David J. Handelsman*, Graham Jones, Dorothy Kouzios and Reena Desai

Evaluation of testosterone, estradiol and progesterone immunoassay calibrators by liquid chromatography mass spectrometry

https://doi.org/10.1515/cclm-2022-1179
Received November 17, 2022; accepted May 9, 2023;
published online June 5, 2023

Abstract

Objectives: In clinical practice, steroid measurements are performed mainly by direct, non-extraction immunoassays adapted to high throughput, automated immunoassay platforms and employing secondary calibrators. The accuracy of such steroid immunoassays is limited by cross-reactivity with structurally related steroids and nonspecific matrix interference as well as the metrological traceability of manufacturer supplied calibrators. The accuracy of steroid immunoassay calibrators has been little investigated by independent chemical methods.

Methods: Steroid concentrations of 41 calibrators (4–6 replicates per calibrator) supplied by four manufacturers for use in testosterone (T), estradiol (E_2), and progesterone (P_4) commercial immunoassays were measured by ultra-pressure liquid chromatography-mass spectrometry (UPLC-MS).

Results: Among 14 non-zero T calibrators, six (43 %) deviated significantly from the label concentration with 29 % outside 20 % of it. Among 14 E_2 calibrators, eight (57 %) deviated significantly, whereas seven (50 %) were outside 20 % of the label concentration. Among 11 P_4 calibrators, eight (73 %) deviated significantly whereas four (36 %) were outside within 20 % of the label concentration.

Conclusions: We conclude that inaccurate calibration of manufacturer's supplied standards may contribute to inaccuracy of commercial direct steroid immunoassays.

Keywords: calibrators; estradiol; liquid chromatography mass spectrometry; progesterone; steroid immunoassay; testosterone.

Introduction

Worldwide, most clinical pathology laboratories currently measure circulating sex steroids, testosterone (T), estradiol (E_2) and progesterone (P_4) by direct immunoassays using serum or plasma without pre-analytical preparation such as extraction or chromatography, to adapt the methodology to high throughput, automated multiplex immunoassay platforms. These non-extraction immunoassays have known limitations in cross-reactivity from steroid precursors or metabolites structurally related to the target steroid as well as non-specific matrix interference from serum or plasma. These assays obtain their metrological traceability through matched calibrator sets obtained from the kit manufacturer rather than using independent gravimetrically pure steroid standards, despite their ready availability. Verification of correct method calibration is a key issue for the validity of quantitative analytical systems [1, 2]. Therefore, in this study we evaluated the values assigned to sex steroid immunoassay calibrators in widely used direct immunoassays for T, E_2 and P_4 by measurement with ultra-pressure liquid chromatography mass spectrometry (UPLC-MS) with validated traceability as an independent chemical method. Details of the assigned values, the higher order reference materials and reference methods on which traceability is based, the claimed measurement uncertainty and the composition of the calibrators was obtained from manufacturers' information.

Materials and methods

Materials

Steroid standards for measurements used Certified Reference Materials for T (M914b, National Measurement Institute, Australia (NMI), Sydney, uncertainty ±1.9 %, coverage factor of 2), E_2 (E060, 1 mg/mL E_2 solution in acetonitrile, Cerilliant, Round Rock, Texas, USA, ±3.0 %; BCR-576, ±4.4 %, BCR-577 ± 5.8 %, BCR-578 ± 5.2, European Commission's Institute for

***Corresponding author: Prof. David J. Handelsman**, MBBS, PhD, ANZAC Research Institute, University of Sydney, Concord, NSW 2139, Australia; and Department of Andrology, Concord Hospital, Concord, NSW 2139, Australia, E-mail: djh@anzac.edu.au
Graham Jones, SydPath, St Vincent's Hospital, Sydney and University of NSW, Darlinghurst, Australia
Dorothy Kouzios, Diagnostic Pathology Unit, Concord Repatriation Hospital, Concord, NSW, Australia
Reena Desai, ANZAC Research Institute, University of Sydney, Concord, NSW, Australia

lyophilized human serum as reference materials) and P_4 (P-069, Cerilliant, USA, ±0.6 %). Stable isotope internal standards were d3-T (D507a, NMI), d4-E_2 (DLM2487-0, Cambridge Isotope Laboratories, USA) and d9-P_4 (D5385, CDN Isotopes, Canada) for calibration and quantification. All chemical of HPLC grade were purchased from Lab-Scan Analytical Science (Dublin, Ireland) unless otherwise specified.

