

COMPARATIVE PERFORMANCE OF TWO IMMUNOASSAY PLATFORMS FOR PEDIATRIC HORMONE TESTING

Trinh Thi Phuong Dung^{1,2}, Nguyen Thi Bang Suong², Cao Viet Tung³, Tran Minh Dien³, Vu Thi Tu Uyen³, Nguyen Quynh Giao¹,
Nguyen Thi Hien¹ and Tran Thi Chi Mai^{1,3}

¹Faculty Medical Technology, Hanoi Medical University, Hanoi, Vietnam

²Ho Chi Minh City University of Medicine and Pharmacy, Ho Chi Minh city, Vietnam

³Vietnam National Children's Hospital, Ho Chi Minh city, Vietnam

Received: 18.06.2025.

Accepted: 01.12.2025.

Corresponding author:

Tran Thi Chi Mai, M.D., Ph.D.

Hanoi Medical University, No.1 Ton That Tung, Dong
Da, Hanoi, Vietnam

E-mail: tranchimai@hmu.edu.vn

ABSTRACT

To compare the analytical performance of the Roche Cobas Pro and Siemens Atellica Solution platforms for measuring serum luteinizing hormone (LH), follicle-stimulating hormone (FSH), estradiol, and testosterone in pediatric samples, with an emphasis on analytical bias, agreement, and clinical interpretability. A method comparison study was conducted using paired serum samples from children (LH: $n = 132$; FSH: $n = 140$; estradiol: $n = 413$; testosterone: $n = 125$). An additional 125 adult female samples were included to extend the estradiol measurement range. Hormone concentrations were measured on both platforms according to CLSI EP09-A3 guidelines. Analytical agreement and bias were evaluated using Passing–Bablok regression, Bland–Altman analysis, and Spearman correlation. Strong correlations were observed for all analytes (Spearman $r \geq 0.899$; $p < 0.0001$). However, clinically relevant biases were detected. Compared with Cobas Pro, Atellica underestimated LH (slope 0.856; intercept -0.424 IU/L) and overestimated FSH (slope 1.087). Estradiol demonstrated the greatest disagreement, with Atellica showing higher values in pediatric samples (mean bias $+24.65$ pmol/L) and adult female samples (mean bias $+142.18$ pmol/L). Among pediatric estradiol measurements, 232/413 samples were below the limit of quantification (LOQ) on one or both platforms; notably, 147 samples were undetectable by Cobas Pro but quantifiable on Atellica. Testosterone showed near-proportional agreement (slope 0.99), although non-linearity was observed at lower concentrations. Although both platforms demonstrate strong analytical correlation, measurement biases, particularly for LH and estradiol, are of sufficient magnitude to influence clinical interpretation in pediatric endocrinology, including puberty assessment and longitudinal hormone monitoring.

Keywords: Pediatric endocrinology, Immunoassay comparison, Luteinizing hormone, Estradiol; Testosterone, Cobas Pro, Atellica Solution.

UDK: 616.43-008.6-053.2
577.175.6

Eabr 2025; 26(4):331-337

DOI: 10.2478/eabr-2025-0021

INTRODUCTION

Accurate measurement of gonadotropins and sex steroids—specifically luteinizing hormone (LH), follicle-stimulating hormone (FSH), estradiol, and testosterone—is essential for the diagnosis and management of pediatric endocrine disorders. These hormones play a central role in sexual development, growth, and reproductive health, and their levels vary substantially with age and sex. Interpreting hormone levels in children and adolescents therefore requires not only high analytical precision but also age- and sex-specific reference intervals to support accurate clinical decision-making (1, 2). In Vietnam, pediatric endocrine services have expanded significantly over the past decade, yet limited access to LC-MS/MS and lack of locally validated immunoassay performance data remain key challenges. Clinical laboratories continue to rely on high-throughput immunoassay platforms, but their accuracy for pediatric hormone quantification has not been systematically evaluated in the Vietnamese setting. This gap raises concerns about potential misclassification, especially in prepubertal children with low hormone levels near the detection limits of conventional assays.

