0021-972X/04/$15.00/0
Printed in U.S.A.

The Journal of Clinical Endocrinology & Metabolism 89(2):525–533

doi: 10.1210/jc.2003-030680

Measurement of Free Testosterone in Normal Women and Women with Androgen Deficiency: Comparison of Methods

KAREN K. MILLER, WILLIAM ROSNER, HANG LEE, JOAN HIER, GEMMA SESMILO, DAVID SCHOENFELD, GREGORY NEUBAUER, AND ANNE KLIBANSKI

Neuroendocrine Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School (K.K.M., J.H., G.S., A.K.) Boston, Massachusetts 02114; College of Physicians and Surgeons, Columbia University and St. Luke's-Roosevelt Hospital Center (W.R.), New York, New York 10019; MGH Biostatistics Center and the Clinical Research Center, Massachusetts General Hospital and Harvard Medical School (H.L., D.S.), Boston, Massachusetts 02114; and General Clinical Research Center, Massachusetts General Hospital and Harvard Medical School (G.N.), Boston, Massachusetts 02114

Androgen deficiency in women is increasingly recognized as a new clinical syndrome and has raised our awareness of the importance of accurate and well-validated measurements of serum free testosterone (T) concentrations in women. Therefore, we compared serum free T levels measured by equilibrium dialysis to those measured by a direct RIA (analog method) and to those calculated from the law of mass action (requires the measurement of total T and SHBG). We also calculated the free androgen index, 100 × T/SHBG, as a simple index known to correlate with free T. Subjects were 147 women with variable androgen and estrogen statuses. All were studied three times in 1 month and included women 1) with regular menses (estrogen positive, T positive), 2) more than 50 yr old and not receiving estrogen (estrogen negative, T positive), 3) receiving estrogen (estrogen positive, T negative), and 4) with severe androgen deficiency secondary to hypopituitarism (estrogen negative, T negative). Calculated values for free T using the laws of mass action correlated well with those obtained from equilibrium dialysis ($r = 0.99$; $P < 0.0001$). However, the agreement depended strongly on the specific assays used for total T and SHBG. In contrast, the direct RIA method had unacceptably high systematic bias and random variability and did not correlate as well with equilibrium dialysis values ($r = 0.81$; $P < 0.0001$). In addition, the lower limit of detection was higher for the direct RIA than for equilibrium dialysis or calculated free T. Free androgen index correlates well with free T by equilibrium dialysis ($r = 0.93$; $P < 0.0001$), but is a unitless number without reference to the physical reality of free T. We conclude that the mass action equation and equilibrium dialysis are the preferred methods for use in diagnosing androgen deficiency in women. (*J Clin Endocrinol Metab* 89: 525–533, 2004)

ANDROGEN DEFICIENCY IS increasingly recognized as an important clinical syndrome in women (1–3). However, because the symptoms and signs of this syndrome are relatively nonspecific, physicians are more than usually dependent upon accurate assays for total and free testosterone (T). In addition, because of the highly variable concentrations of SHBG, the high affinity of SHBG for estrogen, and the large variation in estrogen levels among different populations of women, measurement of serum free T, in addition to total T, is necessary to determine whether abnormal androgen concentrations are present. The ideal method with which to measure free T is controversial. Further, the debate has largely neglected the fact that an accurate measure of free T is totally dependent upon an accurate assay for total T. The issues of sensitivity and accuracy are particularly problematic in women, who have 10-fold lower plasma concentrations of total T and 20-fold lower concentrations of free T than men. Accurate measurement of these steroids are even more challenging in women with androgen deficiency, such as in hypopituitarism.

Abbreviations: CV, Coefficient of variation; FAI, free androgen index; IRMA, immunoradiometric assay; ROC, receiver operator characteristic.

JCEM is published monthly by The Endocrine Society (http://www.endo-society.org), the foremost professional society serving the endocrine community.

To understand how current methods for T have come to be used, it is necessary to explore briefly how they came to be. Early measurements of androgenic activity were made using bioassays until the early to mid 1960s, when a number of alternative methods were developed. Electron capture after gas liquid chromatography (4) and double-isotope derivative formation (involving multiple chromatographs, derivative formation and crystallization) (5) yielded values that were accurate and sensitive. However, the sensitivity of the methods depended upon extracting T from large volumes (10–20 ml) of plasma. The techniques were cumbersome and slow and were not suited for the routine clinical laboratory. In the late 1960s, plasma SHBG was introduced as a reagent for a ligand binding assay for T (6–10). The methodology was much simpler and faster than previous ones, and the original reports took great care to validate the assays. Extraction of T from plasma with an organic solvent and one or more chromatographs were still the rule, but great progress was made in reducing the volume of plasma required to less than 10 ml. Nevertheless, the assays left much room for improvement, because SHBG binds not only T, but also other circulating steroids, and has a relatively fixed K_d (it can be altered by temperature, ionic strength, *etc.*), which is a major determinant of sensitivity. With the primary goal of improving sensitivity and specificity, endocrinologists turned to RIAs (11–

13). The early assays were sufficiently sensitive to measure T in extracts of 0.25 ml male plasma and 1.0 ml female plasma. Thus, by the early 1970s, the basis for today's technologies was in place.

In the ensuing 3 decades, RIAs and other kinds of immunoassays for T evolved that required smaller volumes and less sample preparation, so that today's clinical chemistry laboratories use automated platforms with no extractions into organic solvents and no chromatography. These assays are available in kit form to endocrine and clinical laboratories and are subject to quality control by the College of American Pathologists. However, this quality control does not require vigorous validation. The College of American Pathologists accepts the ability of a given laboratory to obtain the same values (within statistical limits) as other laboratories using the same method or kit, even when different kits give different values for the same sample. There is no requirement for accuracy; further, the community that relies on them cannot examine the way in which many of these assays have been validated, as these methods and data are proprietary secrets, not published in peer-reviewed journals. Remarkably, a new set of normal values is sometimes arbitrarily introduced, rendering the comparison of current results with those of previously obtained results impossible. Tandem mass spectroscopy is under development and is being validated. It is an alternative method that may be available to clinicians and researchers in the near future.