Sample preparation

The steroid UPLC-MS and its validation were described previously [3]. Briefly, aliquots of serum, standards and quality control samples were mixed with deuterated steroid internal standards and methyl-tert butyl ether (MTBE). The organic layer was decanted, the solvent evaporated with dried extracts re-suspended in a methanol:water mix and transferred into a microtitre plate for injection into the UPLC-MS.

UPLC-MS steroid analysis

The UPLC conditions comprised elution of steroids using a methanol/water gradient from a Kinetex Phenyl Hexyl column with a Phenomenex guard cartridge at a column temperature of 45 °C and flow rate of 0.35 mL/min and a total run time of 13 min. The chromatography provided baseline separation for each steroid with estradiol at 5.74 min (in negative ionization mode) and testosterone at 7.07 min and P4 at 10.88 min (in positive ionization mode).

The original mass spectrometry conditions comprised an API-5000 triple-quadrupole mass spectrometer (Applied Biosystems/MDS SCIEX, Foster City, CA/Concord, Ontario, Canada) equipped with an atmospheric pressure photoionization (APPI) source. When the API-5000 mass spectrometer became unavailable, the same profiling method was established on a Sciex 6500+ triple-quadrupole mass spectrometer (SCIEX, Foster City, CA/Concord, Ontario, Canada) equipped with an electrospray ionization (ESI) source operating in both positive and negative modes. The curtain gas – 40psi, GS1-70psi, GS2-60psi, temperature 600deg, IS-4500V, CAD-9. Nitrogen was employed as curtain, nebulizer, collision gases. Unit mass resolution was set in both mass-resolving quadrupole Q1 and Q3. Multiple reaction monitoring (MRM) was used to quantify the steroids in both systems. The MRM's quantified, their dwell time (ms) and collision energies (V) were E_2 (271-> 145, 200, −57 V) and d4-E_2 (275-> 147, 100, 57) in negative ionization mode and T (289-> 109, 200, 35), d3-T (292-> 109, 100, 35) and P_4 (315-> 97, 200, 34) and d9-P_4 (324->100, 100, 34) in positive ionization mode.

The limits of detection (LOD) and limit of quantitation (LOQ) were defined according to FDA [4] and European Medicines Agency [5] guidelines (Supplementary Material, Table 1). Reproducibility was based on the coefficient of variation (%) at four quality control (QC) levels for T and P_4 and three levels for E_2 with 10 within-day measurements as well as replicate analysis of manufacturer's calibrators as shown below. Between-day variation was measured with 15 replicates over 3 days. Extraction recovery was based on the proportion of a pre-extraction spiked sample that was recovered in the analysis and matrix effect is based on the recovery of post extraction spiked samples [6].

Long-term stability of steroid measurements in this method was established from in-house internal quality control (QC) samples for testosterone (414 runs) at high (target 8.0 ng/mL [27.8 nmol/L], 97 ± 0.3 [SEM] %), mid (target 1.6 ng/mL [5.6 nmol/L], 100 ± 0.4 %) and low (target 0.35 ng/mL [1.2 nmol/L], 99 ± 0.5 %) range, for estradiol (345 runs) at high (target 400 pg/mL [1,468 pmol/L], 103 ± 2 %), mid (target 80 pg/mL [294 pmol/L], 99 ± 1 %) and low (target 20 pg/mL [73 pmol/L], 97 ± 1 %) range and for progesterone (65 runs) at high (target 16.0 ng/mL [50.9 nmol/L], 99 α ± 0.7 %), mid (target 3.2 ng/mL [10.2 nmol/L], 98 ± 1.2 %) and low (0.8 ng/mL [2.5 nmol/L], 98 ± 1.5 %) range.

The aim of the present study is to identify whether the true steroid concentrations in the manufacturers' calibrators differ from the assigned values by more than pre-specified amounts. The true value is best described by the CRMs used to calibrate the assays, i.e. the line of identity in the assays. To the pre-specified limits must be added the following uncertainties: uncertainty of purity of reference material; uncertainty of preparation of MS assay calibrators; within run imprecision of the assays (noting that all manufacturer's standards for each steroid were tested in the same run), bias due to manufacturer's calibrator matrix and the stated uncertainty of value assignment of the manufacturer's calibrators. Obtained results will be classified as exceeding the set criteria at the 95 % confidence level after these uncertainties have been included.

Steroid immunoassay calibrators

As direct steroid immunoassays employ manufacturer supplied calibrators, we investigated the accuracy of the calibrators provided with the assays for three direct sex steroid immunoassays and compared the assigned steroid concentrations with Certified Reference Materials (CRM) for T, E_2 and P_4 measured in a validated ultra-pressure liquid chromatography-mass spectrometry method (UPLC-MS) [3] (Table 1). Calibrators were obtained for methods from the following manufacturers: Abbott Architect, Beckman-Coulter Access, Roche Cobas, Siemens Atellica (Supplementary Table 2). The steroid immunoassay calibrators were prepared according to manufacturer's instruction and measured in as many replicates as possible from the stock vials (from 4 to 6 per calibrator standard).

