0021-972X/07/$15.00/0
Printed in U.S.A.

The Journal of Clinical Endocrinology & Metabolism 92(2):405–413

doi: 10.1210/jc.2006-1864

POSITION STATEMENT: Utility, Limitations, and Pitfalls in Measuring Testosterone: An Endocrine Society Position Statement

William Rosner, Richard J. Auchus, Ricardo Azziz, Patrick M. Sluss, and Hershel Raff

St. Luke's/Roosevelt Hospital Center and Columbia University College of Physicians and Surgeons (W.R.), New York, New York 10019; Division of Endocrinology and Metabolism (R.J.A.), University of Texas Southwestern Medical Center, Dallas, Texas 75390; Center for Androgen Related Disorders and Department of Obstetrics-Gynecology (R.A.), Cedars-Sinai Medical Center and Department of Obstetrics-Gynecology and Department of Medicine, The David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, California 90048; Reproductive Endocrine Unit (Department of Medicine) and Department of Pathology (P.M.S.), Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114; and Endocrine Research Laboratory (H.R.), Aurora St. Luke's Medical Center, Division of Endocrinology, Metabolism, and Clinical Nutrition, Medical College of Wisconsin, Milwaukee, Wisconsin 53215

Objective: The objective of the study was to evaluate the current state of clinical assays for total and free testosterone.

Participants: The five participants were appointed by the Council of The Endocrine Society and charged with attaining the objective using published data and expert opinion.

Evidence: Data were gleaned from published sources via online databases (principally PubMed, Ovid MEDLINE, Google Scholar), the College of American Pathologists, and the clinical and laboratory experiences of the participants.

Consensus Process: The statement was an effort of the committee and was reviewed in detail by each member. The Council of The Endocrine Society reviewed a late draft and made specific recommendations.

Conclusions: Laboratory proficiency testing should be based on the ability to measure accurately and precisely samples containing known concentrations of testosterone, not only on agreement with others using the same method. When such standardization is in place, normative values for total and free testosterone should be established for both genders and children, taking into account the many variables that influence serum testosterone concentration. **(*J Clin Endocrinol Metab* 92: 405–413, 2007)**

1. Introduction

The measurement of testosterone (T) in plasma or serum, as done in most laboratories, suffers from a number of serious problems. In women and children, the lack of accuracy and sensitivity has resulted in severely limited utility. For men, most T assays have adequate sensitivity and reasonable clinical utility but are relatively inaccurate. The importance of this issue is highlighted by recent publications in *The Journal of Clinical Endocrinology & Metabolism* including two Original Articles and an Editorial (1), a Position Statement on polycystic ovary syndrome (2), and Clinical Guidelines on Androgen Therapy in Women (3).

The Council of The Endocrine Society, after noting that substantial difficulties exist in the measurement of T in biological fluids, appointed a task force, consisting of the authors of this position paper, to review the problem and make recommendations based on that review. The task force reviewed the literature, gathered data by interview and discussion to assess current practice, evaluated proficiency survey data from the College of American Pathologists (CAP), and came to consensus by both discussion and revision of drafts of the manuscript. The manuscript was reviewed by the Council; their comments were evaluated and included, if appropriate, before finalizing the document.

The problems of sensitivity and specificity of T assays have been addressed by extracting steroids from plasma or serum and separating them chromatographically before subjecting them to immunoassay. These methods are labor intensive and expensive. Hence, high-throughput, relatively inexpensive methods are in wide use that employ whole serum or plasma ("direct" assays) but, for the most part, are too insensitive and inaccurate to measure the total T (TT) concentration in the plasma of women and children. We have the technology to improve the accuracy and precision of T assays and must choose these properties over simplicity and economy.

2. Background

Assays for T in plasma, and their evaluation, pose a number of challenges:

- TT concentrations in plasma vary over 3 orders of magnitude depending on age, gender, and the presence of disease.

First Published Online November 7, 2006

Abbreviations: bio-T, Bioavailable T; CAP, College of American Pathologists; ED, equilibrium dialysis; FAI, free androgen index; FT, free T; ID/GC-MS, isotope-dilution gas chromatography-MS; Kd, dissociation constant; LC/MS-MS, liquid chromatography/MS-MS; MS, mass spectrometry; MS-MS, tandem MS; PCOS, polycystic ovary syndrome; T, testosterone; TT, total T.

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

- The concentration of TT varies with time of day.
- Other steroids of similar structure and abundance in the circulation lead to assay interference.
- Only 1–3% of T is not bound to plasma proteins, raising questions about whether TT or free T (FT) is the most clinically useful measure.
- Age- and gender-corrected normal ranges, using a standardized assay, are generally lacking.
- There is no universally recognized T-calibrating standard.

3. Methods for Measuring TT

The commonly used methods for measuring TT and their strengths and shortcomings are summarized in Table 1. Following, we briefly detail these methods.

a. Immunoassay methods for measuring TT

RIAs and chemiluminescence immunoassays are the most widely used methods for measuring plasma TT. These assays are performed directly on serum or plasma or after extraction and/or chromatography. The more labor-intensive assays that incorporate extraction and chromatography offer several advantages, including removal of interfering proteins, separation of cross-reacting steroids, and use of large serum aliquots to increase sensitivity. Such assays are more accurate and sensitive than direct assays but still require proper validation.

b. Mass spectrometry (MS) methods for measuring TT

MS both identifies and quantifies the analyte and, for TT, routinely incorporates extraction and chromatography before assay (4–9); the specificity of this method has been enhanced by tandem MS (MS-MS), which still must be validated for accuracy, sensitivity, and precision.