The actual concentration of free T in plasma is probably unknowable. The concept of free T arises from a simple consideration of mass action and equilibria. The fact that two molecules may interact reversibly is not in question, because in some cases, the three different species [the complex and each of the uncomplexed (free) molecules] can be observed by physical methods that do not disturb the equilibrium. There is general agreement that, in expert hands, equilibrium dialysis is the best approximation to the concentration of free T in plasma despite considerable technical difficulties. However, there are probably factors *in vivo*, *e.g.* a binding protein for androgens on the vessel wall, that may make the absolute estimate of free T *in vivo* inaccurate. Therefore, there is a degree of arbitrariness about the best way to measure free T.

For the studies in this paper, we have sent samples to a commercial laboratory that performs an RIA for total T after extraction and a single chromatograph. This method is congruent to methods that were validated in the 1970s and is probably accurate. We compare serum free T levels as determined by equilibrium dialysis (Esoterix Endocrinology, Calabasas Hills, CA), the equations based on the law of mass action, and a direct RIA (Diagnostic Systems Laboratories, Inc., Webster, TX). More accessible methods, including the direct RIA of free T and the calculation of a simple free androgen index (FAI; 100 × total T/total SHBG), may not be valid in all populations (14–16). Indeed, that the direct RIA may be seriously flawed has been pointed out on the basis of internal inconsistencies arising from its use (17). Few data are available regarding the use of any of these methods at the lower limits of detection necessary to diagnose androgen deficiency in women. Despite the fact that alternative, easier to perform, lower cost assays, particularly the direct RIA method, have not been validated in women, they are used regularly in clinical laboratories and in research. For example, 48% of the 67 papers that reported free T concentrations, published in the *Journal of Clinical Endocrinology and Metabolism* between January 1998 and November 2000, used the direct RIA method (18).

We assessed serum free T three times in 1 month in women using equilibrium dialysis, a direct RIA, and calculation using the law of mass action and the FAI. We compared these methods in women with normal and reduced androgen and variable estrogen levels. In addition, because all of these methods, except for the direct RIA, depend upon the accuracy of the measurements of total T and SHBG, we measured these variables by two different methods to determine the magnitude of the effect on free T results.

Subjects and Methods

Subjects

One hundred and forty-seven women were studied three times during 1 month. Clinical characteristics of the study subjects are shown in Table 1. All normal premenopausal subjects had histories of regular menstrual cycles since menarche. Confirmation of ovulatory function in premenopausal controls, not receiving estrogen, was documented with a serum progesterone of more than 16 nM in the luteal phase of the menstrual cycle. In subjects receiving estrogen, estrogen preparations included oral contraceptives in women of reproductive age and hormone replacement therapy in the form of an estrogen and a progestin in postmenopausal women. In women with hypopituitarism, the diagnoses of hypogonadism and/or hypoadrenalism were made by the subject's endocrinologist or primary care physician. Subjects with central hypogonadism had all been amenorrheic for at least 1 yr and had serum FSH levels in the premenopausal range or lower. The diagnosis of hypoadrenalism was based on an insulin tolerance test or cosyntropin [ACTH-(1–24)] stimulation test, and signs and symptoms of hypoadrenalism. No subject with hypoadrenalism was receiving supraphysiological doses of glucocorticoids. All hypopituitary subjects had normal serum free T_4 indexes, and all normal subjects had serum TSH levels within the normal range. The study was approved by the subcommittee on human studies of Massachusetts General Hospital, and all subjects gave written consent.

Methods

Healthy subjects with regular menstrual cycles (n = 40) and all subjects taking cyclical progestins were studied three times in the menstrual cycle: 1) d 1–7, 2) d 14, and 3) d 21–23. These times were adjusted for women with cycles of more than 28 d. Hypopituitary and postmenopausal subjects who were not receiving estrogen were also studied three times in 1 month to mimic sample timing in the other groups.

TABLE 1. Clinical characteristics of subjects

Mean age (yr)	45 ± 12
Total no. of patients	147
Hormonal status	
Regular menstrual cycles (estrogen +, T +)	40 (27)[a]
Postmenopausal women not receiving estrogen (estrogen −, T +)	18 (12)
Receiving estrogen (estrogen +, T −)	63 (43)
Oral contraceptive	20 (14)
Low-dose estrogen replacement therapy	43 (29)
Hypopituitarism, not receiving estrogen (estrogen −, T −)	41 (28)
Secondary hypogonadism	21 (14)
Secondary hypogonadism and secondary adrenal insufficiency	15 (10)
Secondary adrenal insufficiency without hypogonadism	5 (3)

[a] No. of patients (percentage of patients).

Subjects were admitted while fasting to the General Clinical Research Center at Massachusetts General Hospital or had blood drawn at an out-patient laboratory at 0800 h after an overnight fast. Height and weight were determined, and all subjects completed a detailed medical questionnaire, which included a menstrual history and medication use.

Hormonal assessment

Total T. Blood was drawn from fasted subjects at 0800 h. Serum T was measured using two methods: RIA after extraction and column chromatography (Esoterix Endocrinology) and an Active Testosterone RIA kit (Diagnostic Systems Laboratories, Inc.). In the column chromatography method, a hexane/ethyl acetate extract of 0.5 ml serum is applied to Al_2O_3 microcolumns. The columns are then washed with hexane-containing ethanol, after which T is eluted using hexane-containing ethanol. Tritiated T was used for recovery. The dried samples were assayed using an antibody raised against T-3-oxime-BSA conjugate. ^{125}I was used as the label, and antibody-bound T was separated with ammonium sulfate. Finally, the concentration of T in each sample was calculated from a curve generated by purified T standards. All samples were run in the same batch.

The Active Testosterone RIA kit is a solid phase RIA that places whole, unextracted serum into polypropylene tubes coated with T-specific antibodies to bind T and ^{125}I-labeled T in a standard RIA format. The characteristics of these assays are shown in Table 2.

SHBG. Esoterix Endocrinology measured SHBG by two different methods, an in-house immunoradiometric assay (IRMA) and an RIA accomplished with a kit from Diagnostic Systems Laboratories. The in-house IRMA was calibrated against a binding capacity assay. The characteristics of these assays are shown in Table 2. All samples for SHBG were run in the same batch.

Free T. Free serum T by equilibrium dialysis (Esoterix Endocrinology) measures percent free T by adding [^3H]T to serum and dialyzing the mixture for 11 h against a phosphate-saline buffer, which is placed on the opposite side of the dialysis membrane, at 37 C using a dialysis cell as described by Nelson and Tomei (19). The percent free T is calculated at equilibrium as [^3H]T in buffer/[^3H]T in serum. [^3H]T is routinely checked for purity by paper chromatography every 2 wk. Serum free T is then calculated by multiplying the percent free T (by equilibrium dialysis) by total T (by RIA after column chromatography). All samples were run in the same batch.