Pediatric hormone testing presents unique analytical challenges due to the naturally low concentrations of sex steroids and gonadotropins in prepubertal and early pubertal children. For instance, serum estradiol levels in young girls and testosterone levels in young boys often fall below 20 pg/mL and 30 ng/dL, respectively—concentrations near or below the functional sensitivity of many commercial immunoassays (3–5). These low levels increase the risk of imprecision, cross-reactivity, and false elevation, leading to potential misdiagnosis or inappropriate treatment.

Among the most commonly used immunoassay platforms in clinical laboratories are the Roche Cobas Pro and Siemens Atellica Solution systems. Both are high-throughput, fully automated analyzers designed for the measurement of a wide range of hormones and clinical chemistry parameters. Previous studies have demonstrated that both systems provide reliable and reproducible measurements of LH and FSH, with well-established pediatric reference intervals that align with international standards (1, 2). However, questions remain about their performance in detecting low hormone concentrations, especially in prepubertal children.

Immunoassays, despite being widely available and cost-effective, are prone to limitations such as cross-reactivity, matrix effects, and reduced specificity at low concentrations. These challenges are especially pronounced when measuring sex steroids in children. Studies have shown that immunoassays may overestimate estradiol and testosterone concentrations when compared to more specific methods like liquid chromatography–tandem mass spectrometry (LC-MS/MS), which is considered the gold standard for hormone quantification (3–5). Such discrepancies may result in diagnostic inaccuracies and suboptimal clinical management.

While LC-MS/MS offers superior analytical specificity and sensitivity, its adoption in routine clinical practice remains limited due to high costs, technical complexity, and the need for specialized personnel. Consequently, immunoassays remain the mainstay of hormone testing in most pediatric settings. It is therefore critical to evaluate the analytical performance and limitations of widely used platforms, particularly in populations where hormone concentrations are near the assay detection limits (6, 7).

The aim of this study was to compare the performance of the Roche Cobas Pro and Siemens Atellica Solution platforms in measuring serum LH, FSH, estradiol, and testosterone in pediatric samples. By assessing analytical agreement and identifying potential biases between platforms, we seek to provide evidence to inform assay selection, guide appropriate clinical interpretation, and support the development of platform-specific pediatric reference intervals in Vietnam.

MATERIALS AND METHODS

Study design, sample size and sampling procedure

This method comparison study was conducted in accordance with the Clinical and Laboratory Standards Institute (CLSI) EP09-A3 guidelines. Paired serum samples were collected from pediatric patients at the Vietnam National Children's Hospital between June 2024 and January 2025. The study included 132 samples for LH, 140 for FSH, 413 for estradiol, and 125 for testosterone. To ensure adequate coverage of the analytical measurement range for estradiol, an additional 125 adult female samples were included. All samples were anonymized and obtained from routine testing; thus, informed consent was waived. Ethical approval was granted by the Institutional Review Board of Vietnam National Children's Hospital (IRB-VN01037/IRB00011976/FWA00028418). This study is part of a broader project aimed at establishing pediatric reference intervals for Vietnamese children.

Sample handling and hormone measurement

Serum samples were excluded if visibly hemolyzed, icteric, or lipemic. Following centrifugation, serum was aliquoted into Eppendorf tubes and stored at -80°C until analysis. All hormone assays were performed on the Roche Cobas Pro and Siemens Atellica Solution platforms within two hours of thawing. Calibration and internal quality control procedures followed the manufacturers' instructions. Daily quality control was performed using Bio-Rad control materials.

Analytical characteristics of each assay—such as measurement principle, traceability, and analytical range—are summarized in Table 1. Detailed calibration procedures and traceability chains have been presented in Supplementary Table S1 to improve clarity and focus. Both platforms use electrochemiluminescence or chemiluminescence immunoassay technologies with standardized calibration materials.