Assessment of calibrator bias was performed using criteria of ±6.4 % for testosterone, ±12.5 % for estradiol with ±2.5 pg/mL [9.2 pmol/L] for values below 20 pg/mL [73 pmol/L] (both taken from the CDC HoST steroid accuracy program bias criteria https://www.cdc.gov/labstandards/hs_standardization.html accessed 3/12/20). For progesterone a pragmatic criterion of ±10 % was used. A wider criterion of values outside ±20 % was also used for all analytes. Additionally, assessment was made relative to the stated uncertainty of the materials.

To verify the validity of steroid analysis in the matrix of the calibrators, spike-recovery experiments (6 replicates) using the certified reference materials as spike and performed in the calibrator matrix (with or without spike) using the zero calibrator, if available, otherwise another calibrator, to estimate the recovery (%) after LCMS measurement including pre-analytical solvent extraction.

Results

Manufacturers' descriptions of the calibrators

In addition to the lot number details of the calibrators tested, Supplementary Table 2 shows the provided traceability information from the manufacturers including the reference material used at the top of the traceability chain.

Supplementary Table 3 shows the calibrator compositions with generally a limited description of the source and nature of the steroid as well as the calibrator matrix. The claimed measurement uncertainty of each level of each calibrator along with the assigned values and results is shown for the 15 T calibrators (Supplementary Table 2), 15 E_2 calibrators (Supplementary Table 3) and 12 P_4 calibrators (Supplementary Table 4) from four manufacturers (Abbott, Beckman, Roche, Siemens). Of note, the Roche calibrators have different values assigned for different lot numbers of reagents making the assigned values approximate only.

Measurement of calibrators

For T (Figure 1, Table 1), the mean of the measurement results for six of the 14 (43 %) non-zero calibrators were outside the 6.4 % bias criteria, with five passing this level at the 95 % probability level with all uncertainties included. Four were more than 20 % from the stated value and for four calibrators, their range of values excluded the assigned value. The two lowest of five Abbott non-zero calibrators were below 80 % of the expected value, the Beckman values ranged between 98 and 113 %; the two Roche calibrators were 82 and 97 % and Siemens were over 129 % of stated values.

For E_2 (Figure 2, Table 2), the mean of measurement results for eight of the 14 (57 %) calibrators were outside the selected limits with six passing this level at the 95 % probability level with all uncertainties included. Seven deviated greater than 20 % from the assigned value and for nine calibrators, their range excluded the assigned value. Additionally, only four were within the claimed measurement uncertainty from the manufacturer. Abbott showed a gradual rise from 92 % to over 130 % with increasing concentration while Beckman showed a fall from over 400–106 % over the range. The higher Roche calibrator was 158 % of the assigned value.

For P_4, (Figure 3, Table 3) all 11 calibrators were below the assigned value with eight of the 11 (73 %) calibrators outside 10 % of the assigned value, five with 95 % probability, and four were outside 20 %.

For testosterone calibrators, the spike-recovery was 93.4 ± 1.1 % (Architect), 91.8 ± 0.7 % (Beckman), 87.4 ± 0.4 % (Roche) and 94.9 ± 0.9 % (Siemens). For estradiol calibrators, the spike-recovery was 107.5 ± 1.6 % (Architect), 95.0 ± 1.3 % (Roche) and 88.8 ± 1.3 % (Siemens). For progesterone calibrators, the spike-recovery was 95.5 ± 9.0 % (Architect), 106.2 ± 3.5 % (Beckman), 103.3 ± 7.8% (Roche) and 107.8 ± 3.5 % (Siemens). The quantitative recovery of spiked standards and uncertainty of recovery was included in assessing the statistical likelihood of exceeding the prescribed bias between the different calibrator measurement results.

Discussion

The present study shows that the assigned concentrations for the manufacturer supplied calibrators used for the three major classes of direct (non-extraction) sex steroid

Table 1: Testosterone immunoassay calibrators.