c. Comparison of TT assay methods

The CAP administers a quality-control program that distributes blinded samples to participating laboratories and gauges accuracy relative to others using the same methodology. The samples they distribute are not in plasma but in material that allows the samples to be stable, although not frozen, and hence more easily distributable to large numbers of laboratories. Assay results may very well be influenced by this artificial matrix. Table 2 shows the CAP results for three samples of TT expected for a normal woman [no. 1; 33 ± 11 ng/dl (1.1 ± 0.4 nM)], a hypogonadal man or an androgenized woman [no. 2; 97 ± 31 ng/dl (3.4 ± 1.1 nM)], and a normal man [no. 3; 465 ± 81 ng/dl (16.1 ± 2.8 nM)]. For no. 1, the coefficients of variation within the same methodology performed at different laboratories ranged from 13–32%. Results for the same sample using the same method varied 2- to 6-fold, demonstrating the unacceptable reliability of these methods for measuring TT in normal women. Examination of "All Instruments" for no. 1 indicates that clinical usefulness is severely compromised. As the concentration of T increases (no. 2 and 3), the coefficients of variation within a methodology decrease. However, there is still considerable variability and lack of standardization between methods as demonstrated by, for example, the range of values for no. 2 [45–365 ng/dl (1.6–12.7 nM)]. Approximately one third of all the laboratories used the same instrument and about two

TABLE 1. Comparison of methods available for measuring TT in the circulation

Method	Strengths	Shortcomings
Direct assay by RIA, ELISA, or CLIA	• Technically simple, rapid, and relatively inexpensive • High throughput and fast turnaround time • Can be automated	• T concentration often overestimated • Susceptible to matrix effects • Not standardized; results and reference intervals are method dependent • Limited accuracy at T < 300 ng/dl • Reference intervals in different populations not well documented • For RIA: generates radioactive waste
RIA after extraction and chromatography	• Extensively used, with well-documented reference intervals in different populations • Relatively large serum volumes can be used for the assay, increasing sensitivity • Potential to assay multiple steroids separated by the chromatography in the same sample aliquotfacilities and waste disposal • T released from steroid-binding proteins during extraction	• Labor intensive, cumbersome, time consuming, and costly • Requires a high degree of technical expertise • Use of organic solvents requires special facilities and waste disposal • Susceptible to matrix effects • Imprecise: measurements must be corrected for recovery • Generates radioactive waste
MS, after extraction and liquid (LC) or gas chromatography (GC)	• Multiple steroids can be measured in the same sample aliquot • Highly accurate when properly validated • Throughput comparable with RIA after extraction and chromatography	• Relatively expensive • Currently standardization still lacking • Limited throughput and relative long turnaround times • Derivatization steps can introduce additional error • Use of organic solvents requires special facilities and waste disposal

CLIA, Chemiluminescent immunoassay. Adapted in part from Ref. 46.

TABLE 2. Selected CAP proficiency survey results for TT (ng/dl)

Instrument	No. of labs	Mean/SD (ng/dl)	CV (%)	Median (ng/dl)	Low/high value (ng/dl)
Test sample 1					
All instruments	**1108**	**32.7/11.4**	**34.9**	**31**	**7–100**
Abbott Architect	5			26	17–32
Bayer ACS: 180	19	30.9/9.5	30.7	34	13–43
Bayer ADVIA Centaur	349	30.3/8.6	28.4	30	9–53
Beckman Acess/2	140	32.4/4.3	13.4	32	22–44
Beckman UniCel Dxl	73	31.3/5.3	17.0	31	12–43
DPC Coat-a-Count	40	27.1/4.2	15.4	26	20–36
DPC Immulite 1000	126	47.8/9.3	19.4	49	26–72
DPC Immulite 2000	9			52	31–56
DPC Immulite 2500	59	51.2/9.2	18.0	51	31–77
Roche Elecsys 1010/2010	70	25.1/7.0	27.7	24	7–43
Roche Elecsys/E170	72	31.2/4.7	15.1	31	20–43
Tosoh AIA-Pack	12	43.8/14.0	32.0	43	17–71
Vitros Eci	83	18.4/2.7	14.5	18	13–26
MS	5	31.8/-		33	27–37
Diagnostic Sys Liquid	5			25	24–33
Diagnostic Sys Solid	5			20	8–23
Test sample 2					
All instruments	**1133**	**97.1/31.3**	**32.2**	**87**	**45–365**
Abbott Architect	8			75	66–108
Bayer ACS: 180	23	97.6/14.1	14.5	95	64–122
Bayer ADVIA Centaur	358	96.9/10.8	11.1	97	65–130
Beckman Acess/2	150	76.8/5.8	7.6	77	62–94
Beckman UniCel Dxl	57	71.4/6.2	8.7	72	56–87
DPC Coat-a-Count	42	79.9/8.0	10.4	76	65–92
DPC Immulite 1000	60	147.1/19.6	13.3	146	103–197
DPC Immulite 2000	133	154.3/17.0	11.0	153	120–200
DPC Immulite 2500	5			157	151–195
Roche Elecsys 1010/2010	82	69.7/10.0	14.3	68	55–105
Roche Elecsys/E170	66	81.1/7.3	9.0	81	60–102
Tosoh AIA-Pack	12	87.8/12.3	14.0	89	71–108
Vitros Eci	85	78.3/6.2	7.9	78	64–91
MS	5	68.6/6.1	8.9	69	60–77
Test sample 3					
All instruments	**1135**	**464.9/80.6**	**17.3**	**449**	**276–744**
Abbott Architect	8			383	353–395
Bayer ACS: 180	23	439.8/42.7	9.7	440	344–509
Bayer ADVIA Centaur	359	424.1/42.6	10.0	421	328–546
Beckman Acess/2	152	402.3/21.6	5.4	402	338–473
Beckman UniCel Dxl	57	377.1/23.2	6.2	379	332–428
DPC Coat-a-Count	42	413.3/35.7	8.6	410	324–516
DPC Immulite 1000	60	550.3/60.5	11.0	546	436–673
DPC Immulite 2000	133	566.0/59.7	10.5	563	423–744
DPC Immulite 2500	5			635	546–667
Roche Elecsys 1010/2010	81	511.8/28.7	5.6	509	451–589
Roche Elecsys/E170	69	550.6/25.7	4.7	546	501–626
Tosoh AIA-Pack	11	636.9/44.8	7.0	652	555–706
Vitros Eci	84	519.4/26.0	5.0	519	453–581
MS	5	354.4/45.4	12.8	365	281–395

