FERTILITY AND STERILITY®
VOL. 69, NO. 2, FEBRUARY 1998

Published by Elsevier Science Inc.
Printed on acid-free paper in U.S.A.

Measurement of total serum testosterone levels using commercially available kits: high degree of between-kit variability

Larry R. Boots, Ph.D., Sally Potter, B.S.,* H. Downing Potter, B.S.,* and Ricardo Azziz, M.D., M.P.H.*†‡*

University of Alabama at Birmingham, Birmingham, Alabama

Objective: The measurement of total serum testosterone has an established clinical role in the management of male hypogonadism and female androgen excess disorders. We studied the between-kit variability and precision of six different commercially available testosterone assays and compared them with an established in-house method.

Design: Laboratory observational prospective study.

Setting: Tertiary university medical center clinical laboratory.

Patient(s): Three groups of samples each of men (n = 36) and women (n = 15) who had high, normal, or low levels of sex hormone-binding globulin (SHBG), respectively, were studied.

Intervention(s): Individual and pooled (male and female) serum samples were analyzed for total testosterone concentration using six different commercially available assays and one in-house method.

Main Outcome Measure(s): The between-kit variability and the effect of the mean (± SD) SHBG level were determined, the results obtained with the use of the kits and the in-house method were compared, and the intraassay variability (i.e., precision) was evaluated.

Result(s): Male samples demonstrated a 26.3%–40.8% variance in the results obtained with different kits, which was greatest for samples with the lowest SHBG levels. For female samples, between-kit variability ranged from 57%–115% (average, 77%). The percent deviation of the results obtained with the use of commercial methods from those obtained with the use of our in-house assay was greater for men (mean variance, 194%) than for women (mean variance, 67%). The female pool intraassay coefficient of variation was 3.8% with the use of the in-house method and ranged from 8.9%–21.2% with the use of the commercial kits. The male pool intraassay coefficient of variation was 3.1% with the use of the in-house method and ranged from 3.3%–5.5% with the use of the commercial kits.

Conclusion(s): Most commercially available kits for measuring the total serum testosterone level demonstrated significant between-kit variability, which was greatest for female samples. Further, samples with the lowest SHBG levels had the highest between-kit variances. These data strongly suggest that the measurement of total serum testosterone using commercial kits may have limited utility, particularly for the detection of hyperandrogenemia. (Fertil Steril® 1998;69: 286–92. ©1998 by American Society for Reproductive Medicine.)

Key Words; Testosterone, radioimmunoassay, androgens, laboratory

Received May 12, 1997; revised and accepted September 18, 1997.

Presented in part at the Annual Meeting of The Endocrine Society, 1995.

Supported in part by grant RO1-HD29364 from the National Institutes of Health, Bethesda, Maryland (R.A.).

* Department of Obstetrics and Gynecology.

† Department of Medicine.

‡ Reprint requests: Ricardo Azziz, M.D., M.P.H., Department of Obstetrics and Gynecology, The University of Alabama at Birmingham, 618 South 20th Street, OHB 549, Birmingham, Alabama 35233-7333 (FAX: 205-934-0914).

0015-0282/98/$19.00
PII S0015-0282(97)00464-0

The measurement of total testosterone in serum has been recommended in the evaluation and management of female androgen excess disorders, particularly for the exclusion of androgen-secreting tumors (1–4). It has been suggested that patients with testosterone levels >200 ng/dL be evaluated for an androgen-secreting tumor. Nonetheless, most published reports studying the value of this androgen measurement use highly specific assay techniques (5).

In contrast, most clinicians who request testosterone levels submit samples to laboratories that make use of commercially available kits for the measurement of this steroid. Hence, a high degree of variability and inaccuracy in testosterone assay results will have a significant impact on the evaluation of female hyperandrogenism. As a result, witness the frequent finding of the hirsute woman whose androgen levels, in particular total testosterone, are normal. This in turn has resulted in many of

these patients being designated as having "idiopathic hirsutism." We have argued that many of the "normal" androgen results that are obtained actually reflect insensitive or inaccurate assay methods and not the absence of hyperandrogenemia (5).