Calibrator	Label concentration	Replicates measured	Measured, ng/mL mean [SEM] (range)	Measured % mean (range)[a] = sign.	Claimed MU
Abbott T A	0 ng/mL	6			NA
Abbott T B	0.03 ng/mL	6	0.023 [0.001] (0.021–0.025)	76 (70–84)[a]	1.0 %
Abbott T C	0.06 ng/mL	6	0.047 [0.001] (0.043–0.050)	78 (72–83)[a]	1.0 %
Abbott T D	0.46 ng/mL	6	0.46 [0.009] (0.44–0.49)	100 (95–107)	0.6 %
Abbott T E	3.61 ng/mL	6	3.67 [0.023] (3.60–3.70)	102 (100–104)	0.6 %
Abbott T F	8.65 ng/mL	6	8.41 [0.148] (7.96–8.93)	97 (92–103)	0.6 %
Beckman T S1	0.5 ng/mL	6	0.49 [0.01] (0.46–0.51)	98 (92–101)	3.6 %
Beckman T S2	1.5 ng/mL	6	1.70 [0.02] (1.65–1.77)	113 (110–118)[a]	3.3 %
Beckman T S3	3.9 ng/mL	6	4.15 [0.05] (3.96–4.32)	98 (51–111)	3.3 %
Beckman T S4	7.7 ng/mL	6	7.85 [0.10] (7.47–8.10)	102 (97–105)	3.4 %
Beckman T S5	16.0 ng/mL	6	15.7 [0.19] (15.0–16.1)	98 (94–101)	3.3 %
Roche T II	0.4 ng/mL	6	0.33 [0.001] (0.32–0.33)	82 (81–83)	4.8 %
Roche T II	11.5 ng/L	6	11.2 [0.18] (10.9–12.0)	97 (95–104)	3.0 %
Siemens L	0.21 ng/mL	4	0.44 [0.01] (0.43–0.46)	212 (204–222)[a]	8.2 %
Siemens H	13.5 ng/mL	4	17.4 [0.23] (16.7–17.8)	129 (124–132)[a]	8.5 %

To convert ng/mL to nmol/L multiply by 3.47. To convert nmol/L to ng/mL divide by 3.47. MU, measurement uncertainty, coverage factor of 2. [a]Result exceeds ±6 % bias with >95 % certainty.

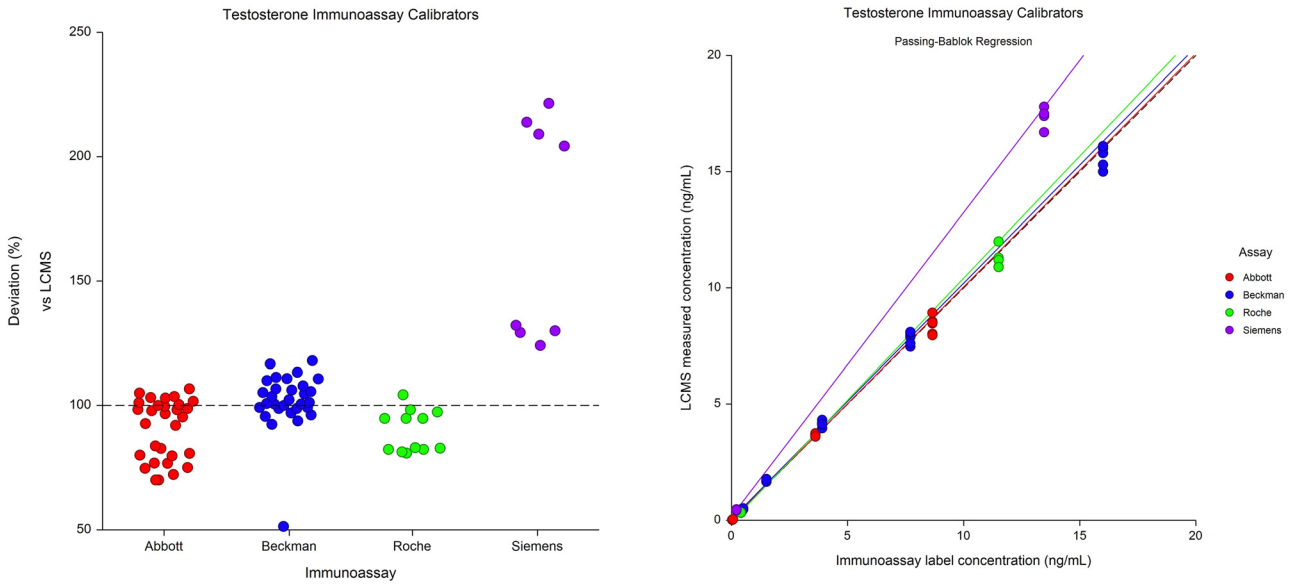

Figure 1: Bland-Altman deviance plot (left) and Passing-Bablok regression plot (right) for testosterone calibrators for five different commercial calibrators. The line of identity is shown as a dashed black line.