Free T was also measured using a direct RIA kit (Diagnostic Systems Laboratories, Inc.). In this method, an ^{125}I-labeled T analog, which does not bind to SHBG, competes for a fixed time with free T in the patient sample for sites on a T-specific antibody immobilized on the wall of a polypropylene tube. The characteristics of this assay are shown in Table 2.

Free T was calculated based on equations derived from the laws of mass action as reported by Sodergard *et al.* (20) and Vermeulen *et al.* (14). These equations require, as input, total T, SHBG, albumin, and the K_d values (37 C) for the interactions of T with SHBG and albumin. We used the following second degree equation for free T: $T_2(K_1 + K_1K_2Z) - T(1 - K_1X + K_2Z + K_1Y) - X = 0$, where T is the free T concentration, X is the total T concentration, Y is the SHBG concentration, Z is the albumin concentration, K_1 is the association constant for T with SHBG (1×10^9 L/mol), and K_2 is the association constant for the T with albumin (3.6×10^4 liter/mol). The albumin concentration was assumed to be 43 g/liter (6.2×10^{-4} mol/liter), according to Vermeulen *et al.* (14), who demonstrated that small differences in albumin values do not significantly affect free T concentrations. We confirmed that our result was equivalent to that calculated by the equations of Vermeulen *et al.* (14) by comparing it to that derived using a computer program in Basic language provided by Jean M. Kaufman and Alex Vermeulen.

The FAI is the quotient of serum total T (measured by RIA after column chromatography) divided by serum SHBG [measured either by IRMA (Esoterix Endocrinology) or RIA (Diagnostic Systems Laboratory)] × 100, *e.g.* (100 T)/SHBG. The FAI is a unitless number when both values are expressed in molar concentrations.

Lower limit of detection. The lower limit of detection of all assays performed by Esoterix Endocrinology was determined as the lowest control that, when assayed 10 times in three different assays, yielded inter- and intraassay percent coefficients of variation (CVs) less than 20%. For assays performed at the Massachusetts General Hospital General Clinical Research Center core laboratory using commercially available kits, the lower limit of detection reported here is that published by the manufacturer of the kits.

Other hormones. Progesterone, TSH, and the free T_4 index were measured by previously described methods (21). Samples from each individual were measured in duplicate and run in the same assay. All other assays were run by the same technician using kits that accommodated 40 samples.

Free T levels as measured by equilibrium dialysis and total T determined by column chromatography have been previously reported in a manuscript that demonstrated androgen deficiency in women with hypopituitarism compared with healthy controls (1).

Statistical analysis

To determine the degree of agreement between different assays, the differences between the methods were plotted against their averages (22). The relationship between equilibrium dialysis and each alternative method were examined by simple linear regression models. The intercepts, slopes, and r values of dialysis were obtained from plots of each alternative method (independent variable) and equilibrium dialysis (dependant variable). To account for the interdependence of the values (three serum samples from each subject), the regression models were refit by the the generalized estimating equations approach (23). Slopes derived in this analysis did not differ from those derived in the simple linear regression, and the statistical significance remained the same.

Assay reproducibility was calculated as the within-subject CV over three consecutive repeated measurements. For this analysis we used only data from amenorrheic subjects and those taking daily acyclic estrogen/progestin regimens (n = 72 × three samples each). The CV calculation was carried out by one-way random subject effects ANOVA of the logarithmically transformed measure where the resulting root mean square error is equivalent to CV in the original scale (24).

To illustrate the differences in diagnostic sensitivity and specificity between the methods, we constructed a receiver operator characteristic (ROC) curve (sensitivity *vs.* 1 − specificity) at a serum free T concentration of 3.47 pM. This cut-off, the fifth percentile for our group of healthy women, was chosen as a clinically relevant serum level for the diagnosis of androgen deficiency in women.

TABLE 2. Assay characteristics

Assay	Sample volume (μl)	% Intraassay CV	% Interassay CV	Lower limit of detection	Normal female range (reported)	Normal female range (our data)
Total T by RIA after column chromatography (nM)	500	2.9–6.1	9–15	0.10	0.35–1.91	0.15–0.98
Total T by direct RIA (nM)	50	7.0	10.2	0.35	0.35–2.77	0.35–1.23
Free T by equilibrium dialysis (pM)	300	6.6	11.9	0.35	3.82–21.86	1.73–8.67
Free T by direct RIA (pM)	50	6.2	9.7	0.87	1.56–11.00	1.01–3.50
SHBG by IRMA (nM)	50	3.5	8.1	10.0	36–185	60–126.5
SHBG by RIA (nM)	50	2.2	4.4	5.0	30–95	26.1–63.4

Results

Median values

Median and 25–75% ranges of T, free T levels, and SHBG, as determined by all assays used, are shown in Table 3. Data are presented as the median and range, instead of the mean ± SD, because data were not normally distributed.

Agreement between methods

The agreement among the various methods used to determine free T (*i.e.* interchangeability) is shown in Fig. 1. The direct RIA method demonstrated high random and systematic variability, *i.e.* bias. Bias was demonstrated by systematic deviation from a horizontal line at 0.0 (*i.e.* if the methods agreed perfectly, the difference between free T as determined by two methods would equal 0.0).

Correlational analysis

Simple linear regression analyses were performed to define further the relationship between free T determined by dialysis and by every other method as well as to quantitate bias (systematic variability) by calculating correction factors (intercept and slope of the linear relationship between values determined by the two methods (Table 4 and Fig. 2). There was a strong linear relationship between free T by dialysis (percent free T by dialysis × total T by RIA after column chromatography) and that calculated by mass action (using T by RIA after column chromatography and SHBG by IRMA; $r = 0.99$; $P < 0.0001$; Table 4 and Fig. 2A). The slope (1.20; $P < 0.0001$) indicates a systematic difference of approximately 20% in the absolute values of free T between the two methods. This difference could arise from systematic errors in equilibrium dialysis or in the calculated value. If the difficulty is in the calculation, then the most likely sources of the difference would be in the measurement of SHBG, the K_d for SHBG and T, or both. Five free T values obtained by equilibrium dialysis were clear outliers, defined as greater than the mean ± 3 SD, but, when remeasured using the same measurement techniques, were no longer outliers. Thus, the gross experimental error was about 1%.