Table 1. Summary of analytical principles, traceability, and analytical ranges of Cobas Pro (Roche) and Atellica Solution (Siemens) platforms

Measurand	Platform	Method principle	Traceability	Analytical range
LH	Atellica	CLIA (sandwich)	NIBSC 80/552	0.07–200 IU/L
LH	Cobas Pro	ECLIA (sandwich)	NIBSC 80/552	0.3–200 IU/L
FSH	Atellica	CLIA	WHO IS 94/632	0.3–200 IU/L
FSH	Cobas Pro	ECLIA	WHO IRP 78/549	0.3–200 IU/L
Testosterone	Atellica	CLIA	ID-LC-MS/MS	0.24–52.05 nmol/L
Testosterone	Cobas Pro	ECLIA	ID-GC/MS	0.087–52.0 nmol/L
Estradiol	Atellica	CLIA	ID-GC/MS (manufacturer)	43.31–11010 pmol/L
Estradiol	Cobas Pro	ECLIA	CRM 6004a via ID-GC/MS	18.4–11010 pmol/L

Precision and trueness verification and monitoring

Precision was evaluated following CLSI EP15-A3 guidelines. Short-term precision was assessed by testing three concentration levels (low, normal, high) five times per day over five consecutive days. Long-term precision was monitored through daily quality control over the course of the study. Trueness was verified via interlaboratory comparison using Bio-Rad materials and peer-group mean comparisons. External quality assurance (EQA) was conducted monthly through participation in the Randox RIQAS Chemistry and Immunoassay program, with satisfactory performance across all assays.

Statistical analysis

Data were analyzed using MedCalc version 23.2.1 (MedCalc Software Ltd., Ostend, Belgium). Method comparison was assessed using Passing–Bablok regression to evaluate systematic and proportional bias. Non-linearity was formally assessed using the cumulative sum (Cusum) test provided within the Passing–Bablok regression framework. Agreement between platforms was visualized using Bland–Altman plots. All regression analyses and graphical outputs were generated using MedCalc version 23.2.1.

Spearman's rank correlation coefficients were used to assess correlation. A p-value of <0.05 was considered statistically significant.

For analytes with results below the limit of quantification (LOQ), values were excluded from quantitative regression and Bland–Altman analyses rather than imputed. The number and distribution of below-LOQ results for each platform were reported separately to preserve analytical validity and avoid bias from arbitrary substitution.

RESULTS

Key analytical agreement metrics for all hormones (correlation coefficients, Passing–Bablok slopes and intercepts, mean biases, and limits of agreement) are summarized in Table 2 to facilitate cross-analyte comparison, while detailed visual assessments are retained in Figures 1–5.

Table 2. Key statistical results across hormones

Hormone	Spearman r	Intercept (95% CI)	Slope (95% CI)	Mean Bias (95% CI)	Cusum Test	Clinical Notes
LH	0.991	−0.42 (−0.50 to −0.35)	0.86 (0.84 to 0.87)	−1.55 IU/L (−1.75 to −1.36)	p < 0.01	Underestimation on Atellica may shift prepubertal values into the prepubertal range
FSH	0.995	0.01 (−0.09 to 0.07)	1.09 (1.07 to 1.10)	+1.35 IU/L (0.92 to 1.78)	p = 0.87	Proportional bias; use caution when switching platforms

Hormone	Spearman r	Intercept (95% CI)	Slope (95% CI)	Mean Bias (95% CI)	Cusum Test	Clinical Notes
Estradiol (peds)	0.898	33.29 (26.55 to 39.02)	0.95 (0.91 to 0.97)	+33.75 pmol/L (16.63 to 50.88)	p = 0.11	Atellica quantifies more low-end results; may influence pubertal staging
Estradiol (adult)	0.968	22.30 (13.66 to 29.11)	1.06 (1.04 to 1.08)	+142.18 pmol/L (97.91 to 186.46)	p = 0.05	Substantial overestimation on Atellica
Testosterone	0.982	-0.27 (-0.39 to -0.15)	0.99 (0.97 to 1.02)	Minimal	p = 0.01	Good agreement; non-linearity at low concentrations

