Inaccurate First-Generation Testosterone Assays Are Influenced by Sex Hormone–Binding Globulin Concentrations

Annemieke C. Heijboer,1* Edo Savelkoul,1 Adrian Kruit,2 Erik Endert,3 and Marinus A. Blankenstein1

Background: The quality of testosterone assays has been a matter of debate for several years. Known limitations of testosterone immunoassays are the cross-reactivity with other steroids and a high variation in the low concentration range. We hypothesized that one of the additional limitations of testosterone immunoassays is an ineffective displacement of testosterone from its binding protein.

Methods: Thirty samples from women not using oral contraceptives (OAC), 30 samples from women using OAC, and 30 samples from pregnant women were used to measure testosterone by an isotope dilution (ID)-LC-MS/MS method and by 6 commercially available testosterone immunoassays (UniCel®, ARCHITECT®, Centaur®, Cobas®, Immulite®, and Liaison®). In addition, sex hormone–binding globulin (SHBG)4 was measured by immunoassay (ARCHITECT).

Results: The first-generation immunoassays (UniCel, Centaur, Immulite, and Liaison) showed inaccurate testosterone results in the method comparisons with the ID-LC-MS/MS method (R between 0.61 and 0.86) and for some assays (UniCel and Liaison) also a very poor standardization (slopes of 0.59 and 0.67, respectively). On average, SHBG concentrations were lowest in women not using OAC and highest in pregnant women, and overall ranged from 18.5 to 633 nmol/L. In the first-generation immunoassays, but not in the second-generation immunoassays, we observed an inverse relationship between SHBG concentrations and deviations in testosterone from the ID-LC-MS/MS results.

Conclusions: Widely used first-generation testosterone immunoassays are influenced by SHBG concentrations, which lead to inaccurate results in samples from patients with high or low SHBG concentrations, respectively. Laboratory specialists, clinicians, and researchers should be aware of this limitation in testosterone assays.

1Department of Clinical Chemistry, Endocrine Laboratory, VU University Medical Center, Amsterdam, the Netherlands; 2Medical Laboratory, Nij Smellinghe Hospital, Drachten, the Netherlands; 3Department of Clinical Chemistry, Laboratory of Endocrinology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands

*Address correspondence to this author at: Department of Clinical Chemistry, VU University Medical Center, PO Box 7057, 1007 MB, Amsterdam, the Netherlands. Fax +31-20-444-3895; e-mail a.heijboer@vumc.nl.

DOI: 10.1373/jalm.2016.020065

© 2016 American Association for Clinical Chemistry

4 Nonstandard abbreviations: SHBG, sex hormone–binding globulin; OAC, oral contraceptives; ID, isotope dilution; LLOQ, lower limit of quantification; IQR, interquartile range.
IMPACT STATEMENT

Female patients on oral contraceptives and pregnant women will benefit from the information presented here. Evidence presented on the inaccuracy of testosterone immunoassays will allow better characterization of hypergonadism in women. Knowledge in the field of endocrinology will be advanced by the information presented.

For years now, the quality of testosterone assays is a matter of debate. In fact, some testosterone assays predict no more accurately than taking a guess (7). The limitations and pitfalls of testosterone immunoassays have led to a position statement from the Endocrine Society (2). Consequently, improvements have been made by some manufacturers by developing second-generation assays. Notwithstanding these urgent calls, other manufacturers still sell, and laboratories still use inaccurate testosterone assays. Known limitations of these inaccurate assays are the cross-reactivity with other steroids and a high variation in the low concentration range (3, 4). Both limitations lead to inaccurately reported testosterone concentrations particularly in samples from women and children and can have serious clinical outcomes.

As testosterone is measured as total testosterone, a major challenge for testosterone assays, especially in view of the short incubation times, is the displacement of testosterone from sex hormone–binding globulin (SHBG). Solvents leading to a complete release of testosterone from SHBG are often not compatible with automated immunoassays. Alternatives might not be sufficiently effective for displacement. The improper release of a hormone from its binding protein has already been identified as one of the limitations of automated 25-OH vitamin D immunoassays (5). In one instance, a commercially available manual testosterone radioimmunoassay, which was also unable to effectively release testosterone from SHBG, led to false conclusions in a clinical trial (6).