A few investigators (6, 7) have reported that sex hormone-binding globulin (SHBG) interferes with the measurement of testosterone when RIA kits are used on unextracted serum. In essence, these reports state that when testosterone is measured in serum that contains low levels of SHBG, the testosterone value may be overestimated, and vice versa. Because androgen excess frequently is associated with suppressed SHBG levels (8), it is possible that kits that use unextracted serum could produce an overestimation of the testosterone level in these patients.

We hypothesized that the measurement of total testosterone using commercial assays on unextracted serum yields results that are not sufficiently accurate or consistent to be useful in the evaluation of women with hyperandrogenism. Further, we hypothesized that the circulating SHBG concentration affects the performance of the commercial testosterone assay kits.

The objectives of the present study were [1] to determine the between-kit variability in total testosterone levels measured by six different commercial assays and to compare the results of these assays to the results obtained with the use of an in-house procedure; [2] to determine the effect of circulating SHBG levels on these two parameters; and [3] to determine the precision (intra-assay variance) of the assays.

MATERIALS AND METHODS

Study Samples

Our laboratory routine specifies that all excess serum samples that are not used in obtaining the clinical measures be stored at −20°C for up to 4 years. From these stored samples, a total of 41 serum aliquots were selected. Each sample was obtained from a single individual (i.e., not pooled or batched) and previously had been assayed for circulating SHBG levels (see later). The study was approved by the Institutional Review Board of the University of Alabama at Birmingham. Samples were selected and grouped according to gender and SHBG level as follows:

1. Males-high: Twelve samples from men with SHBG levels of >220 nmol/L.
2. Males-normal: Twelve samples from men with SHBG levels of 140–220 nmol/L.
3. Males-low: Twelve samples from men with SHBG levels of <140 nmol/L.
4. Females-high: Five samples from women with SHBG levels of >400 nmol/L.
5. Females-normal: Five samples from women with SHBG levels of 210–400 nmol/L.
6. Females-low: Five samples from women with SHBG levels of <210 nmol/L.

Each of these samples was assayed separately for testosterone levels in duplicate using seven different methods: the in-house method and the six commercial assay kits, as described later.

Determining Intra-assay Precision

Aliquots from each of the individual serum samples were combined to produce two pooled samples, one male and one female. Ten aliquots from each of these pooled specimens then were assayed consecutively by our in-house method and by the six different commercial kits, to estimate the degree of intraassay variability.

Hormonal and Sex Hormone-binding Globulin Assays

Sex Hormone-binding Globulin

Sex hormone-binding globulin binding capacity was measured by diffusion equilibrium dialysis using Sephadex G-25 (Pharmacia Biotech AB, Uppsala, Sweden) and ^{3}[H]-testosterone (New England Nuclear, Boston, MA) as the ligand, as previously described (9).

In-house Testosterone Assay

Testosterone was measured using a double antibody RIA method after serum extraction, as previously described (10). In brief, 0.5 mL of serum (0.1 mL when assaying samples from men) and 0.1 mL of ^{3}H-testosterone (approximately 15,000 cpm) was added to 15-mL extraction tubes and diluted to 1.0 mL with deionized H_2O. The steroids were extracted using 5.0 mL of diethyl ether and the extract was dried under air and resuspended in 0.5 mL of 0.15-M phosphate buffer (pH 7.4). Fifty μL of the antibody solution (1:2,000) then was added to subject, control, and standard curve tubes. The cross-reactivity of the antibody has been reported previously (10).