Luteinizing hormone (LH)

Serum LH measurements showed a strong correlation between the Siemens Atellica Solution and Roche Cobas Pro platforms (Spearman r = 0.991, p < 0.0001), but both systematic and proportional biases were present (Figure 1). Passing–Bablok regression demonstrated an intercept of -0.4245 (95% CI: -0.5021 to -0.3488) and a slope of 0.8558 (95% CI: 0.8390 to 0.8713), indicating consistent underestimation by Atellica at increasing concentrations. The Cusum test confirmed significant non-linearity (p < 0.01).

Bland–Altman analysis revealed a mean bias of -1.55 IU/L (95% CI: -1.75 to -1.36), with 95% limits of agreement from -3.77 to +0.67 IU/L (Figure 1). In pediatric practice, this degree of negative bias may result in misclassification of borderline LH values, potentially shifting results from early pubertal into the prepubertal range, particularly when using diagnostic cut-offs for puberty onset. Therefore, the two platforms should not be considered interchangeable for LH measurement.

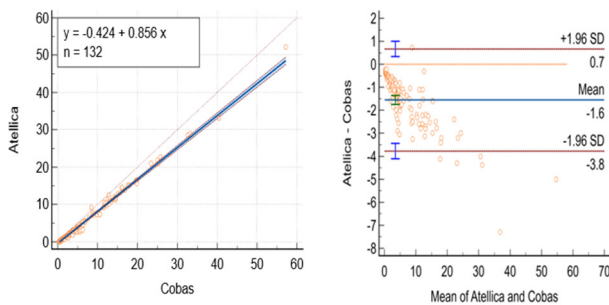


Figure 1. Passing–Bablok regression and Bland-Altman analysis comparing LH concentrations between Atellica Solution and Cobas Pro platforms using 132 pediatric serum samples.

Follicle-stimulating hormone (FSH)

FSH measurements demonstrated excellent correlation (Spearman r = 0.995, p < 0.0001) and no significant systematic bias (intercept: 0.0079; 95% CI: -0.0859 to 0.0737), although a proportional bias was observed (slope: 1.0869; 95% CI: 1.0721 to 1.1024) (Figure 2). The Cusum test did not

detect non-linearity (p = 0.87). Bland–Altman analysis showed a mean bias of +1.35 IU/L (95% CI: 0.92 to 1.78), with 95% limits of agreement from -3.70 to +6.40 IU/L, indicating acceptable agreement for cross-sectional use but caution for longitudinal follow-up using mixed platforms.

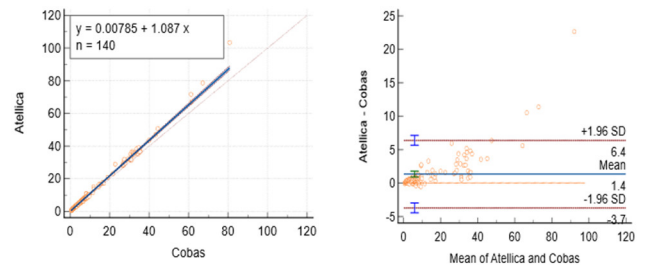


Figure 2. Comparison of FSH concentrations between Atellica Solution and Cobas Pro platforms using 140 pediatric serum samples.

Estradiol – pediatric samples

Among 413 pediatric samples, 232 results were below the limit of quantification (LOQ) on one or both platforms, leaving 181 samples for quantitative comparison. Of the low-level results, 147 were measurable on Atellica but below LOQ on Cobas, reflecting greater analytical sensitivity of Atellica at very low estradiol concentrations (Supplemental Table 1).