FAI (using T by RIA after column chromatography and SHBG by IRMA) also correlated well with free T measured by dialysis ($r = 0.93$; $P < 0.0001$). The slope was 7.38, and the x-intercept was 1.07 ($P < 0.0001$; Table 4 and Fig. 2C). The absolute value of the slope is unimportant because, unlike the other plots in which the units, abscissa, and ordinate are the same, the slope of this plot compares a concentration to an index. Thus, the usefulness of the FAI should be judged

TABLE 3. Results [median (25–75% range)]

	All subjects	Estrogen +, T +	Estrogen −, T +	Estrogen +, T −	Estrogen −, T −
Total T by RIA after column chromatography (nM)	0.5 (0.2–1.0)	0.9 (0.6–1.2)	0.8 (0.4–1.2)	0.4 (0.1–0.9)	0.1 (0.1–0.3)
Total T by direct RIA (nM)	0.8 (0.4–1.2)	1.1 (0.8–1.6)	1.0 (0.4–1.4)	0.7 (0.3–1.0)	0.3 (0.3–0.9)
SHBG by IRMA (nM)	110.0 (64.0–154.0)	91.0 (60.0–126.5)	93.5 (59.0–123.0)	152.0 (115.0–238.0)	49.5 (37.0–89.0)
SHBG by RIA (nM)	53.4 (29.7–81.0)	40.5 (26.1–63.4)	40.0 (26.6–52.8)	77.9 (58.0–118.0)	18.2 (16.8–46.8)
Free T by equilibrium dialysis (pM)	4.5 (1.7–8.7)	8.3 (3.1–13.5)	9.9 (4.2–13.9)	2.4 (0.7–4.9)	2.1 (1.0–4.5)
Free T by direct RIA (pM)	2.3 (1.0–3.5)	3.2 (2.3–4.9)	3.4 (2.1–4.8)	1.7 (0.8–3.0)	1.0 (0.8–1.9)
Free T by mass action equation (pM)	3.8 (1.6–7.1)	6.8 (6.1–13.5)	8.7 (3.8–11.1)	2.0 (0.8–4.3)	1.9 (1.1–3.8)
FAI	0.5 (0.2–0.9)	0.9 (0.6–1.4)	1.2 (0.5–1.5)	0.2 (0.1–0.5)	0.3 (0.1–0.7)

Estrogen +, T +, Healthy women with regular menstrual cycles; estrogen −, T +, healthy postmenopausal women not receiving estrogen; estrogen +, T −, women receiving estrogen; estrogen −, T −, women with hypopituitarism not receiving estrogen.

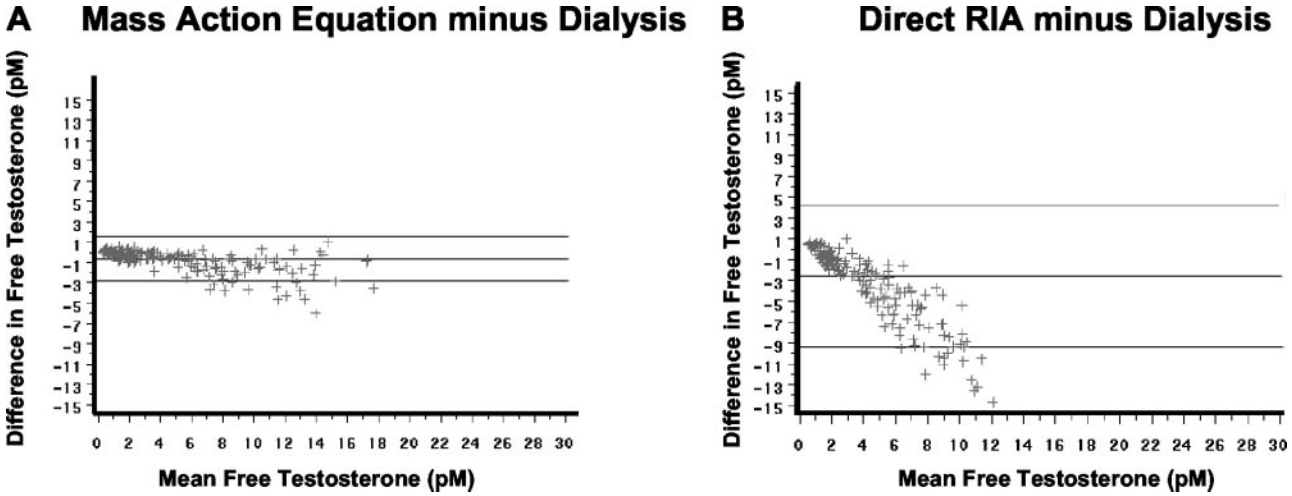


FIG. 1. Agreement between free T by equilibrium dialysis and free T calculated by the mass action equation (A) or by direct RIA (B). Perfect agreement would be indicated by a *horizontal line* through $x = 0.0$ (difference in free T between two methods = 0.0). *Horizontal lines* indicate the mean ± 2 SD.

TABLE 4. Linear regression models: correlation with free T by dialysis

	Direct RIA	Mass action-SHBG by IRMA	FAI-SHBG by IRMA
All subjects (n = 441)			
Correlation coefficient (r)	0.81	0.99	0.93
$y = \alpha + \beta x$	$y = -0.52 + 2.50x$	$y = 0.02 + 1.20x$	$y = 1.07 + 7.38x$
Estrogen +, T + (n = 120)			
Correlation coefficient (r)	0.70	0.97	0.92
$y = \alpha + \beta x$	$y = 1.98 + 2.22x$	$y = 0.09 + 1.20x$	$y = 2.86 + 6.37x$
Estrogen −, T + (n = 54)			
Correlation coefficient (r)	0.74	0.96	0.87
$y = \alpha + \beta x$	$y = 1.46 + 2.36x$	$y = 0.40 + 1.10x$	$y = 2.54 + 6.03x$
Estrogen +, T − (n = 189)			
Correlation coefficient (r)	0.81	0.98	0.97
$y = \alpha + \beta x$	$y = -0.71 + 2.04x$	$y = -0.19 + 1.30x$	$y = -0.07 + 10.73x$
Estrogen −, T − (n = 60)			
Correlation coefficient (r)	0.92	0.99	0.97
$y = \alpha + \beta x$	$y = -0.16 + 2.03x$	$y = -0.19 + 1.24x$	$y = -0.14 + 8.10x$

Estrogen +, T +, Healthy women with regular menstrual cycles; estrogen −, T +, healthy postmenopausal women not receiving estrogen; estrogen +, T −, women receiving estrogen; estrogen −, T −, women with hypopituitarism not receiving estrogen.

by the magnitude of the correlation and the proximity of the intercept to zero.

