male	50	80.0 (21.8)	17.3-167 (4.72-45.6)	13.3 (385)	0.62-32.5 (18.0-938)
18 yrs to 59 years female	50	132 (36.0)	11.4-12357 (3.11-3367)	0.69 (19.9)	0.28-2.55 (8.03-73.6)
60 years and older female	50	21.8 (5.95)	9.25-1152 (2.52-314)	0.72 (20.7)	0.26-4.13 (7.54-119)