The CAP distributed three different unknown samples to the indicated number of laboratories. The measured TT results were returned to CAP and summarized by them (reproduced here in modified form with permission from CAP). Test samples 1, 2, and 3 contain, respectively, concentrations of T similar to those in the plasma of normal women, hypogonadal men, or androgenized women and normal men. Only methods in use in a sufficient number of reporting laboratories (No. of labs) are shown; some do not have sufficient numbers to calculate reliably mean and SD. The coefficient of variation (CV) gives an estimate of the variability both within and between methodologists (All instruments). The CV (Mean/SD) is unitless but is commonly multiplied by 100 and reported as a percent. Also notice that, in some cases, the ranges (Low/high value) are lower or higher than those shown for individual methods. That is because there are several participants in the CAP survey that do not use one of the major methods or instruments listed. To convert to nanomolar, multiply by 0.03467.

thirds used one of three instruments (Table 2). The widespread use of a limited number of methodologies gives promise that standardization is achievable. There are comparable problems with variability and a lack of standards for accuracy for FT.

More reliable MS methods for TT have begun to appear (10–12), although only five of more than 1100 participating laboratories used MS in late 2005. However, the values available from mass spectrometric measurements are consistent with reports of larger bias at low concentrations of TT (7, 9). We propose that the best prospect for a gold standard lies in extraction and chromatography followed by MS or MS-MS in which the chemical structure of the molecule measured is identified.

Taieb *et al.* (7) compared 10 direct commercial immunoassays with isotope-dilution gas chromatography-MS (ID/

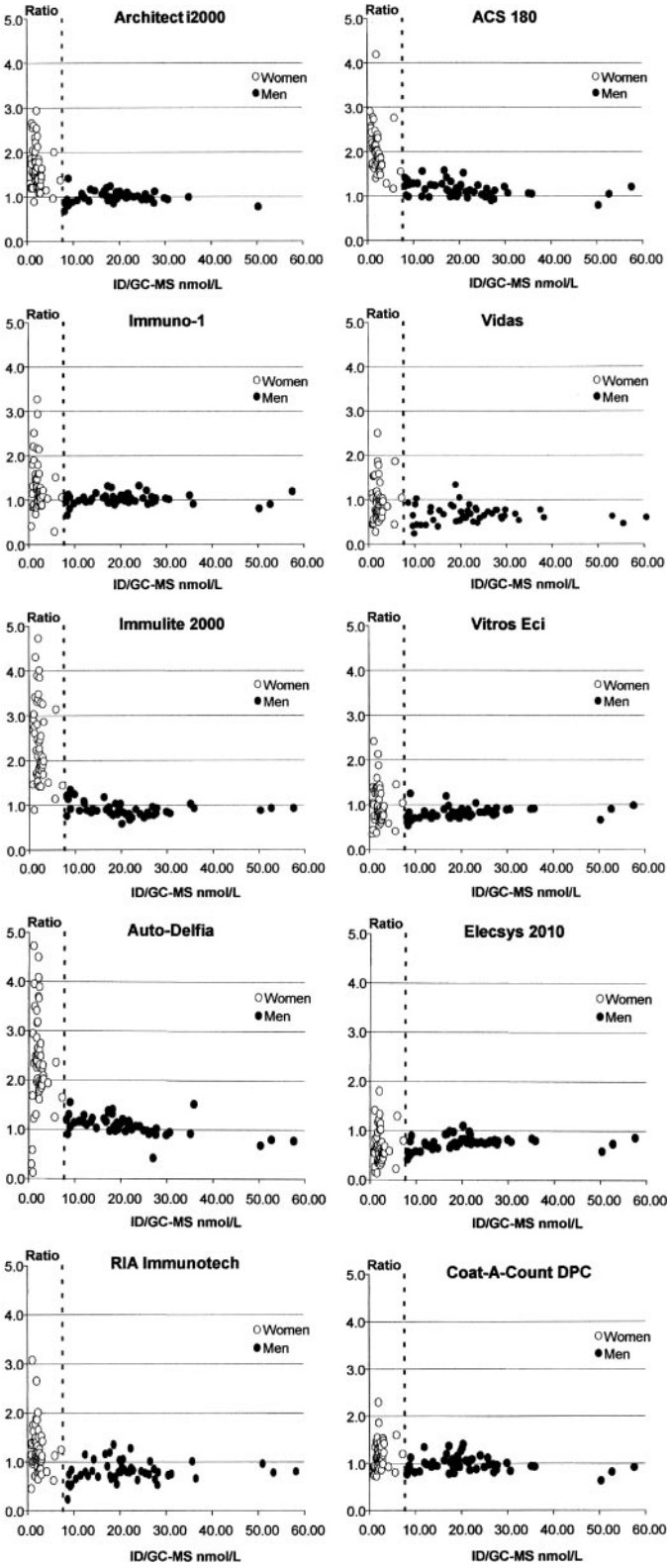

FIG. 1. T concentrations in women (n = 55; ○) and men (n = 50; ●) obtained by 10 immunoassays and ID/GC-MS. The abcissa is the T concentration (nanomoles per liter) measured by ID/GC/MS. (To convert to nanograms per deciliter, multiply by 28.8.) The y-axis is the ratio of T concentration determined by immunoassay divided by T concentration determined by ID/GC-MS. The *vertical dotted line* separates the data for men and women. [Reprinted with permission from Taieb J, Mathian B, Millot F, Patricot MC, Mathieu E, Queyrel N, Lacroix I, Somma-Delpero C, Boudou P 2003 Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women, and children. Clin Chem 49:1381–1395 (7).]