The comparison between free T by dialysis to that by direct RIA was the weakest of all [r = 0.81 ($P < 0.0001$); slope = 2.50 ($P = 0.06$) and x-intercept = −0.52 ($P < 0.0001$); Table 4 and Fig. 2B]. The slope indicates that, on the average, values for free T determined by the direct RIA method were 40% of those determined by equilibrium dialysis. The minimum level of detection of free T by the direct RIA (0.87 pM; Table 2) was higher than that for dialysis (0.35 pM). This is an important part of the range when measuring free T in women, particularly women with androgen deficiency. Therefore, the commonly used direct RIA method demonstrated the highest bias and random variation, the highest lower limit of detection, and the least accurate values of all assays tested.

Similar results were obtained when subsets of data from populations of women with variable free T ranges and estrogen levels were studied (Table 4). Of note, correlation coefficients were highest in the subsets of women with androgen and estrogen deficiency (r = 0.92–0.99) women with androgen and estrogen deficiency (r = 0.92–0.99).

The correlation between total T by RIA after column chromatography and by direct RIA was modest (r = 0.77; $P < 0.0001$; slope = 0.67; intercept = 0.03; Fig. 3). The correlation between SHBG by RIA and IRMA was strong (r = 0.96; $P < 0.0001$; Fig. 3). However, the slope of the regression line was 1.99, which reflected RIA values approximately half of IRMA values. Compared with free T by dialysis values, the calculated values using total T by direct RIA and SHBG by RIA (the alternative methods) correlated less well. The equations and correlation coefficients describing the relationships were as follows. Total T by RIA without chromatography and SHBG IRMA: dialysis *vs.* calculation: r = 0.76, $y = 0.96 + 0.65x$; dialysis *vs.* FAI: r = 0.72, $y = 2.42 + 3.33x$. Total T after chromatography and SHBG RIA: dialysis *vs.* calculation: r = 0.98, $y = -0.12 + 0.70x$; dialysis *vs.* FAI: r = 0.89, $y = 1.84 + 2.71x$. Total T by RIA without chromatography and SHBG by RIA: dialysis *vs.* calculation: r = 0.87, $y = 0.43 + 0.41x$; dialysis *vs.* FAI: r = 0.72, $y = 2.91 + 1.25x$.

Therefore, although the correlation of free T by dialysis (using total T by RIA after chromatography) and calculated values (using the same total T methodology) was strong, the slope of the regression line differed depending upon whether IRMA or RIA was used to assay SHBG. Moreover, the correlations between dialysis values and calculated values, measuring total T by direct RIA, were poor. This reflected the poor correlation between total T by direct RIA with that of RIA after column chromatography.

95% prediction interval

The 95% prediction interval, *i.e.* precision of prediction of dialysis values from values obtained by other methods, was determined (Fig. 2). With equilibrium dialysis, more than 5% of the values fell outside of the 95% prediction interval. For the direct RIA, 6.6% of values were outside of the 95% prediction interval. For calculated values, 7.3% of values were outside of the 95% prediction interval, and for the FAI, 6.1% of values were outside of the 95% prediction interval.

Sensitivity and specificity

ROC curves (sensitivity *vs.* 1 − specificity) were constructed to determine the sensitivity and specificity of each assay at a free T concentration of 3.47 pM (Fig. 4). The maximal area under the curve reflects maximal sensitivity and specificity. Maximal sensitivity and specificity are indicated by the greatest area under the curve or a vertical line at $x = 0.0$, followed by a horizontal line at $y = 1.0$. Figure 4, therefore, demonstrates that the RIA has lower sensitivity and specificity than the calculated methods at a free T concentration of 3.47 pM. A free T concentration of 3.47 pM is the fifth percentile of free T levels, as determined by equilibrium dialysis, in 40 subjects with regular menstrual periods and therefore was chosen as a clinically relevant level for the diagnosis of androgen deficiency in women. Because values were not normally distributed, the mean ± 2 SD did not yield a meaningful normal range (the lower limit as determined in this manner was <0).

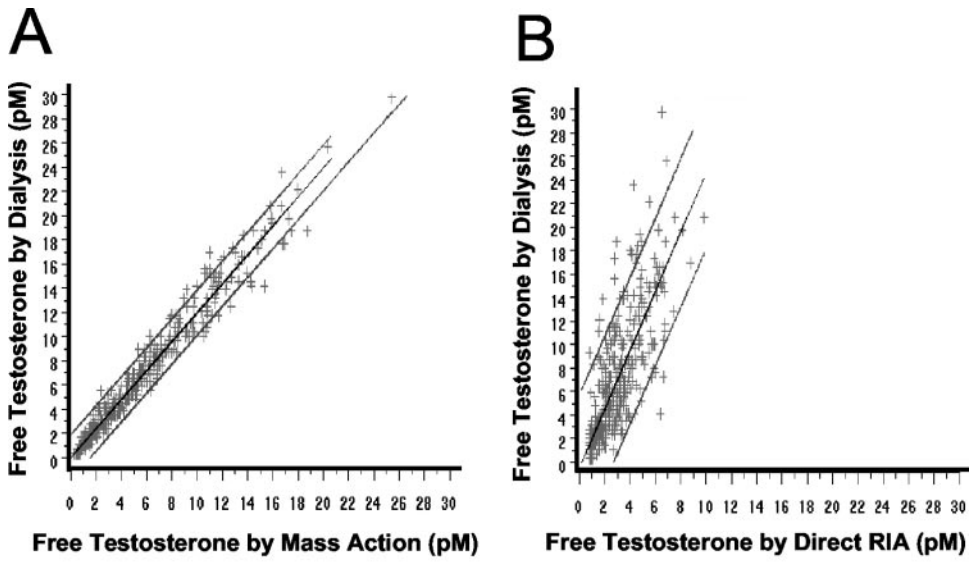
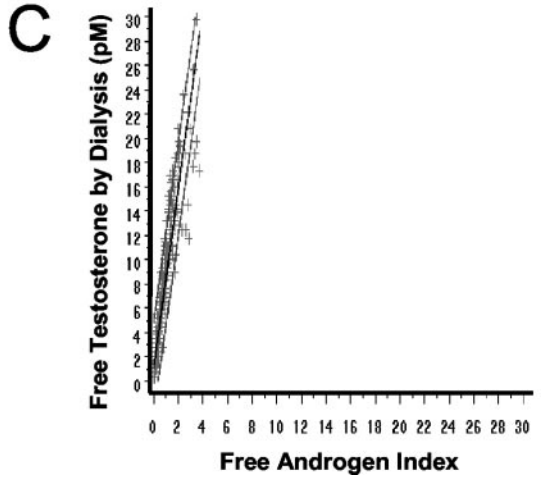