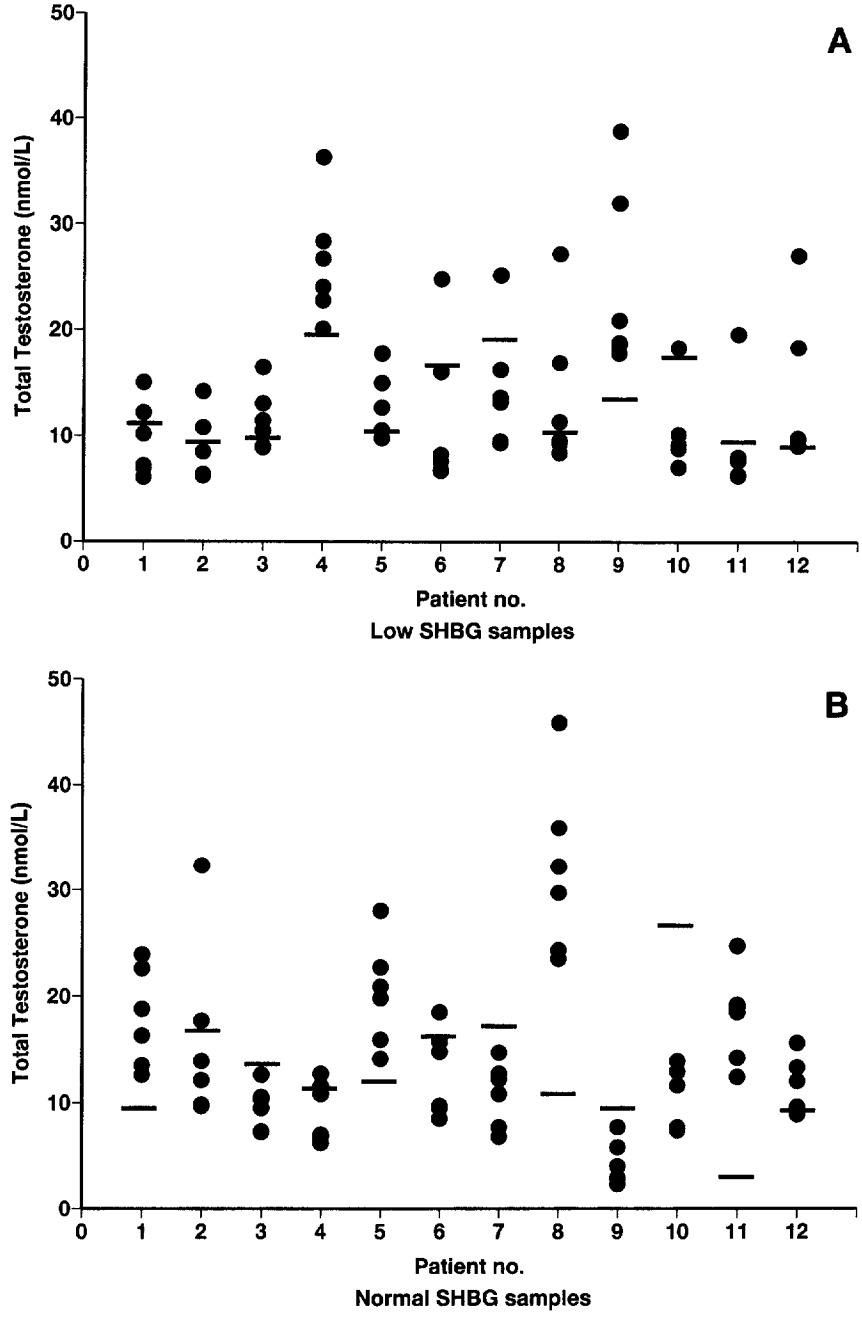
The tubes were incubated for 30 minutes at 4°C. Swirling dextran-coated charcoal (0.1 mL) then was added and the suspension was allowed to sit for 5 minutes on ice. The tubes were centrifuged for 5 minutes at 3,000 rpm and the supernatant was decanted into previously labeled 7-mL glass counting vials. After the addition of 5.0 mL of cocktail, the vials were counted in a scintillation counter. The data were reduced according to standard procedures, and testosterone levels were calculated.