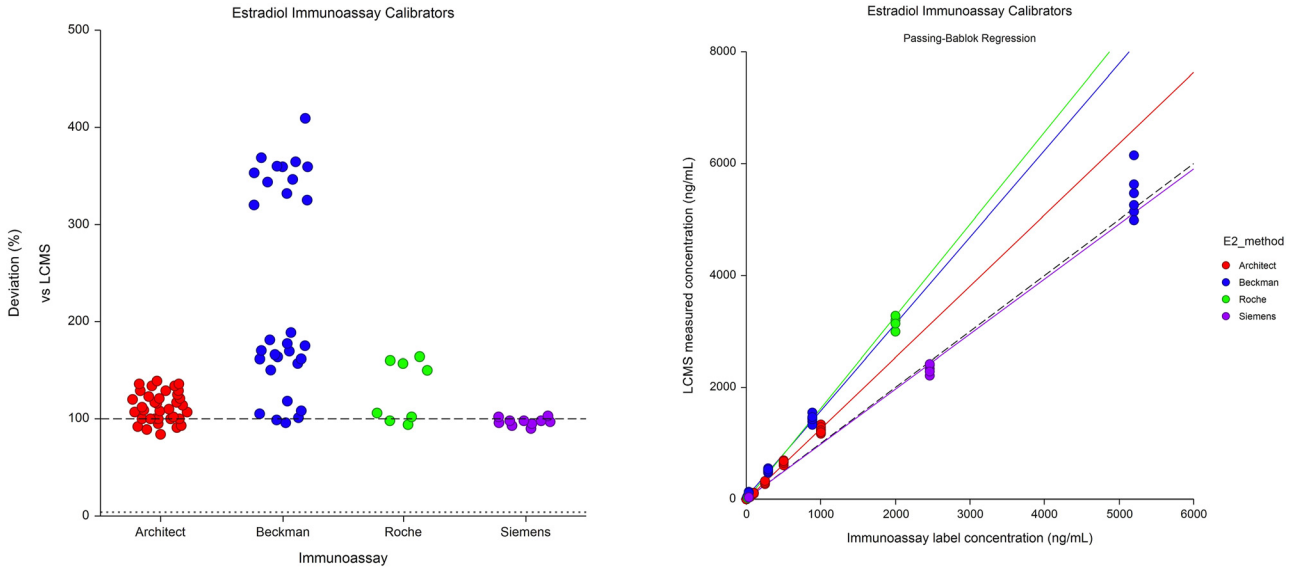

Figure 2: Bland-Altman deviance plot (left) and Passing-Bablok regression plot (right) for estradiol calibrators for four different commercial calibrators. The line of identity is shown as a dashed black line.

immunoassays are frequently, and often markedly, inaccurate. Even by the low standard of being within 20 % of the stated label concentration, only 57 % of T, 50 % of E_2 calibrators and 78 % of P_4 calibrators were within this margin. This finding adds another new dimension to explaining the discrepancies between direct steroid immunoassays displays in relation to steroid mass spectrometry-based methods as well as between different direct immunoassay methods. Although the external calibrators may be designed to bring the immunoassay into compliance with MS-based measurements, it may in fact contribute to the bias as well as adding to the method-specific bias.

Additionally, the assigned uncertainties of the calibrators rarely encompassed the discrepancies and there was wide variation in these between 0.5 % (Abbott E2 higher values) and greater than 10 % for Siemens E2 and P. The uncertainties are all provided with a coverage factor of 2, and indeed some of the very low values seem difficult to

Table 2: Estradiol immunoassay calibrators.

Calibrator	Label concentration	Replicates measured	Measured, pg/mL mean [SEM] (range)	Measured % mean (range)[a] = sign.	Claimed MU
Architect E2 A	0 pg/mL	6			N/A
Architect E2 B	50 pg/mL	6	45.7 [1.2] (42–51)	92 (84–102)	1.5 %
Architect E2 C	100 pg/mL	6	108 [2.7] (95–114)	108 (95–114)	1.5 %
Architect E2 D	250 pg/mL	6	292 [8.0] (268–323)	117 (107–129)	0.5 %
Architect E2 E	500 pg/mL	6	758 [97] (604–1,336)	133 (121–139)[a]	0.5 %
Architect E2 F	1,000 pg/mL	6	1,229 [21] (1,170–1,336)	123 (117–129)	0.5 %
Beckman E2 S1	11 pg/mL	6	37.9 [0.75] (35–40)	344 (320–365)[a]	8.10 %
Beckman E2 S2	32 pg/mL	6	116 [3.58] (104–131)	363 (325–409)[a]	3.20 %
Beckman E2 S3	292 pg/mL	6	510 [11.5] (472–551)	175 (162–189)[a]	1.92 %
Beckman E2 S4	885 pg/mL	6	1,437 [30.4] (1,330–1,550)	162 (150–175)[a]	1.60 %
Beckman E2 S5	5,200 pg/mL	6	5,440 [170] (4,990–6,150)	105 (96–118)	1.88 %
Roche E2 Gen3 Cal 1	20 pg/mL	4	20.0 [0.52] (18.8–21.2)	100 (94–106)	17.7 %
Roche E2 Gen3 Cal 2	2000 pg/mL	4	3,154 [60.0] (2,996–3,280)	158 (150–164)[a]	3.4 %
Siemens 30 Cal L	33.2 pg/mL	5	32.9 [0.47] (31.8–34.2)	99 (96–103)	12.8 %
Siemens 30 Cal H	2,460 pg/mL	56	2,330 [37.9] (2,212–2,419)	95 (90–98)	2.9 %