FIG. 2. Correlation of various methods for measuring free T. Simple linear regression models with 95% prediction intervals. A, Equilibrium dialysis *vs.* mass action with SHBG by IRMA ($y = 0.02 + 1.20x$; r = 0.99; $P < 0.0001$). B, Equilibrium dialysis *vs.* direct RIA ($y = -0.52 + 2.50x$; r = 0.81; $P < 0.0001$)). C, Equilibrium dialysis *vs.* FAI ($y = 1.07 + 7.38x$; r = 0.93; $P < 0.0001$).



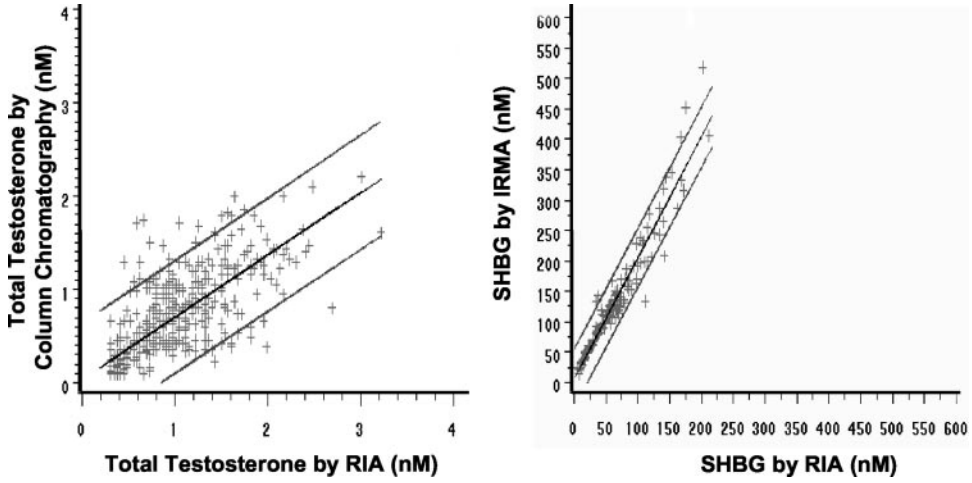


FIG. 3. Comparison of different methods for measuring total T and SHBG. A, Simple regression of serum total T measured by RIA after column chromatography (Esoterix Endocrinology) *vs.* total T by RIA (Diagnostic Systems Laboratories, Inc.) with 95% prediction intervals ($y = 0.67x + 0.03$; r = 0.77; $P < 0.0001$). B, Simple regression of SHBG by in-house IRMA (Esoterix Endocrinology) *vs.* RIA (Diagnostic Systems Laboratories, Inc.) kits ($y = 1.99x + 5.61$; r = 0.96; $P < 0.0001$).

Misclassification

A two by two analysis was performed to determine the number of subjects misclassified by the calculation methods and direct RIA compared with equilibrium dialysis. Four and a half percent of our subjects had free T levels below normal (using a cut-off of <3.47 pM) when calculated using the mass action equation (compared with 5% by equilibrium dialysis), whereas 30.2% and 54.8% would be considered to have androgen deficiency using the direct RIA and FAI methods,

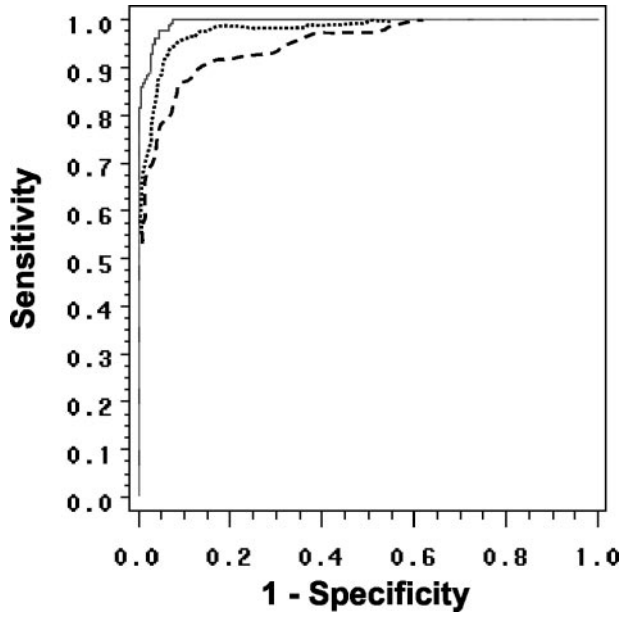


FIG. 4. ROC curves at free T of 3.47 pM for equation based on the laws of mass action (—), FAI (· · ·), and RIA analog (- - -). Maximal sensitivity and specificity are indicated by the greatest AUC or a vertical line at $x = 0.0$, followed by a horizontal line at $y = 1.0$. The two calculated methods have similar sensitivities and specificities, higher than that for RIA analog; 3.47 pM is the fifth percentile of free T in normal women, as measured by equilibrium dialysis.

respectively. The percentage of subjects with below normal free T levels by equilibrium dialysis that would be classified as having normal free T values using the other methods are as follows: 0.9% with the mass action equation, 0.2% with the direct RIA, and 0.0% with FAI. We again chose 3.47 pM, the fifth percentile of normal values as determined by equilibrium dialysis, to indicate a concentration below which androgen deficiency is present. This demonstrates that a higher percentage of patients would be classified as androgen deficient using the direct RIA or the FAI if assay-specific normal ranges are not developed or concurrent controls are not studied.