GC-MS) and also compared samples from 55 women by ID/GC-MS with an extraction and chromatography RIA (13) (Fig. 1). Below approximately 8.0 nM (230 ng/dl), the methods disagreed by up to 5-fold, with immunoassays generally overestimating the T concentration. Some of the methods were better than others, but even the best method showed up to a 2-fold higher T concentration by immunoassay in women. In men, seven of the 10 tested assays had highly statistically significantly different medians compared with ID/GC-MS. The salutary outcome is that three of the commercial assays had medians not significantly different from those measured by the MS-based method in men. Even this salve contains an irritant because ". . . this statistical analysis compared the differences between medians and did not address the scatter of the results between each immunoassay and ID/GC-MS." Thus, concern remains that even the good assays in men are only good on average.

Wang *et al.* (9) obtained results much like those in Fig. 1 by comparing six different direct immunoassays to liquid chromatography/MS-MS (LC/MS-MS). Two of the assays tested were the same as shown in Fig. 1 and four were unique to this study. For T less than 150 ng/dl (<5.2 nmol/liter), the values were neither analytically nor clinically useful. For higher T concentrations, the values have some utility, but the discrepancies among methods are unacceptable. There are three additional studies that support these findings (14–16) and apparently none that contradict them.

4. Methods for Measuring FT

T circulates bound to at least two plasma proteins, SHBG and albumin (17). That which is unbound, FT, is often considered the component that has access to the cell and results in androgenic effects. The situation is more complicated than that (18, 19), but as a practical matter, FT often correlates better with the androgenic state of the patient than does TT (20). In addition, there exists the concept of bioavailable T (bio-T), defined as the concentration of T that is free, plus that which is weakly bound, *e.g.* albumin bound. This is a widely used measurement that we will discuss in concert with the discussion of FT. The commonly used methods for measuring FT, and their strengths and weaknesses, are summarized in Table 3.

a. Measuring FT

An appropriate assay for TT is necessary but not sufficient for the measurement of FT. Because FT is such a tiny portion of TT, an assay that is accurate and precise when measuring very small amounts of T is required. The indirect measurement of FT is accomplished by adding ^{3}H-T to the sample to be assayed and, after equilibrium has been attained, separating bound from free ^{3}H-T, and then measuring free ^{3}H-T. The fraction of free ^{3}H-T is multiplied by the amount of TT, obtained in a separate assay from the same plasma. The

major potential problem with such assays is the possible presence of radiochemical impurities. For example, a 2% contamination with a radioactive substance that does not bind to proteins will cause an apparent doubling of the FT in which 2% of T actually was free. If both direct and indirect methods are done well, they should yield the same answer.

b. Data on FT measurement

The values of FT by equilibrium dialysis (ED) are influenced by a number of variables, the most important of which is the assay for T. However, the details of how ED is done (21) as well as the population being studied all have an impact on

TABLE 3. Comparison of methods available for measuring FT, unbound T, or bio-T in the circulation

Method	Strengths	Shortcomings
Direct RIA	• Simple, rapid, and relatively inexpensive • Requires minimum technical expertise • Can be automated	• Poor accuracy, sensitivity, and between-laboratory comparability: ○ Major biasing effects due to dilution of serum samples ○ Significant binding of the analog to serum proteins ○ Lack of parallelism between measurements of serially diluted serum samples and FT
Physical separation of protein-bound from FT[a]	• Relatively accurate (the equilibrium dialysis assay is considered the gold standard method for quantifying FT) • Relatively sensitive and reproducible	• Relatively expensive • Technically cumbersome and difficult ○ Equilibrium dialysis is influenced by dilution of the serum sample ○ Centrifugal ultrafiltration is subject to adsorption of testosterone to the membrane and difficulty with temperature control ○ Both the dialysis and ultrafiltration methods can be affected by ^{3}H-labeled impurities bound differently from T by SHBG and/or albumin • As for all indirect measures, highly dependent on the accuracy of the TT assay • At this time, none of the methods are sensitive enough to accurately measure free T directly in women and children
Ammonium sulfate precipitation to measure bio-T	• Technically simple	• Can be inaccurate due to: ○ Use of impure ^{3}H-T ○ Incomplete precipitation of globulins ○ Lack of uniformity of methodology between labs
Calculation: The FAI (T/SHBG)	• Simple • Good correlation with physical separation measures in women	• Poor correlation with physical separation measures in men • Highly dependent on the accuracy and sensitivity of the total T and SHBG assays
Calculation: using algorithms based on law of mass action[b]	• Simple • Excellent correlation with physical separation measures	• Highly dependent on the accuracy and sensitivity of the total T and SHBG assays • Assumptions and reference intervals not standardized
Calculated: using empirical equations	• Excellent correlation with physical separation measures • Relatively sensitive	• Equations are derived from computer modeling based on known concentrations of T, SHBG, and FT obtained in individual laboratories • Hundreds to thousands of samples are needed to generate the equations • Lack of transportability of the equations among laboratories

Adapted in part from Refs. 36, 46, and 47. FAI, Free androgen index.

[a] Separation by means of a membrane (*i.e.* ED) or a filter (*i.e.* centrifugal ultrafiltration). Either use ^{3}H-T and multiply by total T concentration (obtained in separate assay) to measure percent FT or measure FT directly.

[b] Requires total T and SHBG (and possibly albumin) concentrations and the Kd between T and SHBG and albumin.

the results. Before citing some examples of how FT is measured, our agreement with the following should be noted: ". . . it is clear that even the best available and scrupulously performed measurement procedures have technical and fundamental limitations and that, consequently, the scientific community will have to accept that there will remain a degree of arbitrariness about the best way to measure free hormone concentrations" (8). We believe that the degree of arbitrariness can be small and that the best approaches approximate the FT concentration.

c. Direct and indirect measurement of FT

Both Adachi *et al.* (22) and Van Uytfanghe *et al.* (8) measured FT by assaying for T in an ultrafiltrate of plasma or serum. The former study gave inadequate attention to the measurement of TT, thus invalidating the use of their results for this stringent review. The latter used a method based on GC-MS for the measurement of TT and paid close attention to the validation of their method. Even then, these investigators concluded that further sensitivity must be attained for the method to be useful in women.