To convert pg/mL to pmol/L multiply by 3.67. To convert pmol/L to pg/mL divide by 3.67. MU, measurement uncertainty, coverage factor of 2. [a]Result exceeds ±12.5 % bias with >95 % certainty.

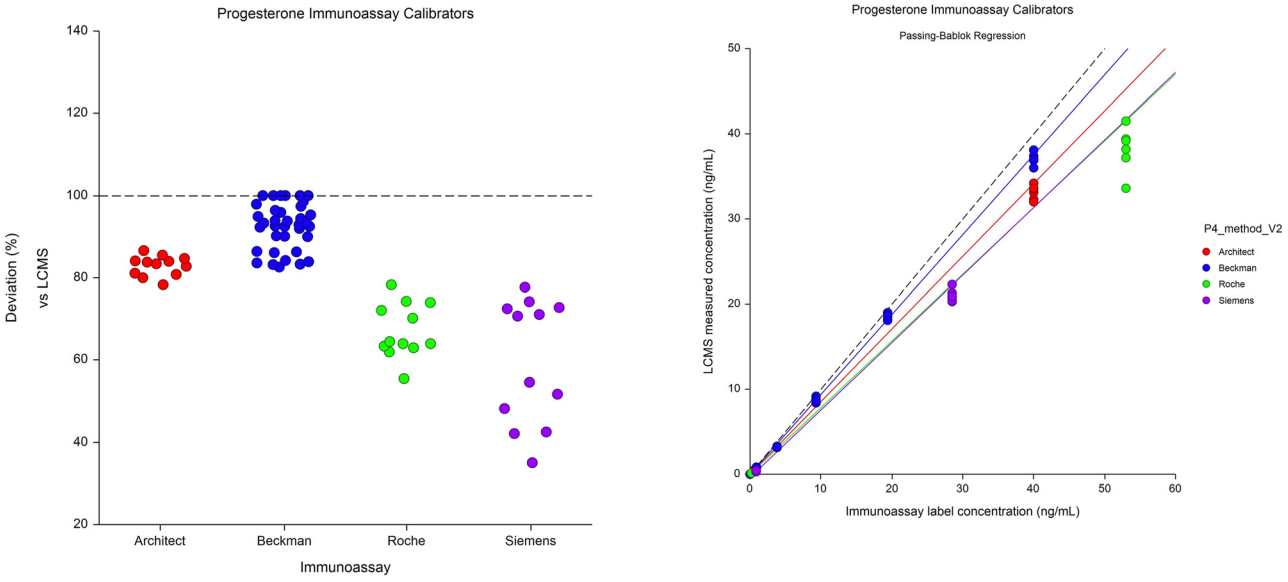

Figure 3: Bland-Altman deviance plot (left) and Passing-Bablok regression plot (right) for progesterone calibrators for four different commercial calibrators. The line of identity is shown as a dashed black line.

reconcile with the required experimental approach to value assignment.

The first steroid immunoassays were reported in 1969–70 [7, 8] a decade after the Nobel Prize-winning invention of peptide immunoassay [9, 10]. The decade delay was due to the need for additional technical requirements for valid immunoassay of non-immunogenic steroids. These additional steps required for valid original steroid immunoassay included pre-assay solvent extraction and chromatography together with use of authentic steroid tracers, the triplet validity criteria for steroid immunoassay. Pre-assay solvent extraction reduced non-specific matrix interference effects of serum/plasma while chromatography reduces reduce cross-reactivity from steroid precursors or metabolites that are structurally related to the target steroid. Hence, while the original, in-house steroid immunoassays of the 1970s and 1980s remain valid, subsequent simplified methods, direct (non-extraction)

Table 3: Progesterone immunoassay calibrator measurements.