Commercial Testosterone Assays

Six commercial kits were studied. All measured testosterone using a polyclonal RIA method, on nonextracted serum. Those that used a double antibody method included an iodine-125–testosterone tracer, a rabbit antitestosterone antibody, and a precipitating solution containing goat anti-rabbit gamma globulin. Those that used a coated tube

FIGURE 1

Total testosterone levels in samples of 12 men with either low (**A**), normal (**B**), or high (**C**) sex hormone-binding globulin (SHBG) mean levels, assayed by six commercial kits (*individual points*) and an in-house assay method (*individual lines*).



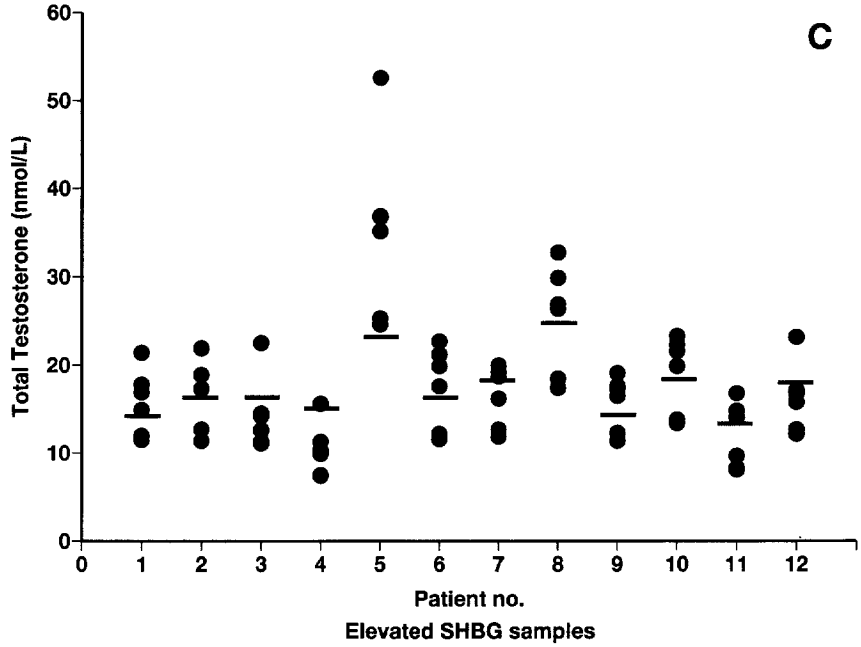
method included an iodine-125–testosterone tracer and a rabbit antitestosterone immunoglobulin immobilized to the inside of polypropylene tubes. The assays studied were as follows:

A. Double antibody method. Standards: 0, 0.1, 0.4, 0.8, 1.6, 3.2, 6.4, 12.8, and 25.6 ng/mL (0, 0.4, 1.4, 2.8, 5.5, 11.1, 22.2, 44.4, and 88.8 nmol/L) (catalog no. 135; Pantex, Santa Monica, CA). Expected adult ranges per package insert: men, 3–12 ng/mL (10.4–41.6 nmol/L); women, 0.25–0.85 ng/mL (0.86–2.95 nmol/L).

B. Coated tube method. Standards: 0, 0.1, 0.4, 0.8, 1.6, 3.2, 6.4, 12.8, and 25.6 ng/mL (0, 0.4, 1.4, 2.8, 5.5, 11.1, 22.2, 44.4, and 88.8 nmol/L) (catalog no. 335M; Pantex). Expected

FIGURE 1

Continued.



adult ranges per package insert: men, 3–12 ng/mL (10.4–41.6 nmol/L); women, 0.25–0.85 ng/mL (0.86–2.95 nmol/L).

C. Double antibody method. Standards: 0, 0.1, 0.5, 2.5, 10.0, and 25.0 ng/mL (0, 0.4, 1.7, 8.7, 34.7, and 86.7 nmol/L) (catalog no. DSL-4100; Diagnostic Systems Laboratories, Inc., Webster, TX). Expected adult ranges per package insert (mean ± 2 SD): men, 2.8–8.8 ng/mL (9.7–30.5 nmol/L); women, 0.1–0.8 ng/mL (0.34–2.77 nmol/L).

