

A critical appraisal of accuracy and cost of laboratory methodologies for the diagnosis of hypogonadism: the role of free testosterone assays

Alvaro Morales, MD, Christine P. Collier, PhD, Albert F. Clark, PhD

Queen's University, Kingston, Ontario, Canada

MORALES A, COLLIER CP, CLARK AF. A critical appraisal of accuracy and cost of laboratory methodologies for the diagnosis of hypogonadism: the role of free testosterone assays. *The Canadian Journal of Urology*. 2012;19(3):6314-6318.

The biochemical diagnosis of male hypogonadism remains a controversial issue. The problem is compounded by the variety of laboratory assays available to measure serum testosterone (T) and the limited understanding, among clinicians, of their relative diagnostic validity. It is widely accepted that only the testosterone not bound to sex hormone-binding globulin is metabolically active. Therefore, for diagnostic purposes it is frequent practice to order the measurement of free T (FT) or bioavailable T (BAT).

Our objective is to describe the methods available for measuring FT and to review the literature to determine the relevance of ordering FT as a diagnostic laboratory tool in cases of suspected hypogonadism. We also provide our

biochemical approach in evaluating men with T deficiency. The limited information available in this regard is frequently confined to the biochemistry literature. The few reliable studies indicate that analog-based measurement of FT offers no diagnostic or financial advantage over automated assay for total T. The manuscript also describes "How we do it." For optimal diagnostic accuracy and financial responsibility, total T and calculated FT (cFT) should be the tests employed for initial and confirmatory diagnosis respectively. Measurement of bioavailable T is an alternative option but not germane to the points to which we are calling attention in this paper.

While clinicians should be discouraged from ordering FT assays, laboratories performing it should indicate what method was used and warned about possible reliability concerns. FT assays should no longer be a reimbursable test.

Key Words: testosterone, hypogonadism, diagnosis, free testosterone

Introduction

The development of hypogonadism, also known as testosterone deficiency syndrome (TDS), associated with aging in men is a highly controversial but unquestionable reality, with growing incidence mostly due to the expanding aging population.¹ Although the condition

has been traditionally considered the realm of specialists (urologists or endocrinologists), family practitioners are increasingly taking responsibility for its diagnosis and management. Confusion and controversy exists at all levels about TDS, ranging from the appropriateness of the name and approach to diagnosis, to the choices for treatment and manner of long term follow up.²

Much of the controversy centers on the biochemical diagnosis. Uncertainty originates on the variety of tests available to measure serum testosterone (T) and the discrepancies on "normal values". In addition, there are significant laboratory methodological biases which

Accepted for publication April 2012

Address correspondence to Dr. Alvaro Morales, 59 Lakeshore Blvd., Kingston, ON K7M 6R4 Canada

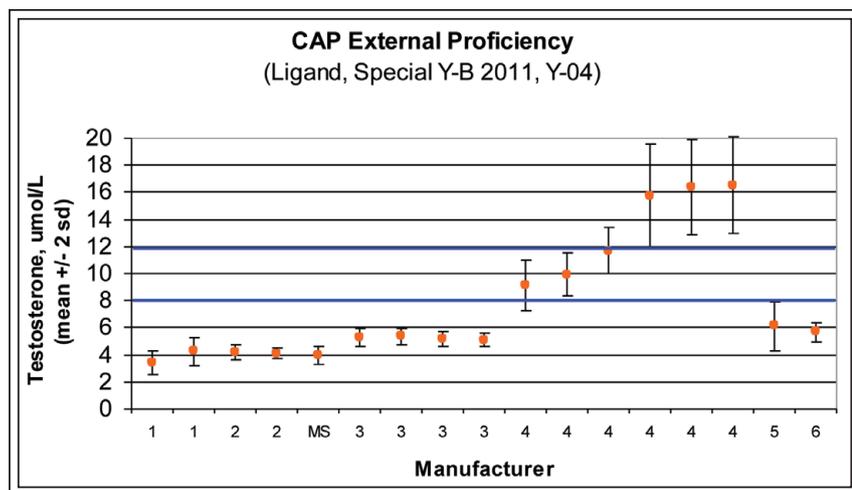


Figure 1. Inter-laboratory comparison for testosterone measurements – same sample – different instruments.

Data from the 2011 (Y-B, Y-04) College of American Pathologists' Ligand Special External Proficiency Testing Program for 17 different instruments from six manufacturers (identified by number only), as well as from labs using mass spectrometry (MS – the gold standard reference method). A total of 1500 labs participated, with 72 (1), 306 (2), 21 (MS), 403 (3), 568 (4), 114 (6) labs reporting for each manufacturer respectively. This figure demonstrates that T concentrations tend to be consistent within a manufacturer (except for #4), but not consistent across all manufacturers. This challenge sample, around the male hypogonadal decision limit, highlights the need for method (manufacturer) specific decision limits (blue lines indicate the borderline range in our laboratory, using manufacturer #2).