Whether currently available automated testosterone immunoassays are sufficiently effective in liberating testosterone from SHBG to accurately measure total testosterone concentrations is not known. Therefore, the goal of our study was to study the influence of SHBG concentration on the performance of 6 commonly used automated testosterone immunoassays.

MATERIALS AND METHODS

Samples

Serum was obtained by drawing an extra tube (serum separating tube) of blood from patients who already underwent a venipuncture for diagnostic purposes in our outpatient clinic (VU University Medical Center, Amsterdam). This study was approved by the local medical ethical committee. After obtaining informed consent, at a random time, blood was drawn from 30 women who were not using oral contraceptives (OAC) or who were not pregnant (age range 24–45 years), from 30 women using OAC (age range 20–45 years), and from 30 pregnant women (age range 24–43 years and pregnancy term range 5–39 weeks). All samples were anonymized immediately after withdrawal and handled identically. After centrifugation, serum was separated, aliquoted, and frozen at −20 °C until analyses. Storage time did not exceed 5 months, taking into account the stability.

Methods

Isotope dilution LC-MS/MS method. In all serum samples, total testosterone was measured using an isotope dilution (ID)-LC-MS/MS method, as de-
scribed earlier (7). In short, the serum samples underwent a liquid–liquid extraction using hexane ether (4:1) after addition of internal standards: [13C3]-testosterone (Ceriliant) and [13C3]-androstenedione (Ceriliant), before injection on an Acquity 2D-UPLC system coupled with a Xevo TQ-S tandem mass spectrometer (Waters). This method is very accurate, sensitive, and well standardized, as was shown in previous studies (7–9). Intraassay variation in both the female and male concentration range was 4%. Analyses were performed in the Endocrine Laboratory of the VU University Medical Center.

**Serum total testosterone immunoassays.** Total testosterone in all serum samples was measured using 6 automated immunoassays: UniCel® DxI 600 (Beckman Coulter, lot no. 436922), ARCHITECT® i2000 (Abbott Diagnostics, lot no. 10405UP00), Centaur® XP (Siemens Diagnostics, lot no. 036176), Cobas® 6000 (Roche Diagnostics, lot no. 18351201), Immulite® 2000 (Siemens Diagnostics, lot no. 505) and Liaison® (Diasorin, lot no. 131680). For lack of formal terminology to identify the different iterations of assays on the market, the UniCel, Centaur, Immulite, and Liaison methods are referred to as “first-generation” testosterone assays respectively. The lower limit of quantification (LLOQ) of the methods is shown in Table 1. Analyses using the Immulite 2000 were performed in the Laboratory of Endocrinology of the Academic Medical Center of the University of Amsterdam, analyses using the UniCel DxI 600 were performed in the Medical Laboratory of the Nij Smellinghe Hospital, and analyses using the ARCHITECT, i2000, Centaur XP, Cobas 6000, and Liaison were performed at the VU University Medical Center.

**Serum SHBG assay.** In all serum samples, SHBG was measured using the ARCHITECT i2000 platform (Abbott Diagnostics, lot no. 00614L00) (10). Intraassay variation was below 6% over the whole concentrations range.

**Statistics**

All serum total testosterone concentrations measured using the different immunoassays were compared to the concentrations measured using the ID-LC-MS/MS methods. We used Passing–Bablok regression analysis and calculated Pearson correlation coefficients for both the method comparisons, and the correlation between the SHBG concentration and the deviation of the 6 automated immunoassays with the LC-MS/MS method. The Cusum test for linearity was used to determine whether the method

---

<table>
<thead>
<tr>
<th>Table 1. LLOQ: the slope and intercept derived from Passing-Bablok regression analysis, the correlation coefficient ($R$), all compared with the ID-LC-MS/MS method, based on N samples per method.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
</tr>
<tr>
<td>First-generation assays</td>
</tr>
<tr>
<td>UniCel*</td>
</tr>
<tr>
<td>Centaur*</td>
</tr>
<tr>
<td>Immulite</td>
</tr>
<tr>
<td>Liaison*</td>
</tr>
<tr>
<td>Second-generation assays</td>
</tr>
<tr>
<td>ARCHITECT</td>
</tr>
<tr>
<td>Cobas</td>
</tr>
</tbody>
</table>

To convert testosterone concentrations to ng/mL, multiply by 0.3.
* Significant deviation from linearity.
comparison was linear. All statistical analyses were performed using Medcalc (version 11.6, Medcalc Software). P < 0.05 was considered statistically significant.