Calibrator	Label concentration	Replicates measured	Measured, ng/mL mean [SEM] (range)	Measured % mean (range)[a] = sign.	Claimed MU
P4 Architect Cal 1	0.7 ng/mL	6	0.58 [0.01] (0.55–0.61)	83 (78–87)	1.4 %
P4 Architect Cal 2	40 ng/mL	6	33.1[0.34] (32–34)	83 (80–86)	0.7 %
Beckman P_4 S0	0 ng/mL	6	0	100	
Beckman P4 S1	0.9 ng/mL	6	0.79 [0.02] (0.75–0.83)	88 (109–136)	3.84 %
Beckman P4 S2	3.8 ng/mL	6	3.20 [0.02] (3.14–3.28)	84 (104–117)[a]	2.86 %
Beckman P4 S3	9.3 ng/mL	6	8.76 [0.10] (8.4–9.2)	94 (99–108)	2.42 %
Beckman P4 S4	19.4 ng/mL	6	18.6 [0.15] (18.1–19)	96 (106–111)	1.88 %
Beckman P4 S5	40 ng/mL	6	37.1 [0.28] (36–38)	93 (101–109)	6.58 %
Roche P4 III Cal 1	0.20 ng/mL	6	0.12 [0.004] (0.11–0.13)	62 (97–107)[a]	11.3
Roche P4 III Cal 2	53 ng/mL	6	38.2 [0.48] (33.6–41.5)	72 (36–40)[a]	4.1 %
Siemens E Cal L	0.86 ng/mL	6	0.38 [0.01] (0.29–0.45)	46 (35–55)[a]	17.8 %
Siemens E Cal H	28.5 ng/mL	6	21.0 [0.30] (20.3–22.3)	73 (71–78)[a]	6.4 %

To convert ng/mL to nmol/L multiply by 3.18. To convert nmol/L to ng/mL divide by 3.18. MU, measurement uncertainty, coverage factor of 2. [a]Result exceeds ±10 % bias with >95 % certainty.

immunoassays, were developed to include steroids into automated multiplex platform immunoassays. This eliminated the triplet of validity criteria at the cost of assay validity. Consequently numerous studies show that direct steroid immunoassays display method-specific bias relative to mass spectrometry-based reference methods for T [11–14] as well as E_2 and P_4 [15–23]. It would therefore be surprising if direct testosterone immunoassays could meet the requirements of accuracy-based proficiency or quality control programs [24].

These discrepancies persist despite ongoing work for assay standardisation at the international Level. The Joint Committee for Traceability in Laboratory Medicine (JCTLM) has been promoting traceability since 2002 and currently lists higher-order reference materials (pure and matrix matched), reference measurement procedures and reference measurement services for all three measurands (JCTLM database. https://www.bipm.org/jctlm/ accessed 10/11/2020). Investigation of the information supplied by the manufacturers indicates that five reference methods and three reference materials are listed on the JCTLM database, however only one (Beckman-Coulter E2) specifically stated this in their product information. Additionally, the CDC HoST program also supports accuracy of steroid measurements (https://www.cdc.gov/labstandards/hs.html). However, currently the Siemens Centaur testosterone is the only sex steroid immunoassay verified in this program, a fact that may reflect the difficulty of direct steroid immunoassays with their inherent method-specific bias meeting the accuracy-based standards of the CDC HOST program.

These significant deviations of measured sex steroid concentrations from the calibrator labels identify another feature of direct (non-extraction) sex steroid immunoassays which contributes to their method-specific bias relative to MS-based reference measurements limiting their accuracy and validity. The present findings highlight the desirability to widening availability of LC-MS-based steroid measurement for clinical practice involving steroid measurements, which requires greater uptake by pathology laboratories of steroid LC-MS methodologies. The introduction of more accurate sex steroid measurements for clinical practice is hindered by the cost of expensive equipment and of skilled technicians required for LC-MS measurements, and service provision to match or exceed the high throughput, rapid result availability of current automated immunoassay platforms.

The strengths of this study include that it measured a wide range of steroid immunoassay calibrators by a validated UPLC-MS method which represents the reference level standard for steroid measurement. Our experiments excluded the possibility that matrix effects of the calibrators led to inaccuracy consistent with the fact that pre-analytical solvent extraction as standard for steroid LCMS is expected to eliminate matrix effects.

A limitation of the current study is that the effect of these calibrators on assay results was not assessed. It may be that the selection of assigned values for these materials leads to conformity of the immunoassay system. However, the need to make such post-analytical adjustments indicates likely issues with masking the analytical specificity of the assay and matrix effects. Additional limitations of this study includes that it did

not investigate all available external calibrators and some more accurate ones may have been missed.