NOTE: *Inter-laboratory differences for patient samples should not be as pronounced as observed in this survey which uses "spiked" sera. In the CAP "PT Accuracy Survey", which uses minimally processed human sera, "80% of testosterone results are within 15% of the true value".*

are specific to the manufacturers of the test kits, Figure 1, and yet population reference ranges are often similar. To complicate matters, the literature has been at cross purposes, with expert clinical groups^{3,4} developing guidelines and recommendations based on assumptions about laboratory testing, and laboratories not providing adequate information to the clinician on the differences amongst the tests. Even publications devoted to continuing medical education frequently do not go beyond listing the tests available without providing guidance as to their diagnostic usefulness (<http://www.medscape.org/viewarticle/74920>).

Thus, it is evident that a great deal of clarification and education are badly needed for specialists and family physicians, as well as those responsible for laboratory testing. As an initial step, we would like to draw attention to the laboratory diagnosis of TDS, and specifically address the validity and cost/benefit of the "free" testosterone (FT) assay.

Methods for FT assessment

Expert professional societies recommend FT measurement for the initial diagnosis, or as confirmatory testing in equivocal cases of hypogonadism.^{3,4} FT is on Ontario's list of Ministry-approved and funded tests (MOH, L608), and is thus routinely ordered by physicians. Proficiency testing data from the College of American Pathologists (www.cap.org) indicates that FT determinations are also popular across most of North America. The challenge is that there are actually five distinct methods for determination of FT.⁵ Unfortunately the literature recommending the diagnostic utility of FT, rarely differentiates among them.

a) *Equilibrium dialysis*, is a reference method which is based on the differential passage of low molecular weight substances through a membrane with a predetermined molecular weight. Thus FT moves through the membrane from the serum sample to the dialysate over time, but T bound to protein (e.g. albumin or sex hormone binding globulin,) is retained in the serum. The FT in the dialysate is measured by a radioactive isotope detection

process. Results are affected by assay temperature, pH and sample dilution. Equilibrium dialysis is a manual time consuming and costly method, which is only done in a handful of reference or research laboratories across North America (<http://www.mayomdicallaboratories.com/test-catalog/Performance/83686>), and thus not routinely available to most clinicians.

b) *Ultrafiltration* is also considered a reference method. FT from the serum is forced by ultracentrifugation through a selective membrane into the dialysate and then measured by a radioactive isotope detection process. Although the method is faster than equilibrium dialysis and is more automated, it is still considered a manual and time consuming assay, which explains why it is also only available in a few laboratories.

c) *Analog FT*. Based on the limited availability and high cost of performing the reference methods for FT, "direct" measurement of FT by radioimmunoassay (RIA) was developed as commercially available kits for

the diagnosis of hirsutism in women. In this single-step, non-extraction method, a radio-labeled testosterone “analog” competes with FT in the serum sample for binding to an antibody that has been immobilized on the surface of the assay tube. The premise of the test is that the T analog has a low affinity for sex-hormone binding globulin (SHBG) and for albumin, and thus does not bind to them during the assay. However, values for healthy men are almost 10 fold lower compared to reference FT methods, necessitating method specific reference intervals as well as prompting serious concern about the accuracy of the method. The premise of the assay is that the T analog does not interact with other proteins in the sample, a requirement that most commercial kits fail to fulfill. There is a variety of FT immunoassays available.

Over a decade ago, Rosner⁶ was one of the first to call attention to the problems of the direct analog method for FT, concluding then that “evidence conclusively shows that the direct RIA (analog based assay] of FT is seriously inaccurate, underestimating T concentration by many-fold”. Subsequent comparative studies with equilibrium dialysis by Winters⁷ and Vermeulen⁸ confirmed these discrepancies, and in 2008, Fritz and coworkers⁹ again concluded that as the analoge-based assay “does not detect or quantify FT, it should not be used as a FT assay”. It is not entirely clear what constituent(s) are measured by the analog-based methods, but FT measured by this assay simply correlates with total T and not with FT as measured by other methods. Fritz et al⁹ speculated that the assay non-specificity may be due to protein-T complexes binding to the T antibody, while other studies suggest that the results are affected by the level of SHBG.