D. Coated tube method. Standards: 0, 0.1, 0.5, 2.5, 10.0, and 25.0 ng/mL (0, 0.4, 1.7, 8.7, 34.7, and 86.7 nmol/L) (catalog no. DSL-4000; Diagnostic Systems Laboratories, Inc.). Expected adult ranges per package insert (mean ± 2 SD): men, 2.8–8.8 ng/mL (9.7–30.5 nmol/L); women, 0.1–0.8 ng/mL (0.34–2.77 nmol/L).

E. Coated tube method. Standards: 0, 0.2, 0.5, 1.0, 2.5, 7.5, and 20.0 ng/mL (0, 0.7, 1.7, 3.5, 8.7, 26.1, and 69.6 nmol/L) (catalog no. 1558; Incstar, Stillwater, MN). Expected adult ranges per package insert: men, 1.88–8.96 ng/mL (6.5–31.1 nmol/L); women, not detectable–0.62 ng/mL (not detectable–2.15 nmol/L).

F. Coated tube method. Standards: 0, 0.2, 1.0, 4.0, 8.0, and 16.0 ng/mL (0, 0.7, 3.5, 14.0, 28.0, and 55.0 nmol/L) (Coated Tube Solid Phase Assay Procedure; Diagnostic Products Corp; Los Angeles, CA). Expected adult ranges per package insert: men, 2.45–18.36 ng/mL (8.5–63.6 nmol/L); women, 0.22–0.80 ng/mL (0.76–2.77 nmol/L).

Statistical Calculations

Between-kit variability was calculated for each individual sample by calculating the percent coefficient of variation (CV) ([standard deviation/mean] × 100) for the results obtained with each kit. When the kit results were read as below detectable levels, the lowest value detectable according to the kit was assigned as the value. We calculated a percent deviation for each individual sample for the results obtained using each commercial assay from the results obtained for the same sample using our in-house testosterone assay method (i.e., percent deviation = |(commercial result for sample-in-house result for same sample)/in-house result for same sample| × 100). The CVs between groups were compared by analysis of variance after first-log transformation of the individual sample CVs. The intraassay precision for each type of assay was calculated by determining the percent CV of the 10 consecutive results obtained in the male and female pooled specimens. Other mean (± SD) values (e.g., SHBG) were compared by analysis of variance.

RESULTS

Between-kit Variance of Commercial Assay Methods for the Measurement of Total Testosterone in Men

As expected, mean (± SD) SHBG levels were 122 ± 9 nmol/L, 161 ± 17 nmol/L, and 294 ± 79 nmol/L, for samples in the low, normal, and high SHBG groups ($P < 0.001$). Individual results for total testosterone values obtained for each individual sample and for each commercial kit are depicted in Figure 1. Male samples averaged a between-kit variance of 32.6% ± 11.5%,

which decreased as the mean (± SD) level of SHBG increased. Samples with a high SHBG level had a mean (± SD) CV of 26.3% ± 3.7%, those with a normal SHBG level had a mean (± SD) CV of 30.7% ± 10.2%, and those with a low SHBG level had a mean (± SD) CV of 40.8% ± 13.5%, which represented a statistically significant difference ($P < 0.05$).

The percent deviation of testosterone results obtained with the use of commercial methods from the values obtained with the use of our in-house assay averaged 194% (Table 1). The degree of variance from the values obtained with the use of our in-house assay appeared to be unaffected by the mean (± SD) circulating SHBG level. Finally, the male pool intra-assay CV was 3.1% using the in-house method and ranged from 3.3%–5.5%.

Between-kit Variance of Commercial Assay Methods for the Measurement of Total Testosterone in Women

Mean (± SD) SHBG levels varied by sample group, with samples with low SHBG levels averaging 166 ± 11 nmol/L, those with normal SHBG levels averaging 262 ± 36 nmol/L, and those with high SHBG levels averaging 544 ± 79 nmol/L ($P < 0.001$). Individual results for total testosterone obtained for each individual sample and for each commercial kit are depicted in Figure 2. Between-kit variability ranged from 57%–115% (Table 2), with an average variability of 77%. Again, the degree of between-kit variability decreased minimally as the mean (± SD) level of SHBG increased, such that the average between-kit CV was 75% ± 5.8% in samples with high levels of the globulin, 79% ± 16.5% in those with normal levels, and 85.4% ± 24.9% in those with low levels, although the difference failed to reach statistical significance.