RESULTS

One of the 90 serum samples showed a testosterone concentration below the LLOQ in all methods including the ID-LC-MS/MS method (LLOQ 0.10 nmol/L; to convert testosterone concentrations to ng/mL, multiply by 0.3). Another serum sample measured a testosterone concentration of 15.9 nmol/L using ID-LC-MS/MS (7.5 nmol/L by UniCel, 15.1 nmol/L by Centaur, 12.4 nmol/L by Immulite, 6.2 nmol/L by Liaison, 15.8 nmol/L by ARCHITECT, and 13.7 nmol/L by Cobas). This concentration is exceptionally high for women, even during pregnancy. It may therefore be either from a man, a transgender person, or a woman with an unspecified adrenal or ovarian disease. Because the samples were anonymized, no further action was taken to investigate the cause of this high concentration. Both samples were considered outliers and were therefore excluded from statistical analysis. In a number of samples, many different immunoassays measured testosterone concentrations below the LLOQ. For this reason, the samples that were compared varied in number. The total number used for each method is summarized in Table 1.

Table 1 shows the slope and intercept derived from the Passing–Bablok analysis and Pearson correlation coefficient (r). In Fig. 1, the Passing–Bablok regression is shown. According to the Cusum test for linearity, testosterone results from UniCel, Centaur, and Liaison did not show a linear relationship with the LC-MS/MS results.

SHBG concentrations were significantly higher in women using OAC [median and interquartile range (IQR): 117 nmol/L (73–177)] and pregnant women [381 nmol/L (210–518)] compared to women not using OAC [62 nmol/L (36–93)] (see Fig. 2). Pregnant women had significantly higher SHBG concentrations compared to women using OAC.

Testosterone concentrations were significantly higher in women using OAC (median and IQR) [0.45 nmol/L (0.29–0.72)] and pregnant women [1.96 nmol/L (1.31–3.17)] compared to women not using OAC [0.70 nmol/L (0.54–0.98)] (see Fig. 2). Pregnant women had significantly higher concentrations than women using OAC.

The correlation and correlation coefficient between the SHBG concentration and the deviation of each automated immunoassay from the ID-LC-MS/MS method are shown in Fig. 3. In the first-generation immunoassays, but not in the second-generation immunoassays, we observed a statistically significant inverse relationship between SHBG concentrations and deviations in testosterone from the ID-LC-MS/MS results.

DISCUSSION AND CONCLUSION

In this study, we showed the limitations of some of the currently commercially available automated immunoassays. Standardization problems in combination with a high inaccuracy were observed. The first-generation immunoassays (UniCel, Centaur, Immulite, and Liaison) showed inaccurate results expressed as low to very low correlation coefficients when compared to our ID-LC-MS/MS method (r between 0.61 and 0.86) and for some assays (UniCel and Liaison) a poor standardization (slopes of 0.59 and 0.67, respectively). The correlation coefficients are comparable with earlier method comparison studies in women. Since then, however, the standardization seems to have changed (3, 4). Concerted efforts in standardization of testosterone assays in recent years, in combination with the testosterone standardization program of the CDC (17), might be the reason for these apparent changes. Our study clearly shows that the second-generation immunoassays (ARCHITECT and Cobas) have a better standardization (slopes of 0.99 and 1.01, respectively) and less variation (r of 0.95 and 0.93, respectively).
Fig. 1. Passing–Bablok regression analysis and Bland–Altman plots for the 6 automated immunoassays. UniCel, Centaur, Cobas, and Liaison showed a significant deviation from linearity. In the Passing–Bablok figures: on the x axis, the testosterone concentrations were measured using ID-LC-MS/MS and, on the y axis, the testosterone concentrations using the respective immunoassays are shown. For the Bland–Altman plots: on the x axis, the testosterone concentrations were measured using ID-LC-MS/MS and, on the y axis, the % deviation of the respective immunoassays compared to the ID-LC-MS/MS assay are shown. (A and B), UniCel; (C and D), Centaur; (E and F), Immulite; (G and H), Liaison; (I and J), ARCHITECT; (K and L), Cobas. To convert testosterone concentrations to ng/mL, multiply by 0.3.
This improvement is in accordance with previous studies (3, 12).