d) *Calculated FT*. The fourth method involves use of a simple formula (<http://www.issam.ch>) which permits the calculation of FT (cFT) from total T, SHBG, and sometimes albumin. This method is considered a good reflection of FT by equilibrium dialysis.³ As results may vary among laboratories based on their individual methodologies for total T, Figure 1, SHBG and albumin this problem may be compounded if a laboratory uses kits from different manufacturers. For this reason, it is particularly important for laboratories to validate their individual cFT reference intervals. Some laboratories don't measure albumin directly, assuming normality (e.g. 43 g/L or 4.3 g/dL) instead. Although this assumption is valid for most patients, men with severe hepatic or renal disease, may benefit from having measured albumin included in the calculation.

e) *Free androgen index (FAI)*. Included here only for the sake of completeness, FAI simply calculates the ratio of total T and SHBG. It is unreliable and should not be employed.⁹

It is worth mentioning that use of the less reliable direct analog assay for FT has finally started to decrease significantly in laboratories and in the literature. This is probably related to the fact that SHBG assays have not been routinely available until recently on many of the automated immunoassay analyzers. From 2010 to 2011, use of the FT analog method fell from 50% to 33% in the CAP external proficiency surveys.

It is anticipated that as a calculation, cFT is not a reimbursable test. Overall, the combined cost of total T, SHBG and albumin is considerably less than FT by equilibrium dialysis or ultracentrifugation (or even total T and FT by the analogue RIA method). An advantage of the cFT approach is that SHBG is also reported giving the clinician a second piece of reliable information along with the total T result.⁵

Discussion

The wide use of the FT analoge-based immunoassay by laboratories is based on the ease of its performance as an automated test, and on its significantly lower cost than the non-automatable definitive measurement of FT by equilibrium dialysis. Its' popularity is also due to lack of specific knowledge by ordering physicians regarding the differences and limitations of this methodology, and to the recommendation by professional associations that FT be measured in the initial diagnosis or in equivocal cases of hypogonadism,¹⁰ without specific clarification of the preferred method to be used. Not surprisingly, then, the practicing physician often orders the correct test but the laboratory, due to availability, performs a “suboptimal” assay, without indicating which method was used.¹¹

Clinical laboratories have a responsibility to clearly define for their clients (physicians and patients) the analytical method used, to indicate the confidence or measurement of uncertainty of the results and, if necessary include pertinent helpful interpretative comments in the reports. Laboratory personnel are aware that any analytical result is subject to error even with repeated measurements of the same material. For assays of gonadal steroids in men, the difficulties are magnified by the well-recognized concept of intra-individual biological variability.¹² By the same token, physicians involved in the management of TDS have a duty to be knowledgeable about the tests that they use in routine practice. This issue is relevant since, by law, laboratories are obligated to perform the tests ordered; but nothing prevents a clinical biochemist from contacting the ordering physician to clarify or discuss the requested test.

Four basic tenets should be kept in mind when using laboratory data in the investigation of hypogonadism: 1) reference intervals from different laboratories should not

be interchanged, 2) results have an inherent measurement uncertainty, 3) there is considerable intra-individual variation in the results and 4) the same laboratory should be used when repeated measurements are necessary for diagnosing or monitoring a patient's progress or establishing their homeostatic set point. The difference in "FT" methods highlights these essential considerations and emphasizes the importance of good communication between physicians using a laboratory and biochemists responsible for ensuring quality service and patient care.

The recent 2011 CAP "Proficiency Testing Accuracy Survey" clearly showed the persistent problem of continued variability in both total T and SHBG measurements confirming the need for "additional standardization in the case of testosterone and harmonization in the case of SHBG". This reinforces the concept that "normal ranges" (reference intervals) must be laboratory specific, and that patient monitoring must be performed consistently using the same laboratory.¹¹

The way we do it

In the laboratory

We were aware of the issue of the analogue FT assay from the start, and thus have never performed the test or approved requests for referred-out testing. We did originally refer out test requests for bioavailable testosterone (BAT), however, as soon as our instrument manufacturer introduced their test kit for SHBG, we changed to providing a calculated cBAT using the Vermeulen calculation¹⁰ which is also found on <http://www.issam.ch>. As cFT and cBAT are calculated directly from the same equation and input information, we decided to only report one of these estimates (cBAT), so that there were no inconsistent interpretations based on slight variations from their respective decision limits. For an example, see Figure 2.

Requests for total T and cBAT are accepted. Our physicians are aware that cBAT is reported as a calculation. The repeat frequency for total T is set to 7 days

(as opposed to 1 to 3 months), to accommodate the repeat testing we encourage for the diagnosis of hypogonadism. Our laboratory requisition has the option to order: 1) testosterone, or 2) "testosterone with cBAT if T abnormal". This is meant to encourage the testing of total T only on the first sample, while the second and subsequent samples can have reflexive testing if needed. Our approach has been to use a borderline range for hypogonadism to accommodate T's inherent within-individual biological variation. This allows a more scientific and realistic approach to interpretation of results than the simplistic and unreliable concept of either normal or abnormal. An accurate diagnosis may require repeated testing, a more ethical option than subjecting a man to years of T supplementation that he didn't need in the first place.