As predicted by their lower mean (± SD) circulating testosterone levels, the mean (± SD) deviation was less for female than for male samples, averaging 67% (Table 3). The results for an individual sample could deviate by as much as 400% from those obtained with the use of our in-house testosterone assay. Again, the percent deviation from our in-house assay results was similar, regardless of the mean (± SD) SHBG level. In determining assay precision, the female pool intraassay CV was 3.8% with the use of the in-house method and ranged from 8.9%–21.2% with the use of the commercial kits.

DISCUSSION

A number of investigators (11, 12) have recently become concerned with the level of accuracy of currently used clinical chemistry and hormonal assays. In this regard, our data clearly demonstrate that the measurement of total testosterone in serum with the use of commercial kits is highly variable and probably inaccurate. This is of particular importance in the study of women, in whom the results of different kits varied by an average of 77%. Therefore, the results for an individual female sample could range from within normal limits to well above normal (according to the expected ranges reported by the kit insert) and those for an individual male sample could range from subnormal to normal, simply by varying the kit used to measure total testosterone. The degree of between-kit variability, although still statistically significant, was considerably less for male samples, averaging 32%.

Among female samples, the results obtained with the use of commercial kits deviated by as much as 400% from those obtained with the use of our in-house assay. The kits that demonstrated the least deviation (kits D and F) had mean (± SD) variances from the in-house method results ranging from 22%–63%, depending on the mean (± SD) SHBG value. Compared with our in-house assay system, some of the kits tended to overestimate the total testosterone level consistently, whereas others underestimated it.

Although it may be reasonable to assume that those kits with lower values also would have a lower range of expected values reported in their package insert, this did not appear to be the case. In fact, the expected ranges included in the package inserts were relatively similar for all six commercial assay methods, with the exception of

TABLE 1

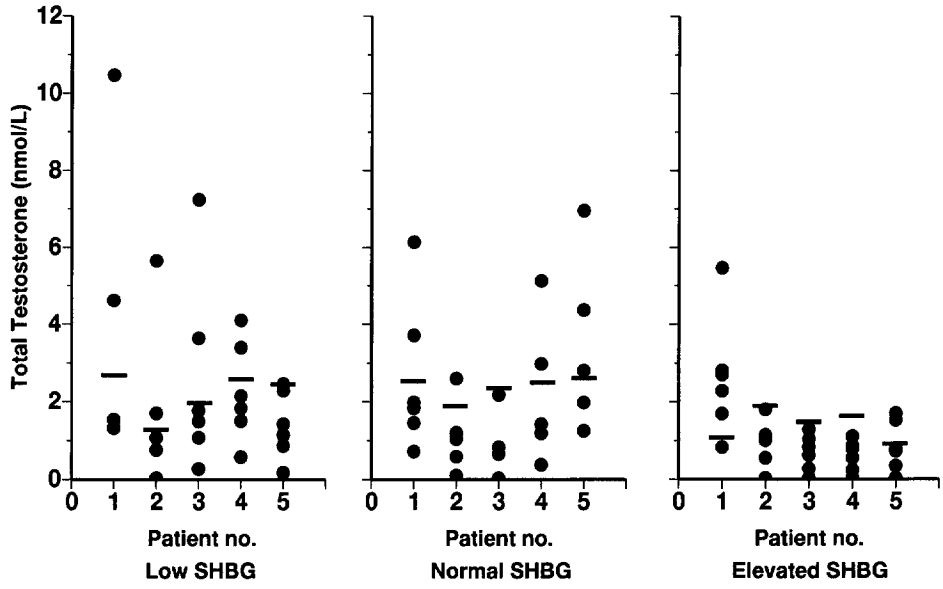
Percent deviation of testosterone values obtained with the use of commercial assays from the values obtained with the use of an in-house method in male samples.