We hypothesized that one of the additional limitations of the commercially available testosterone immunoassays is the ineffective displacement of testosterone from its binding protein. According to Fig. 3, as shown by statistically significant inverse relationship between SHBG concentrations and deviations in testosterone from the ID-LC-MS/MS results, our study strongly suggests that all 4 first-generation testosterone immunoassays are incapable of completely releasing testosterone from its binding proteins. This may lead to false low results in patients with high SHBG concentrations and to false high results in patients with low SHBG concentrations, respectively. In an earlier study, we showed that using another immunoassay, which also suffered from this displacement issue, led to false conclusions in a clinical trial (6). This example illustrates the clinical relevance of this displacement issue.

Based on our study, we concluded that not all automated testosterone assays are suitable to measure testosterone in patients with either high or low SHBG concentrations. In this study, we showed that pregnant women and women using OAC have higher SHBG concentrations (6- and 2-fold, respectively). This confirms the results from previous studies, showing a 5- to 10-fold increase during pregnancy (13, 14) and a 2- to 4-fold increase in women using OAC (15, 16), which is due to estrogens stimulating the liver to produce SHBG. Increased serum SHBG concentrations are also seen in samples from patients with anorexia nervosa, hyperthyroidism, and Cushing disease (17). Patients with obesity, polycystic ovary syndrome, and metabolic syndrome all have lower SHBG concentration, which is probably due to hepatic fat content inhibiting SHBG production (17).

All assays evaluated in our study have their own reference intervals and although applying assay-specific reference values would be a fitting approach to the differences in standardization, it would neither resolve the poor correlation nor the influence of SHBG concentration that we observed.

Despite the profound and clinically relevant limitations of the first-generation testosterone immunoassays, these are still commonly used worldwide. In 2015, in the German external quality assessment scheme for testosterone, more than 40% of the 700+ participants measured tes-
Testosterone with one of the first-generation testosterone immunoassays included in our study.

In the College of American Pathologists survey, more than 50% of the 1400+ participants measured testosterone with one of the first-generation testosterone immunoassays. In the UK National and Dutch external quality assessment schemes, approximately 30% of the participants measured testosterone with one of the first-generation testosterone immunoassays.

The 2 second-generation testosterone assays tested here were not influenced by the SHBG concentrations, which makes these particular assays more suitable for clinical use and a step forward in the accurate determination of serum testosterone. Unfortunately, these assays still have their drawbacks, as they seem to suffer from cross-reactivity with other steroids, albeit less severe than the first-generation testosterone assays (3). For this reason, interpretation of testosterone measurements based on second-generation immunoassays may continue to be a challenge in specific patient groups, such as neonates in whom high concentrations of several uncommon steroid hormones may cross-react in the immunoassays.

In conclusion, first-generation testosterone immunoassays are influenced by SHBG concentrations, which lead to inaccurate results in samples from subjects with high or low SHBG concentrations, respectively. Laboratory specialists, clinicians and researchers should be aware of this limitation in first-generation testosterone immunoassays, which are still used worldwide, and should not use these assays in the low concentration range.

Fig. 3. Correlation between the SHBG concentration and the difference between the results of each of the 6 automated immunoassays and the ID-LC-MS/MS method. (A), UniCel; (B), Centaur; (C), Immulite; (D), Liaison; (E), ARCHITECT; (F), Cobas. To convert testosterone concentrations to ng/mL, multiply by 0.3.
First-Generation Testosterone Assays Are Influenced by SHBG

**Author Contributions:** All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

**Authors’ Disclosures or Potential Conflicts of Interest:** Upon manuscript submission, all authors completed the author disclosure form. Employment or Leadership: None declared. Consultant or Advisory Role: None declared. Research Funding: A.C. Heijboer, Dutch Foundation for Quality Assessment, Endocrinology section. Expert Testimony: None declared. Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

**Acknowledgments:** We thank the medical technologists of the endocrine laboratories of the VU University Medical Center, the Academic Medical Center of the University of Amsterdam, and the medical laboratory of Nij Smellinghe Hospital for their excellent technical support.

**REFERENCES**