To our knowledge, there is only one major study in which T was measured directly when FT was separated by ED (23). Although the work was meticulously performed, T was measured by a direct RIA using ^{125}I-T as tracer, and the method was not validated against a criterion (gold) standard method. All of the other communications in which FT was measured by ultrafiltration or ED used the indirect method. Thus, when FT obtained by ED is compared with other methods, we are considering studies that use the indirect method. Otherwise, studies that measure FT do so by calculating it, using a surrogate for it [either the free androgen index (FAI) or bio-T], or a direct assay for FT. There is considerable confusion in the use of terms. For example, the term FT was used when what was reported was a surrogate (24, 25) or the calculated FAI was specified in the title and then reported as FT in nmol/liter (26).

d. Measurement of bioavailable T (bio-T)

Bio-T is albumin bound plus FT, the fraction thought to be available to tissues. It is measured by adding ^{3}H-T to serum and precipitating SHBG bound ^{3}H-T with ammonium sulfate. The fraction of ^{3}H-T not precipitated is used to calculate bio-T by multiplying by the TT obtained in a separate assay. Variations in precipitation and assay methodology makes comparison of results between different communications difficult (27). Furthermore, the concept itself may be misleading and confusing. Despite this, bio-T has been reported to correlate with FT by ED from fairly (28) to extremely well (29) and to be a useful index of some biological changes (28).

e. The Free Androgen Index (FAI)

The FAI is the unitless quotient T/SHBG and depends on appropriate measurements for T and SHBG. Therefore, there is a reasonable correlation between FAI and FT, particularly in women. However, FT depends on not only the ratio but also the absolute concentration of both T and SHBG (30). Thus, the correlation will depend on these variables and will be biased, depending on the concentration of T as, at lower levels, measurements are less precise. This is illustrated by the range of correlations obtained by different investigators (29, 31, 32). Having measured both T and SHBG, FT should be calculated, which is easily done using a fixed formula in a spreadsheet or using a FT calculator (*e.g.* http://www.issam.ch/freetesto.htm).

f. One-step direct assay for FT

This assay, which uses an ^{125}I-T analog, has been consistently found to be inaccurate (29, 32, 33), and its use is highly questionable.

g. Calculation of FT concentration

One can also use the law of mass action to calculate concentrations of T that are free or bound to SHBG and albumin (29, 34). The calculation depends on the measurement of TT, total SHBG and total albumin, and the use of the equilibrium dissociation constants (Kd) for the binding of SHBG and T and albumin and T. Within rather broad limits, the concentration of the low affinity binding protein, albumin, varies too little to significantly affect FT levels (29). Although the Kd for SHBG-T is about 1 nM, this number needs to be verified and universally agreed on. In addition, a universal standard for SHBG has not been agreed on. With those caveats, calculation of FT is the most useful estimate of FT in plasma (29, 32) except in pregnancy (29).

Calculation of FT compares extraordinarily well with FT measured by ED (29, 32, 35). Miller *et al.* compared calculation with ED in more than 400 women with a variety of disorders and looked at five clinical subgroups (32). The correlation coefficients between ED and calculation were greater than 0.96; the intercepts were not different from zero, but the slopes indicated a 20% bias between the measured and calculated values due either to a systematic error in the method used for ED or in the calculation. The error in the calculation, in turn, could arise from using the wrong Kd, errors in the measurement of SHBG, or both. Studies comparing calculation with ultrafiltration mostly (36), but not always (8), show excellent correlation.

5. Uses of T Assays

For scientific, analytical purposes, T assays need to be as accurate and reproducible as possible and as sensitive as necessary for the job at hand. We must be able to compare studies from multiple laboratories using different methods. Clinically used assays must be held to the same high standard. The diagnosis and management of the hyperandrogenic woman, or the hypoandrogenic woman or man, will depend to a large degree on highly accurate androgen measurements. In addition, because there is a diurnal variation in plasma T that may be superimposed on variations over smaller time intervals, samples should be multiple (three should suffice) and be obtained between 0800 and 1000 h.

a. Evaluating adult males

The most common use of clinical T assays is to diagnose hypogonadism in men for which almost any assay will do.

Whether the subtle decrease in plasma T in the aging male is normal or represents hypogonadism will not be answered with the use of almost any assay. Furthermore, the evaluation of the risks and benefits of T replacement requires sensitive and accurate assays. When the TT lies near the lower limit of the normal range, a calculated FT may prove useful.

b. Evaluating adult females

The measurement of T in women is used for evaluating states of androgen excess to both exclude androgen-producing tumors and help in the diagnosis of other hyperandrogenic states. Most commercial assays are adequate for identifying, but not accurately quantifying, elevated TT in women. However, these assays frequently fail to detect the moderately androgenized patient, *e.g.* most patients with polycystic ovary syndrome (PCOS) (20). FT, or one of its surrogates, correlates better with the clinical presentation of these patients than does TT. FT measurements may be the most sensitive marker of hyperandrogenemia; they are above the upper normal limit in 60–70% of women with clinical signs and symptoms of hyperandrogenism (20).

The influence of T on female sexual desire and sense of well-being has received considerable attention; two studies have shown no correlation between circulating T concentrations and female sexual function (37, 38), although there is evidence that T replacement in ovariectomized premenopausal women (39) and those with hypopituitarism (40) confers some benefit. The normative values for T and FT across a woman's life span are not adequate. Thus, the measurement of serum T for the evaluation of poor libido in women is unlikely to be informative, and we recommend that such measurements not be used for this purpose until improved methods are available.

c. Evaluating children

In boys, TT measurements are used during adolescence in the evaluation of early or delayed puberty or at birth during the evaluation of undervirilized males. In girls, TT assays are used to assess and treat disorders of sexual development and in the evaluation of contrasexual pubertal development. As in women, TT determination in children should be carried out only with assays of sufficient sensitivity and in conjunction with appropriate normative data. FT in children is of limited value.