Sample group	A	B	C	D	E	F
Low SHBG	232 (10–352)	204 (8–273)	149 (1–117)	128 (1–107)	168 (11–144)	172 (8–119)
Normal SHBG	317 (12–606)	199 (<1–443)	165 (4–306)	151 (6–255)	208 (3–429)	232 (3–450)
Elevated SHBG	208 (<1–59)	204 (3–59)	133 (6–50)	133 (9–50)	205 (4–51)	288 (4–127)

Note. SHBG = sex hormone-binding globulin.
Values are means with ranges in parentheses.
* See text for particulars of assay.

FIGURE 2

Total testosterone levels in samples of five women with either low, normal, or high sex hormone-binding globulin (SHBG) levels, assayed by six commercial kits (*individual points*) and an in-house assay method (*individual lines*).



kit E, whose reported lower limit of expected female values was below that detectable by the assay (<0.03 nmol/L). Of particular concern was the fact that when a manufacturer produced more than one type of kit (e.g., double antibody and coated tube methods), the same expected range of values usually was reported in both package inserts, suggesting that a specific attempt to establish the expected (and presumably normal) range of testosterone levels for each assay method had not been made.

TABLE 2

Between-kit variability in the measurement of testosterone levels in female samples using six different commercial kits.

Sample group	Sample no.	Between-kit variability: Mean ± SD (nmol/L)	Between-kit variability: Percent CV
Elevated SHBG	1	2.6 ± 1.6	60
	2	0.9 ± 0.6	63
	3	0.7 ± 0.5	69
	4	0.6 ± 0.4	69
	5	0.8 ± 0.6	75
Normal SHBG	6	2.6 ± 2.0	78
	7	1.1 ± 0.8	75
	8	0.7 ± 0.8	106
	9	2.2 ± 1.6	75
	10	3.4 ± 2.1	61
Low SHBG	11	4.5 ± 4.3	96
	12	1.7 ± 2.0	115
	13	2.6 ± 2.5	97
	14	2.3 ± 1.3	57
	15	1.4 ± 0.9	62

Note. SHBG = sex hormone-binding globulin.

The degree of between-kit variability was correlated negatively with the mean (± SD) SHBG level of the sample analyzed for both female and male samples (although it was statistically significant only for male samples). This is of particular concern in the evaluation of patients with hyperandrogenism, in whom testosterone values most often are requested, because these patients often have lower circulating SHBG levels (8).

However, unlike a previous report (7), the mean (± SD) circulating SHBG level did not appear to affect the accuracy of the test result (i.e., the degree of deviation from our in-house results). In fact, two of the manufacturers (Incstar and Diagnostic Products Corp.) also had evaluated the effect of SHBG concentrations on their testosterone results and had reported no observable effect in their package inserts.

Finally, in contrast to the degree of between-kit variability and variance from the results obtained with our in-house assay, the intraassay variability of commercial RIA kits (i.e., a measure of assay precision) was acceptable for most of the kits, although it was higher than that of the in-house method. As expected, in view of their lower mean (± SD) levels and smaller sample numbers, variances in the female samples were greater than those in the male samples.

TABLE 3

Percent deviation of testosterone values obtained with the use of commercial assays from the values obtained with the use of an in-house method in female samples.

Sample group	A	B	C	D	E	F
Low SHBG	189	43	48	22	88	30
	(6–329)	(1–91)	(22–44)	(7–39)	(47–93)	(14–51)
Normal SHBG	89	47	56	35	81	33
	(3–151)	(24–126)	(29–85)	(1–75)	(1–36)	(1–71)
Elevated SHBG	111	64	71	60	83	63
	(4–410)	(12–152)	(58–85)	(22–113)	(1–24)	(19–161)

Note. SHBG = sex hormone-binding globulin.
Values are means with ranges in parentheses.
* See text for assay specifics.