Research funding: None declared.
Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: The authors state no conflict of interest.
Research ethics: Not applicable.
Informed consent: Not applicable.
Data availability: Not applicable.

References

1. Killeen AA, Long T, Souers R, Styer P, Ventura CB, Klee GG. Verifying performance characteristics of quantitative analytical systems: calibration verification, linearity, and analytical measurement range. Arch Pathol Lab Med 2014;138:1173–81.
2. Pum J. A practical guide to validation and verification of analytical methods in the clinical laboratory. Adv Clin Chem 2019;90:215–81.
3. Skiba MA, Bell RJ, Islam RM, Handelsman DJ, Desai R, Davis SR. Androgens during the reproductive years: what is normal for women? J Clin Endocrinol Metab 2019;104:5382–92.
4. FDA CDER. Bioanalytical method validation. Silver Springs Maryland, MD, USA: FDA CDER; 2018.
5. European Medicines Agency. Guideline on bioanalytical method validation. Amsterdam, The Netherlands: European Medicines Agency; 2011.
6. Matuszewski BK, Constanzer ML, Chavez-Eng CM. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. Anal Chem 2003;75:3019–30.
7. Abraham GE. Solid-phase radioimmunoassay of estradiol-17 beta. J Clin Endocrinol Metab 1969;29:866–70.
8. Furuyama S, Mayes DM, Nugent CA. A radioimmunoassay for plasma testosterone. Steroids 1970;16:415–28.
9. Yalow RS, Berson SA. Immunoassay of endogenous plasma insulin in man. J Clin Invest 1960;39:1157–75.
10. Yalow RS, Berson SA. Assay of plasma insulin in human subjects by immunological methods. Nature 1959;184(21 Suppl):1648–9.
11. Taieb J, Mathian B, Millot F, Patricot MC, Mathieu E, Queyrel N, et al. Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women, and children. Clin Chem 2003;49:1381–95.
12. Sikaris K, McLachlan RI, Kazlauskas R, de Kretser D, Holden CA, Handelsman DJ. Reproductive hormone reference intervals for healthy fertile young men: evaluation of automated platform assays. J Clin Endocrinol Metab 2005;90:5928–36.
13. 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;89:534–43.
14. 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; 92:405–13.
15. 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;98:1376–87.
16. Handelsman DJ, Newman JD, Jimenez M, McLachlan R, Sartorius G, Jones GR. Performance of direct estradiol immunoassays with human male serum samples. Clin Chem 2014;60:510–7.
17. Shankara-Narayana N, Zawada S, Walters KA, Desai R, Marren A, Handelsman DJ. Accuracy of a direct progesterone immunoassay. J Appl Lab Med 2016;1:294–9.
18. Coucke W, Devleeschouwer N, Libeer JC, Schiettecatte J, Martin M, Smitz J. Accuracy and reproducibility of automated estradiol-17beta and progesterone assays using native serum samples: results obtained in the Belgian external assessment scheme. Hum Reprod 2007;22:3204–9.
19. Chen Y, Kinney L, Soldin SJ. Performance evaluation of Siemens ADVIA Centaur enhanced estradiol assay and a split sample comparison with liquid chromatography-tandem mass spectrometry. Clin Biochem 2012;45:811–5.
20. Patton PE, Lim JY, Hickok LR, Kettel LM, Larson JM, Pau KY. Precision of progesterone measurements with the use of automated immunoassay analyzers and the impact on clinical decisions for in vitro fertilization. Fertil Steril 2014;101:1629–36.
21. Wang Y, Zhang T, Zhao H, Zhou W, Zeng J, Zhang J, et al. Measurement of serum progesterone by isotope dilution liquid chromatography tandem mass spectrometry: a candidate reference method and its application to evaluating immunoassays. Anal Bioanal Chem 2019;411: 2363–71.
22. Sluss PM, Hayes FJ, Adams JM, Barnes W, Williams G, Frost S, et al. Mass spectrometric and physiological validation of a sensitive, automated, direct immunoassay for serum estradiol using the Architect. Clin Chim Acta 2008;388:99–105.
23. Yang DT, Owen WE, Ramsay CS, Xie H, Roberts WL. Performance characteristics of eight estradiol immunoassays. Am J Clin Pathol 2004; 122:332–7.
24. Cao ZT, Botelho JC, Rej R, Vesper H. Accuracy-based proficiency testing for testosterone measurements with immunoassays and liquid chromatography-mass spectrometry. Clin Chim Acta 2017;469:31–6.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2022-1179).