For quantifiable samples, estradiol concentrations correlated well between platforms (Spearman r = 0.898, p < 0.0001), but a clear systematic bias was present (Figure 3). Passing–Bablok regression yielded an intercept of 33.29 pmol/L (95% CI: 26.55 to 39.02) and a slope of 0.945 (95% CI: 0.912 to 0.974). Bland–Altman analysis demonstrated a mean bias of +33.75 pmol/L (95% CI: 16.63 to 50.88), with wide limits of agreement (-195.09 to +262.60 pmol/L). Such variability is clinically relevant in early puberty, where estradiol concentrations approach assay detection limits and small absolute differences may influence pubertal staging.

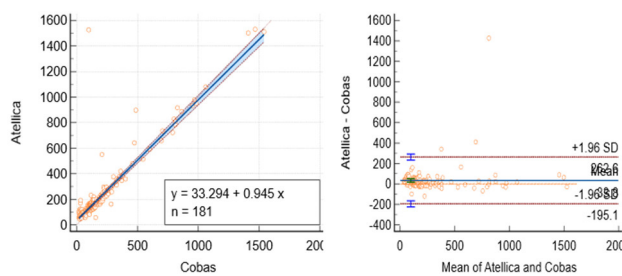


Figure 3. Comparison of estradiol concentrations between Atellica Solution and Cobas Pro platforms in 181 pediatric serum samples.

Estradiol – adult female samples

In adult female samples, estradiol measurements remained strongly correlated (Spearman $r = 0.968$, $p < 0.0001$) but showed substantial positive bias on Atellica (Figure 4). Passing–Bablok regression indicated systematic overestimation (slope: 1.063; intercept: 22.30 pmol/L), and Bland–Altman analysis revealed a mean bias of +142.18 pmol/L (95% CI: 97.91 to 186.46), with asymmetric limits of agreement (–47.99 to +632.36 pmol/L). Borderline non-linearity was observed (Cusum $p = 0.05$), suggesting reduced interchangeability at higher estradiol ranges.

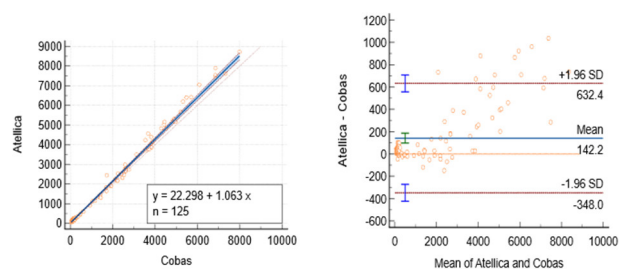


Figure 4. Comparison of estradiol concentrations in 125 adult female samples measured by Atellica Solution and Cobas Pro.

Testosterone

Testosterone concentrations showed excellent agreement between platforms (Spearman $r = 0.982$, $p < 0.0001$) with near-unity proportionality (slope: 0.9901; intercept: –0.2723) (Figure 5). Despite minimal overall bias, the Cusum test indicated significant non-linearity ($p = 0.01$), particularly at lower concentrations. Although agreement is generally acceptable, careful interpretation is advised during longitudinal monitoring in pediatric patients, where testosterone levels are often near the lower analytical range.

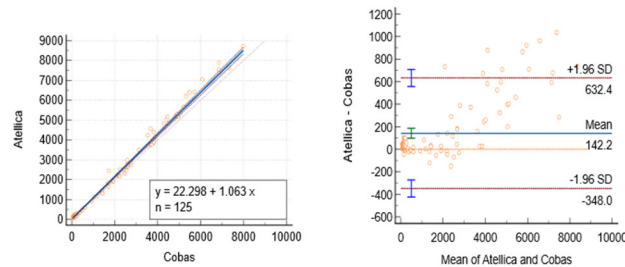


Figure 5. Comparison of testosterone concentrations between Atellica Solution and Cobas Pro platforms using 125 pediatric and adolescent serum samples.