Repeatability

Repeatability is the variability of measurements among different sera from each individual subject measured in the same assay. We measured the coefficient of variation, which reflects both intraassay and intrapatient variabilities over time. Data from 72 subjects, each of whom had three serum samples during 1 month, were used. Only data from amenorrheic subjects and subjects receiving continuous estrogen/progestin regimens were used, as serum free T levels were unlikely to change over the course of 1 month in this subset. The coefficient of variation for free T by dialysis was 23.7%, for calculated free T using SHBG by IRMA it was 19.5%, for FAI it was 20.5%, and for direct RIA it was 20.0%. These values reflect variability among the three samples from each patient as well as methodological variability. The coefficients of variation for calculated free T and FAI are highly dependent on the SHBG and total T coefficients of variation. The coefficient of variation for equilibrium dialysis is highly dependent on the total T assay (free T = % free T by dialysis × total T) and the purity of the [^{3}H]T. The purity of the label is important, because if the [^{3}H]T is impure, it will bind SHBG in the sample with lower affinity and result in spuriously high values for free T.

Discussion

We have demonstrated that the calculation of free T, using equations derived from the law of mass action, is fairly comparable to its measurement by equilibrium dialysis. The FAI is a reasonable index of, but not equal to, free T. Because it is an index, it does not indicate the actual physical concentration of free T. By contrast, the direct RIA method fared poorly. It had the highest coefficient of variations and was not sufficiently sensitive to measure reproducibly and accurately the low plasma concentrations of T present in many women with androgen deficiency. It is important to note that the accuracy of both the equilibrium dialysis and the calculation methods depend strongly on the validity of the total T assay, and the accuracy of the calculation also depends on the SHBG assay. Although it was assumed in this study that total T by RIA after column chromatography was the more accurate of the two methods used to measure total T, neither method was validated by another rigorous gravimetric method such as mass spectrometry. Therefore, importantly, none of the methods studied can be considered a gold standard. This is a critical issue because total T as measured by the two different methods correlated poorly. Because there is substantial data in older literature (using methods closely related to the RIA after chromatography) that support the accuracy of such measurements, we think that they are more likely to be correct, or at least closer to the real values. Likewise, two different SHBG assays tested resulted in greatly different free T levels when used in the mass action equation. Because calculated values of free T are comparable to those obtained from equilibrium dialysis, they should prove useful for clinical research and the clinical care of women with androgen deficiency. The direct RIA has little to recommend it, as it is inaccurate and less sensitive than the other methods.

Even the best available methods have limitations, as they do not directly measure free T in serum. The high variability of all of the free T assays reflects their limitations. At this time, one group has reported successfully measuring free T directly in the dialysate after equilibrium (25). However, the extremely low quantities of free T that must be detected make this difficult. Assays that directly measure free T in the dialysate are not commercially available and are not accessible to most clinicians or researchers. Although the use of equilibrium dialysis has no insuperable obstacles, it requires a high degree of technical expertise, exquisite attention to temperature control (if the temperature of the sample increases, the proportion of unbound T will increase), and particular care in the maintenance of dialysis cells. In addition, purification of the label must be ensured by HPLC or column chromatography because an unpure label results in reduced binding of SHBG, leading to a spuriously high free T result. Moreover, the dilution factor, due to the addition of a small volume of tritiated T to the sample (7% of the total serum

volume in this case), is not taken into account routinely in equilibrium dialysis and may be a further source of error. Finally, it should be noted that five of 441 of our samples assayed were outliers, the values of which corrected upon repeat assay. Thus, we presume experimental error was involved in the initial measurement.

Free T calculated from the law of mass action should be accurate when based on valid total T and SHBG assays. However, it should be noted that SHBG binding abnormalities and/or substantial amounts of steroids that compete for binding sites on SHBG could lead to an overestimation of free T. The former has yet to be described, but the latter exists in pregnancy, where presumably high concentrations of steroids that compete for binding sites on SHBG result in a calculated value substantially different from those obtained by equilibrium dialysis. In a group of women in the third trimester of pregnancy (n = 16), the calculated mean free T was significantly lower than that obtained with dialysis (14). The same study reported no such disagreement in a small number of men, patients with hyperthyroidism (a high SHBG state), postmenopausal women, and women with hyperandrogenism (14).

The direct RIA is widely available and frequently used in clinical research (18), but its validity has been previously questioned in small studies (14, 15). The validity of the use of the direct RIA method in clinical research has been called into question because it leads to internally inconsistent conclusions (17). In addition, as is the case for calculations, the direct RIA has not been validated in populations with binding globulin abnormalities, as it has not been shown to be free of protein effects.

The validity of the FAI as an accurate reflection of free T has been challenged. The ratio of FAI to free T by dialysis was shown to vary from 0.12–0.26 in one small study of healthy men and therefore to be an unreliable reflection of the serum free T level (14). Another study demonstrated a low correlation coefficient of 0.435 for FAI *vs.* free T by centrifugal ultrafiltration in another small group of men, but a somewhat better correlation in a small group of women (r = 0.858) (16). In our study there was a good correlation between FAI and free T determined by dialysis or calculation. The FAI can be altered by alterations in either T or SHBG. Unlike free T (whether by calculation or dialysis), using this quotient alone can be misleading. Therefore, the utility of the absolute values is limited.

Androgen deficiency may be a syndrome with important clinical consequences for women, including effects on bone, libido, and quality of life (2, 3, 26–29). Serum total T levels are not sufficient to diagnosis androgen deficiency in women, because dramatic increases in SHBG are common, most frequently due to increases in estrogen from endogenous or exogenous sources (30). This results in relatively lower free T than total T serum levels in these patient populations, as demonstrated by our data. Normative data that address the use of total and/or free T to define this syndrome are not available, but if the example of hirsutism is applicable, free T will be the diagnostic tool of choice. Therefore, accurate measurement of serum free T levels is necessary to diagnose androgen deficiency in women. Our data demonstrate that a calculated value for free T, using the mass action equation, is an acceptable substitute for measurement by equilibrium dialysis in normal women, women with androgen deficiency, and women with variable estrogen levels. It is important to note that the accuracy of both the equilibrium dialysis and the calculation depend strongly on the validity of the total T measurement on which they are based. In addition, the validity of the calculated method depends on the validity of the SHBG assay and standard SHBG binding conditions, which may not be present in all populations of women. The direct RIA for free T did not perform as well as the calculation and is not sufficiently accurate to be used in clinical research. FAI is unitless and correlates well with both the calculated free T and equilibrium dialysis, but does not furnish a physically meaningful result. Normal ranges specific to all assays and methods must be developed, given their dependence on inconstant T and SHBG assays.