6. Normal Ranges for T

a. In adult males

Although the measurement of TT in normal men does not represent a problem in sensitivity, a precise definition of the lower limit of normal TT for adult males remains elusive. Because the clinical presentation of hypogonadism is highly variable, particularly in the setting of comorbid conditions, hypogonadism is often a laboratory rather than clinical diagnosis. TT greater than 320 ng/dl (11.1 nM) is considered normal (41). TT less than 200 ng/dl (6.9 nM) is diagnostic of hypogonadism, but TT 200–320 ng/dl (6.9–11.1 nM) is equivocal. The agreement among platform assays is marginal in the difficult range of 200–320 ng/dl (6.9–11.1 nM) in which a difference of 10% might alter clinical decisions. Although standardization (or replacement) of platform assays with MS-based methods holds the promise of obviating the assay-based confusion at the lower end of the normal range, the variability in FT consequent to alterations in SHBG must be considered. Because T secretion is pulsatile and varies diurnally, more than a single measurement is sometimes required to make a therapeutic decision.

For values in the equivocal zone, the determination of FI or bio-T is recommended to distinguish eugonadism from hypogonadism. An FT of 6.5 ng/dl (0.23 nM) and a bio-T of 150 ng/dl (5.2 nM) are considered the lower limits of normal (41). Measurement of FT or bio-T does not avoid the problem of TT assay standardization, because both use TT as part of the measurement.

b. In adult women

The need for defining an accurate lower range for T in women has recently become significant. Platform and conventional RIAs are unreliable in this range, whereas immunoassay after extraction and chromatography or LC/MS-MS appears capable of yielding meaningful data.

In constructing normal ranges, care must be taken to exclude subjects with PCOS, or other forms of androgen excess. T distributions are bimodal in families of PCOS subjects (42) and normal ranges show a tail at the high end of the distribution. T in women varies not only with the menstrual cycle but also with age, race, and body mass index (38).

The FAI is often used as a surrogate for FT, and the FAI correlates well with FT in women (32) but not men. Because T production is regulated by gonadotropin feedback in men, changes in SHBG, which alter FT concentrations, will be compensated by autoregulation of T production but not so in women. In addition, much circulating T in women is derived from the peripheral conversion of adrenal dehydroepiandrosterone and dehydroepiandrosterone sulfate (43) that also is not subject to feedback control. Because SHBG is present in such large excess in women (10–100:1), FT concentrations are driven primarily by SHBG abundance. In addition, T excess in women lowers SHBG concentrations, which raises the FT concentration and contributes to the strong correlation of 1/SHBG with FT.

c. In children

The testes secrete large amounts of T during the first year of life, but gonadal steroidogenesis is very low in both boys and girls thereafter until the start of puberty. Consequently, T concentrations are extremely low throughout childhood; the measurement of T in children poses the same problems as those in women. Consequently, assay of T in children should use immunoassay after extraction and chromatography or LC/MS-MS (5). One recent report indicates that derivatization before LC/MS-MS improves assay characteristics (4). Total and FT reference intervals must specify gender, age, and Tanner stage, as has recently been done (5) because all these variables influence T concentrations. Normative data for infants are difficult to obtain, so historical data are used (44, 45).

7. Summary of Key Findings and Recommendations

This review demonstrates that the manner in which most assays for TT and FT are currently performed is decidedly unsatisfactory. The technology exists to perform accurate, precise, and reproducible assays for T, and we should move forward to ensure that these assays become the standard by which all assays are validated. We have summarized our findings in Tables 1 and 3.

- Our most salient recommendation is: laboratory proficiency testing should be based on the ability to accurately and precisely measure a sample containing a known concentration of T and not only on agreement with peers using the same method. When such standardization is in place, normative values for TT and FT should be established taking into account all the appropriate variables, *e.g.* gender, age, race, stage of puberty, time of day, *etc*. We believe that this goal can be accomplished. It has been done for cholesterol.

In the interim we offer the following recommendations to physicians ordering and using androgen assays:

- Know the type and quality of the assay that is being used and the properly established and validated reference intervals for that assay. Reference intervals should be established by each laboratory in collaboration with endocrinologists, using well-defined and characterized populations.
- In the absence of other information, direct assays (those performed on whole serum) perform poorly at low T concentrations (*i.e.* in women, children, and hypogonadal men) and should be avoided. Assays after extraction and chromatography, followed by either MS or immunoassay, are likely to furnish more reliable results and are currently preferred.
- Assays for T may behave differently in controls and affected individuals, perhaps reflecting differences in the endocrine milieu of patients.
- Most assays will distinguish between T concentrations in classic hypogonadism and those in normal men. Serum TT, preferably obtained on more than one morning sampling, is the recommended screening test for hypogonadism.
- Assuming a high-quality assay and well-defined reference intervals, a serum TT, preferably drawn during the early follicular phase of the menstrual cycle, is recommended as the initial test in seeking out androgen-producing tumors in women.
- Calculated FT, using high-quality T and SHBG assays with well-defined reference intervals, is the most useful, clinically sensitive marker of hyperandrogenemia in women and can be used in concert with clinical end points in the diagnosis and follow-up of such patients.
- In the absence of pituitary insufficiency, the use of T assays in the evaluation of sexual dysfunction or fatigue in adult women is not supported by published evidence and is strongly discouraged.
- In children, reference intervals must be adjusted for gender, age, and stage of adolescent development and must be specific for the assay method, until a universal standard is available.
- FT measurements in children are of limited value. Evaluations of androgen excess, virilization, intersex disorders, or contrasexual maturation are the only indications for T measurement in girls. Several indications exist for T measurements in boys, including assessment of gonadal failure, disorders of sexual development or puberty, and monitoring response to treatment.