DISCUSSION

This study evaluated the analytical comparability of the Roche Cobas Pro and Siemens Atellica Solution platforms for measuring serum LH, FSH, estradiol, and testosterone in pediatric samples. Although strong correlations were observed across all analytes, clinically meaningful systematic and proportional biases, particularly for LH and estradiol, were identified, indicating that results obtained from the two platforms are not directly interchangeable. These findings are consistent with previous reports highlighting inherent limitations of immunoassays for hormone quantification in children and reinforce the importance of assay-specific interpretation in pediatric endocrinology.

For LH, the consistent underestimation observed on the Atellica platform relative to Cobas Pro has direct implications for clinical decision-making. Diagnostic cut-offs used to distinguish prepubertal from pubertal status, either in basal assessment or during GnRH stimulation testing, are often narrow. A mean bias of approximately –1.5 IU/L may therefore shift borderline values below clinically relevant thresholds, potentially delaying recognition of pubertal onset or altering the classification of central versus peripheral pubertal disorders. Similar inter-platform variability has been reported in pediatric cohorts using Cobas-based reference intervals in the CALIPER study and other comparative analyses (1, 2).

FSH measurements showed strong agreement and retained linearity, although a proportional overestimation of approximately 9% was observed on Atellica. While this bias may be acceptable for cross-sectional interpretation, small absolute differences in FSH during early puberty may still influence Tanner staging or GnRH test interpretation, especially during longitudinal follow-up. Comparable degrees of assay-dependent variability for gonadotropins have been described in pediatric immunoassay comparison studies (1, 8).

Estradiol demonstrated the greatest analytical divergence between platforms, particularly at low concentrations typical of prepubertal children. The higher analytical sensitivity of Atellica allowed quantification of a substantial number of

samples that were below the LOQ on Cobas Pro; however, this was accompanied by a significant positive bias and wide limits of agreement. These findings align with published immunoassay–LC-MS/MS comparison studies reporting up to twofold overestimation of estradiol at low concentrations in prepubertal girls (4, 5, 9, 10). In clinical practice, such bias may influence assessment of early pubertal development, bone maturation, or decisions regarding pubertal suppression, underscoring the need for caution when interpreting low-level estradiol results.

Testosterone measurements exhibited the closest agreement between platforms, consistent with previous method comparison studies (6, 11). Nevertheless, the presence of non-linearity at low concentrations highlights persistent analytical challenges in pediatric testing, where prepubertal testosterone levels frequently approach assay detection limits. Even minor analytical deviations in this range may result in disproportionate clinical impact, a phenomenon well described in immunoassay versus LC-MS/MS comparisons (3).

From a technical perspective, differences in calibration traceability, antibody specificity, and susceptibility to cross-reactivity are likely contributors to the observed biases. Although both platforms are traceable to recognized reference materials, manufacturers employ different calibration hierarchies and antibody designs, which may variably detect hormone isoforms, metabolites, or structurally related compounds. These effects are amplified at low concentrations and in complex pediatric serum matrices, particularly for estradiol and testosterone.

Overall, these findings emphasize that while automated immunoassays remain indispensable in routine pediatric practice, their clinical interpretation must be platform-specific. Inadvertent switching of analytical platforms during follow-up may introduce artificial trends that mimic true biological change and lead to misinterpretation of disease progression or treatment response (1, 2).

LIMITATIONS

Several limitations should be acknowledged. First, LC-MS/MS was not included as an external reference method, limiting assessment of absolute analytical accuracy, particularly for estradiol and testosterone. Future validation against LC-MS/MS would strengthen confidence in platform-specific performance. Second, the study population was derived from a single center, which may limit generalizability to other settings or ethnic groups. Third, within-run and between-run reproducibility between platforms was not directly compared, which may influence clinical reliability during serial monitoring. Fourth, differences in sample matrix (pediatric serum versus adult female serum) may also affect assay comparability, particularly for estradiol, and should be interpreted cautiously. Finally, potential interferences such as heterophilic antibodies or biotin were not systematically evaluated and may contribute to variability in real-world clinical samples.