Our results are similar to those obtained a decade ago by Wheeler and colleagues (13, 14), who studied 13 commercial RIA kits for the measurement of testosterone and found many of them to be of limited value. Nonetheless, those reports primarily investigated kits used in Europe, and only one of these was a solid-phase separation system (Coat-A-Count by Diagnostic Products Corp., which also was studied in the present report).

The present study, which evaluated more recently developed commercial RIA kits, also concludes that most demonstrate significant between-kit variability and unacceptable accuracy. This is of greater concern in the evaluation of women with hyperandrogenism, in whom the between-kit variability is greatest, the SHBG level is lowest, and the normal range is most narrow.

These data strongly suggest that the measurement of total serum testosterone with the use of commercial kits may have limited clinical usefulness, particularly for confirmation of androgen excess. For example, the diagnosis of the polycystic ovary syndrome, according to the 1990 National Institute of Child Health and Human Development consensus conference (15), entails the presence of oligo-ovulation with clinical evidence of either hyperandrogenism (i.e., hirsutism, acne, alopecia) or hyperandrogenemia (e.g., elevated testosterone levels). Hence, an insensitive assay, or a highly variable result, may have the effect of reducing the number of patients in whom polycystic ovary syndrome is detected, particularly among young or Asian patients, who have less peripheral evidence of hyperandrogenism.

References

1. American Society for Reproductive Medicine. The evaluation and treatment of androgen excess. Guidelines for practice. Birmingham (AL): American Society for Reproductive Medicine, 1995.
2. Evaluation and treatment of hirsute women. Washington, D.C.: American College of Obstetricians and Gynecologists, 1995; Technical Bulletin, No. 203.
3. Kessel B, Liu J. Clinical and laboratory evaluation of hirsutism. Clin Obstet Gynecol 1991;34:805–16.
4. Rittmaster RS. Medical treatment of androgen-dependent hirsutism. J Clin Endocrinol Metab 1995;80:2559–63.
5. Meldrum DR, Abraham GE. Peripheral and ovarian venous concentrations of various steroid hormones in virilizing ovarian tumors. Obstet Gynecol 1979;53:36–43.
6. Slaats EH, Kennedy JC, Krujjswijk H. Interference of sex hormone-binding globulin in the "Coat-A-Count" Testosterone No-Extraction Radioimmunoassay. Clin Chem 1987;33:300–2.
7. Masters A, Hahnel R. Investigation of sex hormone-binding globulin interference in direct radioimmunoassays for testosterone and estradiol. Clin Chem 1989;35:979–84.
8. Azziz R. Hirsutism. In: Sciarra JJ, Ditts PV, editors. Gynecology and obstetrics. vol 5. Philadelphia: JB Lippincott Co., 1994:1–22.
9. Pearlman WH, Crepy O. Steroid-protein interaction with particular reference to testosterone binding by human serum. J Biol Chem 1967; 242:182–9.
10. Azziz R, Bradley EL Jr, Potter HD, Parker CR Jr, Boots LR. Chronic hyperinsulinemia and the adrenal androgen response to acute corticotropin-(1-24) stimulation in hyperandrogenic women. Am J Obstet Gynecol 1995;172:1251–6.
11. Potischman N, Falk RT, Laiming VA, Siiteri PK, Hoover RN. Reproducibility of laboratory assays for steroid hormones and sex hormone-binding globulin. Cancer Res 1994;54:5363–7.
12. Tietz NW. Accuracy in clinical chemistry—does anybody care? Clin Chem 1994;40:859–961.
13. Wheeler MJ, Shaikh M, Jennings RD. An evaluation of 13 commercial kits for the measurement of testosterone in serum and plasma. Ann Clin Biochem 1986;23:303–9.
14. Wheeler MJ, Lowy C. Warning on serum testosterone measurement. Lancet 1987;2:514–5.
15. Zawadzki JK, Dunaif A. Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. In: Dunaif A, Givens JR, Haseltine F, Merriam GR, editors. Polycystic ovary syndrome. Boston: Blackwell Scientific Publications, 1992:377–84.