Other laboratories have struggled with similar issues as the ones discussed herein. The Beth Israel Deaconess Medical Center in Boston, Massachusetts, for instance, has presented their physicians with a comprehensive review with detailed explanations and their suggested approach (http://home.caregroup.org/departments/pathology/lab_manual/addinfo5%Free_Testosterone.pdf). Thus, laboratory concerns about the appropriateness of FT analogue assays is beginning to be recognized and addressed.

In the clinic

Men at high risk (type 2 diabetes, HIV disease, obesity, opioids abuse) as well as men with a picture compatible with TDS, submit two samples of blood in the morning at 2 week intervals for determination of total T and SHBG. If the mean of the results is within 20% we consider the biological variation (CVi) to be acceptable and they can be classified with normal or low T concentrations. If the CVi is > 20%, further sampling is needed.¹² Those who consistently repeat within the "borderline" interval (8 nmol/L-12 nmol/L or 250 ng/dL-350 ng/dL) are candidates for a 3 month diagnostic trial of T therapy once other conditions and contraindications have been ruled out.

Take the example of a 62-year-old man with a body mass index of 34 and well controlled diabetes, decreased sexual desire, tiredness and fatigue. Two separate morning serum total T determinations, 2 weeks apart are reported as 10.5 nmol/L and 11.2 nmol/L. The SHBG is 80 nmol/L. Using the formula available free at <http://www.issam.ch>, the physician only needs to enter the values for total T (average in this case 11 nmol/L) and the value for SHBG. Albumin is considered a stable figure and does not require actual measurement (except in those men with severe liver or renal disease). Entering the numbers in the blank spaces will provide the clinician with results for both calculated FT (cFT), in this case 0.10 ng/dL and calculated BT (cBT) in this particular man 2.5 ng/dL. The site also gives the interested practitioner examples and a complete description of the formula.

Figure 2. As an illustrative case.

In summary, as direct analog-based FT assays do not add value to the investigation of men suspected of TDS beyond the simpler and cheaper total T determination, we recommend that:

- physicians be strongly discouraged from ordering FT, unless they can justify and have access to a reference FT method,
- laboratories indicate the FT method used, and
- neither FT nor cFT should be a reimbursable test. □

References

1. Morales A, Heaton JPW, Carson C. Andropause: A misnomer for a true clinical entity. *J Urol* 2000;163(3):705-712.
2. Eggertson L. Brouhaha erupts over testosterone-testing advertising campaign. *CMAJ* 2011;183(16):1161.
3. Bhasin S, Cunningham GR, Hayes FJ et al. Testosterone therapy in men with androgen deficiency syndrome: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2010; 95(6):2536-2559.
4. Wang C, Nieschlag E, Swerdloff R et al. Investigation, treatment, and monitoring of late-onset hypogonadism in males: ISA, ISSAM, EAU, EAA, and ASA Recommendations. *J Androl* 2009;30(8):1-9.
5. Wheeler MJ, Barnes SC. Measurement of testosterone in the diagnosis of hypogonadism in the ageing male. *Clin Endocrinol* 2008;69(9):515-525.
6. Rosner W. Errors in the measurement of plasma free testosterone. *J Clin Endocrinol Metab* 1997;82(6):2014-2015.
7. Winters SJ, Kelley DE, Goodpaster B. The "analog" free testosterone assay: are the results in men clinically useful? *Clin Chem* 1998;44(10):2178-2182.
8. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for evaluation of free testosterone in serum. *J Clin Endocrinol Metab* 1999;84(10):3666-3672.
9. Fritz KS, McKean AJS, Nelson JC, et al. Analog-based free testosterone test results linked to total testosterone concentrations, not free testosterone concentrations. *Clin Chem* 2008;54(3): 512-516.
10. Collier CP, Clark AF, Bain J et al. Functional testosterone: biochemical assessment of hypogonadism in men – Report from a multidisciplinary workshop hosted by the Ontario Society of Clinical Chemists. *Aging Male* 2007;10(4):211-216.
11. Collier CP, Morales A, Clark AF et al. The significance of biological variation in the diagnosis of testosterone deficiency and considerations of the relevance of total, free and bioavailable testosterone determinations. *J Urol* 2010;183(6):2294-2299.
12. Black A, Day A, Morales A. The reliability of the clinical and biochemical assessment in symptomatic late onset hypogonadism. Can a case be made for a 3-month therapeutic trial? *Br J Urol* 2004; 94(5):1066-1070.