Acknowledgments

Received August 24, 2006. Accepted October 30, 2006.

To whom correspondence and reprint requests should be addressed: William Rosner, M.D., Department of Medicine, St. Luke's/Roosevelt Hospital Center, 1000 Tenth Avenue, AJA 403, New York, New York 10019. E-mail: wr7@columbia.edu.

Disclosures: W.R. has previously consulted for Solvay. R.J.A. received lecture fees from Columbia Laboratories. R.A. consults for Procter and Gamble, Merck, and Quest Diagnostics. P.M.S. consults for Diagnostic Systems Laboratories and Diagnostic Products Corp. and received lecture fees from Fujirebio, Abbott Diagnostics, Bayer Diagnostics, and Roche Diagnostics and grant support (2005 to present) from Roche Diagnostics. H.R. has nothing to declare.

References

1. **Matsumoto AM, Bremner WJ** 2004 Serum testosterone assays—accuracy matters. J Clin Endocrinol Metab 89:520–524
2. **Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, Janssen OE, Legro RS, Norman RJ, Taylor AE, Witchel SF** 2006 Position statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society Guideline. J Clin Endocrinol Metab 91:4237–4245
3. **Wierman ME, Basson R, Davis SR, Khosla S, Miller KK, Rosner W, Santoro N** 2006 Clinical practice guideline: androgen therapy in women: an Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 91:3697–3710
4. **Dorgan JF, Fears TR, McMahon RP, Aronson FL, Patterson BH, Greenhut SF** 2002 Measurement of steroid sex hormones in serum: a comparison of radioimmunoassay and mass spectrometry. Steroids 67:151–158
5. **Kushnir MM, Rockwood AL, Roberts WL, Pattison EG, Bunker AM, Fitzgerald RL, Meikle AW** 2006 Performance characteristics of a novel tandem mass spectrometry assay for serum testosterone. Clin Chem 52:120–128
6. **Rauh M, Groschl M, Rascher W, Dorr HG** 2006 Automated, fast and sensitive quantification of 17α-hydroxy-progesterone, androstenedione and testosterone by tandem mass spectrometry with on-line extraction. Steroids 71:450–458
7. **Taieb J, Mathian B, Millot F, Patricot MC, Mathieu E, Queyrel N, Lacroix I, Somma-Delpero C, Boudou P** 2003 Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women, and children. Clin Chem 49:1381–1395
8. **Van Uytfanghe K, Stockl D, Kaufman JM, Fiers T, Ross HA, De Leenheer AP, Thienpont LM** 2004 Evaluation of a candidate reference measurement procedure for serum free testosterone based on ultrafiltration and isotope dilution-gas chromatography-mass spectrometry. Clin Chem 50:2101–2110
9. **Wang C, Catlin DH, Demers LM, Starcevic B, Swerdloff RS** 2004 Measurement of total serum testosterone in adult men: comparison of current laboratory methods versus liquid chromatography-tandem mass spectrometry. J Clin Endocrinol Metab 89:534–543
10. **Thienpont LM, Van NB, Stockl D, Reinauer H, De Leenheer AP** 1996 Determination of reference method values by isotope dilution-gas chromatography/mass spectrometry: a five years' experience of two European Reference Laboratories. Eur J Clin Chem Clin Biochem 34:853–860
11. **Cawood ML, Field HP, Ford CG, Gillingwater S, Kicman A, Cowan D, Barth JH** 2005 Testosterone measurement by isotope-dilution liquid chromatography-tandem mass spectrometry: validation of a method for routine clinical practice. Clin Chem 51:1472–1479
12. **Chang YC, Li CM, Li LA, Jong SB, Liao PC, Chang LW** 2003 Quantitative measurement of male steroid hormones using automated on-line solid phase extraction-liquid chromatography-tandem mass spectrometry and comparison with radioimmunoassay. Analyst 128:363–368
13. **Fiet J, Gosling JP, Soliman H, Galons H, Boudou P, Aubin P, Belanger A, Villette JM, Julien R, Brerault JL** 1994 Hirsutism and acne in women: coordinated radioimmunoassays for eight relevant plasma steroids. Clin Chem 40:2296–2305
14. **Stanczyk FZ, Cho MM, Endres DB, Morrison JL, Patel S, Paulson RJ** 2003