Acknowledgments

We thank the staff of the Massachusetts General Hospital General Clinical Research Center for their dedicated patient care, Steven Grinspoon for his helpful comments regarding the manuscript, Alex Vermeulen and Jean M. Kaufman for their assistance in providing further details about the mass action equation that is described in their paper (14), and the patients who participated in the study.

Received April 17, 2003. Accepted October 29, 2003.

Address all correspondence and requests for reprints to: Karen K. Miller, M.D., Neuroendocrine Unit, Bulfinch 457B, Massachusetts General Hospital, Boston, Massachusetts 02114. E-mail: kkmiller@partners.org.

This work was supported in part by NIH Grants M01-RR-01066, M01-RR-01066-27S1, and R03-DK-59297 and FDA Grant FD-R-001981.

References

1. **Miller K, Sesmilo G, Schiller A, Schoenfeld D, Burton S, Klibanski A** 2001 Androgen deficiency in women with hypopituitarism. J Clin Endocrinol Metab 86:561–567
2. **Davis S** 1999 Androgen replacement in women: a commentary. J Clin Endocrinol Metab 84:1886–1891
3. **Snyder P** 2001 Editorial: the role of androgens in women. J Clin Endocrinol Metab 86:1006–1007
4. **Nakagawa K, McNiven N, Forchielli E, Vermeulen A, Dorfman R** 1966 Determination of testosterone by gas-liquid chromatography using an electron capture detector. I. Responses of halo-alkyl derivatives. Steroids 7:329–340
5. **Rivarola M, Migeon C** 1966 Determination of testosterone and androst-4-ene-3,17-dione concentration in human plasma. Steroids 7:103–117
6. **Mayes D, Nugent C** 1968 Determination of plasma testosterone by the use of competitive protein binding. J Clin Endocrinol Metab 28:1169–1176
7. **Horton R, Kato T, Sherins R** 1967 A rapid method for the estimation of testosterone in male plasma. Steroids 10:245–256
8. **Rosenfield R, Eberlein W, Bongiovanni A** 1969 Measurement of plasma testosterone by means of competitive protein binding analysis. J Clin Endocrinol Metab 29:854–859
9. **Rudd B, Rosenfield R, Bongiovanni A, Eberlein W** 1969 The measurement of urinary testosterone glucuronide by competitive protein binding and validation by a double isotope derivative assay. Steroids 13:227–245
10. **Heyns W, Verhoeven G, Van Baelen H, De Moor P** 1969 Measurement of testosterone in plasma by competitive protein binding. Ann Endocrinol 30: 153–158
11. **Furuyama S, Mayes D, Nugent C** 1970 A radioimmunoassay for plasma testosterone. Steroids 16:415–428
12. **Chen J, Zorn E, Hallberg M, Wieland R** 1971 Antibodies to testosterone-3-bovine serum albumin, applied to assay of serum 17-β-ol androgens. Clin Chem 17:581–584
13. **Dufau M, Catt K, Tsuruhara T, Ryan D** 1972 Radioimmunoassay of plasma testosterone. Clin Chim Acta 37:109–116
14. **Vermeulen A, Verdonck L, Kaufman J** 1999 A critical evaluation of simple methods for the estimation of free testosterone in serum. J Clin Endocrinol Metab 84:3666–3672
15. **Winters S, Kelley D, Goodpaster B** 1998 The analog free testosterone assay: are the results in men clinically useful? Clin Chem 44:2178–2182

16. **Kapoor P, Luttrell B, Williams D** 1993 The free androgen index is not valid for adult males. J Steroid Biochem Mol Biol 45:325–326
17. **Rosner W** 1997 Errors in the measurement of plasma free testosterone [Letter]. J Clin Endocrinol Metab 82:2014–2015
18. **Rosner W** 2001 An extraordinarily inaccurate assay for free testosterone is still with us [Letter]. J Clin Endocrinol Metab 86:2903
19. **Nelson J, Tomei R** 1988 Direct determination of free thyroxin in undiluted serum by equilibrium dialysis/radioimmunoassay. Clin Chem 34:1737–1744
20. **Sodergard R, Backstrom T, Shanbhag V, Carstensen H** 1982 Calculation of free and bound fractions of testosterone and estradiol-17β to human plasma proteins at body temperature. J Steroid Biochem 16:801–810
21. **Jordan D, Flood J, Laposata M, Lewandrowski K** 1992 Case records of the Massachusetts General Hospital. Weekly clinicopathological exercises. Normal reference laboratory values. N Engl J Med 327:718–724
22. **Bland J, Altman D** 1986 Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310
23. **Zeger S, Liange K, Albert P** 1988 Models for longitudinal data: a generalized estimating equations approach. Biometrics 44:1049–1060
24. **Rosner B** 1995 Fundamentals of biostatistics, 4th Ed. Belmont, CA: Duxbury Press
25. **Sinha-Hikim I, Arver S, Beall G, Shen R, Guerrero M, Sattler F, Shikuma C, Nelson J, Landgren B, Mazer N, Bhasin S** 1998 The use of a sensitive equilibrium dialysis method for the measurement of free testosterone levels in healthy, cycling women and in human immunodeficiency virus-infected women. J Clin Endocrinol Metab 83:1312–1318
26. **Arlt W, Callies F, van Vlijmen J, Koehler I, Reincke M, Bidlingmaier M, Huebler D, Oettel M, Ernst M, Schulte H, Allolio B** 1999 Dehydroepiandrosterone replacement in women with adrenal insufficiency. N Engl J Med 341:1013–1020
27. **Davis SR, McCloud P, Strauss BJ, Burger H** 1995 Testosterone enhances estradiol's effects on postmenopausal bone density and sexuality. Maturitas 21:227–236
28. **Slemenda C, Longcope C, Peacock M, Hui S, Johnston C** 1996 Sex steroids, bone mass, and bone loss. A prospective study of pre-, peri- and postmenopausal women. J Clin Invest 97:14–21
29. **Miller K** 2001 Androgen deficiency in women. J Clin Endocrinol Metab 86: 2395–2401
30. **Pugeat M, Crave J, Tourniaire J, Forest M** 1996 Clinical utility of sex hormone-binding globulin measurement. Horm Res 45:148–155

JCEM is published monthly by The Endocrine Society (http://www.endo-society.org), the foremost professional society serving the endocrine community.