CONCLUSION

In conclusion, although both the Roche Cobas Pro and Siemens Atellica Solution platforms provide high-throughput hormone measurements suitable for pediatric endocrine testing, their results are not interchangeable, particularly for LH and estradiol. The magnitude of observed bias is sufficient to influence clinical classification in pediatric puberty assessment. Clinicians are advised to avoid cross-platform comparison during longitudinal follow-up and to apply platform-specific pediatric reference intervals consistently. Rather than relying solely on method harmonization at the calibration level, these findings highlight the need for harmonized, platform-specific pediatric reference standards supported by clinical outcome data. This study provides a foundation for developing standardized, platform-aware approaches to pediatric hormone interpretation and supports future external validation using LC-MS/MS to improve diagnostic accuracy and clinical confidence.

REFERENCES

1. Mary Kathryn B, Victoria H, Shervin A, Felix L, Felix L, Barry H, et al. Paediatric reference intervals for 17 Roche cobas 8000 e602 immunoassays in the CALIPER cohort of healthy children and adolescents. *Clinical Chemistry and Laboratory Medicine*. 2019;57(12):1968-79.
2. Mary Kathryn B, Paul SH, Donna L, Paul S, Alexandra H, Khosrow A. Pediatric reference intervals for endocrine markers and fertility hormones in healthy children and adolescents on the Siemens Healthineers Atellica immunoassay system. *Clinical Chemistry and Laboratory Medicine*. 2021;59(8):1421-30.
3. Joëlle T, Bruno M, Françoise M, Marie-Claude P, Elisabeth M, Nicole Q, et al. Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women, and children. *Clinical Chemistry*. 2003;49(8):1381-95.
4. Ashley Di M, Mehrdad Y, Victoria H, Matthew MK, Mary CN, Mary Kathryn B, et al. Highly sensitive tandem mass spectrometric measurement of serum estradiol without derivatization and pediatric reference intervals in children and adolescents. *Clinical chemistry and laboratory medicine*. 2023;0(0).
5. Carina A-L, Charlotte B, Emilia S, Henrik R. Methodological considerations in determining sex steroids in children: comparison of conventional immunoassays with liquid chromatography-tandem mass spectrometry. *Clinical Chemistry and Laboratory Medicine*. 2023;62:85-96.
6. Stefan AW, Ulrich W, Homoki J, Walter MT. 17 α -Hydroxyprogesterone, 4-Androstenedione, and Testosterone Profiled by Routine Stable Isotope Dilution/Gas Chromatography-Mass Spectrometry in Plasma of Children. *Pediatric Research*. 1995;38(1):76-80.
7. Alexandra K, Maik W, Paul Martin H, Felix GR. Principles and clinical applications of liquid chromatography — tandem mass spectrometry for the determination of adrenal and gonadal steroid hormones. *Journal of Endocrinological Investigation*. 2011;34(9):702-8.

8. David JH, Graham RDJ, Kouzios D, Reena D. Evaluation of testosterone, estradiol and progesterone immunoassay calibrators by liquid chromatography mass spectrometry. *Clinical chemistry and laboratory medicine*. 2023;0(0).
9. Leticia Ribeiro O, Leticia Ribeiro O, Carlos Alberto L, Guilherme G-F, Guilherme G-F, Jose Luiz C, et al. Androgens by immunoassay and mass spectrometry in children with 46,XY disorder of sex development. *Endocrine connections*. 2020;9(11):1085-94.
10. Valentin B, Hermann S, Lorenz R, Christoph S. Design and Validation of a Sensitive Multisteroid LC-MS/MS Assay for the Routine Clinical Use: One-Step Sample Preparation with Phospholipid Removal and Comparison to Immunoassays. *International Journal of Molecular Sciences*. 2022;23(23):14691-.
11. Faye BV, Frederick AS, Rafael S, Sihe W. Measurement of serum testosterone using high-performance liquid chromatography/tandem mass spectrometry. *Clinical Chemistry and Laboratory Medicine*. 2006;44(1):70-5.