Limitations of direct estradiol and testosterone immunoassay kits. Steroids 68:1173–1178
15. **Boots LR, Potter S, Potter D, Azziz R** 1998 Measurement of total serum testosterone levels using commercially available kits: high degree of between-kit variability. Fertil Steril 69:286–292
16. **Sikaris K, McLachlan RI, Kazlauskas R, de Kretser D, Holden CA, Handelsman DJ** 2005 Reproductive hormone reference intervals for healthy fertile young men: evaluation of automated platform assays. J Clin Endocrinol Metab 90:5928–5936
17. **Rosner W** 1999 Sex hormone-binding globulin. In: Knobil E, Neill JD, eds. Encyclopedia of reproduction. San Diego: Academic Press; 471–475
18. **Mendel CM** 1989 The free hormone hypothesis: a physiologically based mathematical model. Endocr Rev 10:232–274
19. **Mendel CM** 1992 The free hormone hypothesis. Distinction from the free hormone transport hypothesis. J Androl 13:107–116
20. **Chang WY, Knochenhauer ES, Bartolucci AA, Azziz R** 2005 Phenotypic spectrum of polycystic ovary syndrome: clinical and biochemical characterization of the three major clinical subgroups. Fertil Steril 83:1717–1723
21. **Westphal U** 1971 Steroid-protein interactions. New York: Springer-Verlag; 1–567
22. **Adachi K, Yasuda K, Fuwa Y, Goshima E, Yamakita N, Miura K** 1991 Measurement of plasma free steroids by direct radioimmunoassay of ultrafiltrate in association with the monitoring of free components with [14C]glucose. Clin Chim Acta 200:13–22
23. **Sinha-Hikim I, Arver S, Beall G, Shen RQ, Guerrero M, Sattler F, Shikuma C, Nelson JC, Landgren BM, Mazer NA, 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
24. **Harman SM, Metter EJ, Tobin JD, Pearson J, Blackman MR** 2001 Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. J Clin Endocrinol Metab 86:724–731
25. **Hogervorst E, Bandelow S, Combrinck M, Smith AD** 2004 Low free testosterone is an independent risk factor for Alzheimer's disease. Exp Gerontol 39:1633–1639
26. **Sutton-Tyrrell K, Wildman RP, Matthews KA, Chae C, Lasley BL, Brockwell S, Pasternak RC, Lloyd-Jones D, Sowers MF, Torrens JI** 2005 Sex hormone-binding globulin and the free androgen index are related to cardiovascular risk factors in multiethnic premenopausal and perimenopausal women enrolled in the Study of Women Across the Nation (SWAN). Circulation 111:1242–1249
27. **Emadi-Konjin P, Bain J, Bromberg IL** 2003 Evaluation of an algorithm for calculation of serum "bioavailable" testosterone (BAT). Clin Biochem 36:591–596
28. **Morley JE, Patrick P, Perry III HM** 2002 Evaluation of assays available to measure free testosterone. Metabolism 51:554–559
29. **Vermeulen A, Verdonck L, Kaufman JM** 1999 A critical evaluation of simple methods for the estimation of free testosterone in serum. J Clin Endocrinol Metab 84:3666–3672
30. **Rosner W** 1975 Some theoretical considerations regarding filter disk assays. Anal Biochem 67:422–427
31. **Kapoor P, Luttrell BM, Williams D** 1993 The free androgen index is not valid for adult males. J Steroid Biochem Mol Biol 45:325–326
32. **Miller KK, Rosner W, Lee H, Hier J, Sesmilo G, Schoenfeld D, Neubauer G, Klibanski A** 2004 Measurement of free testosterone in normal women and women with androgen deficiency: comparison of methods. J Clin Endocrinol Metab 89:525–533
33. **Rosner W** 2001 An extraordinarily inaccurate assay for free testosterone is still with us. J Clin Endocrinol Metab 86:2903
34. **Sodergard R, Backstrom T, Shanbhag V, Carstensen H** 1982 Calculation of free and bound fractions of testosterone and estradiol-17α to human plasma proteins at body temperature. J Steroid Biochem Mol Biol 16:801–810
35. **Rinaldi S, Geay A, Dechaud H, Biessy C, Zeleniuch-Jacquotte A, Akhmedkhanov A, Shore RE, Riboli E, Toniolo P, Kaaks R** 2002 Validity of free testosterone and free estradiol determinations in serum samples from postmenopausal women by theoretical calculations. Cancer Epidemiol Biomarkers Prev 11:1065–1071
36. **Ly LP, Handelsman DJ** 2005 Empirical estimation of free testosterone from testosterone and sex hormone-binding globulin immunoassays. Eur J Endocrinol 152:471–478
37. **Davis SR, Davison SL, Donath S, Bell RJ** 2005 Circulating androgen levels and self-reported sexual function in women. JAMA 294:91–96
38. **Santoro N, Torrens J, Crawford S, Allsworth JE, Finkelstein JS, Gold EB, Korenman S, Lasley WL, Luborsky JL, McConnell D, Sowers MF, Weiss G** 2005 Correlates of circulating androgens in mid-life women: the study of women's health across the nation (SWAN). J Clin Endocrinol Metab 90:4836–4845
39. **Davis SR, van der Mooren M, van Lunsen RH, Lopes P, Ribot J, Rees M, Moufarege A, Rodenberg C, Buch A, Purdie DW** 2006 The efficacy and safety of a testosterone patch for the treatment of hypoactive sexual desire disorder in surgically menopausal women: a randomized, placebo controlled-trial. Menopause 13:387–396
40. **Miller KK, Biller BM, Beauregard C, Lipman JG, Jones J, Schoenfeld D, Sherman JC, Swearingen B, Loeffler J, Klibanski A** 2006 Effects of testosterone replacement in androgen-deficient women with hypopituitarism: a randomized, double-blind, placebo-controlled study. J Clin Endocrinol Metab 91:1683–1690
41. **Vermeulen A** 2005 Hormonal cut-offs of partial androgen deficiency: a survey of androgen assays. J Endocrinol Invest 28:28–31
42. **Sam S, Legro RS, Essah PA, Apridonidze T, Dunaif A** 2006 Evidence for metabolic and reproductive phenotypes in mothers of women with polycystic ovary syndrome. Proc Natl Acad Sci USA 103:7030–7035
43. **Arlt W, Justl HG, Callies F, Reincke M, Hubler D, Oettel M, Ernst M, Schulte HM, Allolio B** 1998 Oral dehydroepiandrosterone for adrenal androgen replacement: pharmacokinetics and peripheral conversion to androgens and estrogens in young healthy females after dexamethasone suppression. J Clin Endocrinol Metab 83:1928–1934
44. **Forest MG, Sizonenko PC, Cathiard AM, Bertrand J** 1974 Hypophyso-gonadal function in humans during the first year of life. 1. Evidence for testicular activity in early infancy. J Clin Invest 53:819–828
45. **Forest MG, Cathiard AM, Bertrand JA** 1973 Total and unbound testosterone levels in the newborn and in normal and hypogonadal children: use of a sensitive radioimmunoassay for testosterone. J Clin Endocrinol Metab 36:1132–1142
46. **Stanczyk FZ** 2006 Androgen measurements: methods, interpretation and limitations. In: Azziz R, Nestler JE, Dewailly D, eds. Androgen excess disorders in women. 2nd ed. Totowa, NJ: Humana Press
47. **Morris PD, Malkin CJ, Channer KS, Jones TH** 2004 A mathematical comparison of techniques to predict biologically available testosterone in a cohort of 1072 men. Eur J Endocrinol 151:241–249

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