HHS Public Access

Author manuscript

J Sex Med. Author manuscript; available in PMC 2017 July 19.

Published in final edited form as:

J Sex Med. 2016 July ; 13(7): 1029–1046. doi:10.1016/j.jsxm.2016.04.068.

Challenges in Testosterone Measurement, Data Interpretation, and Methodological Appraisal of Interventional Trials

Landon W. Trost, MD[a,d] and **John P. Mulhall, MD**[d]

[a]Department of Urology, Mayo Clinic, Rochester, MN

[d]Department of Urology, Memorial Sloan-Kettering Cancer Center, New York, NY

Abstract

Introduction—Male hypogonadism is a common condition with an increasing body of literature on diagnosis, implications, and management. Given the significant variability in T from a physiologic and assay perspective, a thorough understanding of factors impacting T values and study methodology are essential to appropriately interpret reported study outcomes. However, despite the large number of T publications, there are no reference materials, which consolidate all relevant and potentially confounding factors necessary to appropriately interpret T studies.

Aims—To create a resource document that reviews sources of T variability, free versus total T, assay techniques and questionnaires, and study methodology relevant to interpreting outcomes.

Methods—A PubMed search was performed of all T literature published on T variability, assay techniques, and T-specific questionnaires. Results were summarized in the context of their impact on interpreting T literature outcomes and methodology.

Main Outcome Measures—The effect of various factors on T variability and their relevance to study methodology and outcomes.

Results—Several factors impact measured T levels including aging, circadian rhythms, geography, genetics, lifestyle choices, comorbid conditions, and intra-individual daily variability. The utility of free T (fT) over total T is debatable and must be compared using appropriate threshold levels. Among various assay techniques, mass spectrometry and equilibrium dialysis are gold standards. Calculated empirical estimates of fT are also commonly utilized and accepted. Hypogonadism-specific questionnaires have limited utility in screening for hypogonadism, and their role as objective end-points for quantifying symptoms remains unclear. Numerous aspects of study methodology may directly or indirectly impact reported outcomes including design (randomized, prospective, retrospective), duration, populations studied (age, comorbid conditions), low T threshold, therapeutic agent utilized, objective measures/end-points selected, and statistical interpretation.

Corresponding Author: John P. Mulhall, MD, Professor of Urology, Memorial Sloan-Kettering Cancer Center, Department of Urology, 353 E 68th St, Rm 531, New York, NY 10065, mulhalj1@mskcc.org.

Conflicts of Interest: None

Conclusions—Critical appraisal of T literature requires an understanding of numerous factors resulting in T variability, study design and methodology, and limitations of assay techniques and objective measurement scales.

Introduction

Male hypogonadism represents a clinical condition characterized by one or more hypogonadal symptoms in the setting of low serum testosterone (T). It is distinguished from "classical hypogonadism," which is defined as insufficient T production due to disruption of the hypothalamic-pituitary-gonadal (HPG) axis.[1] The prevalence of symptomatic hypogonadism remains poorly defined, with estimates ranging from 2.1–17% in population-based surveys of men aged 30–87 years depending upon criterion utilized.[2–4] Likely due to a variety of factors, including increased recognition and availability of therapies, the number of patients diagnosed and treated for hypogonadism over the past decade has increased significantly, with an exponential rise in prescribed T in many developed countries.[5–7]

Given the prevalence and increasing recognition of the condition, it is prudent for specialists managing hypogonadism to have a thorough understanding of the methodology and limitations associated with currently available diagnostic tools and contemporary research findings. Several challenges exist in the diagnosis and treatment of hypogonadism including variability in T assays, lack of consensus on normal T levels, and poor objective measures for symptom assessment and therapeutic benefit. Similarly, although abundant research is available on hypogonadism and associated comorbidities, relatively limited data are available regarding the effect of T supplementation on these conditions. Variations in study methodology, inclusion/exclusion criteria, populations evaluated, absence of validated objective measures, and small cohorts have all hindered the quality of data obtained and restricted generalizability of findings.

To address the many challenges and limitations with hypogonadism management, the current review is outlined to provide practitioners with a concise overview of T assessment in general, characteristics and limitations of T assays and relevant questionnaires, and the impact of T physiology on laboratory assessment. Various factors of clinical research methodology will be reviewed including design, population, study agent, selected end-points, and adverse event (AE) reporting, with emphasis on their impact on interpretation of study findings. Further mention will also be made of methodology for data quality assessment and statistical interpretation. The objective of the manuscript is therefore to equip the practitioner with a practical and readily accessible resource to enhance understanding of hypogonadism assessment and aid in the comprehension and critical appraisal of contemporary literature. The physiologic effects of T deficiency and beneficial/adverse impacts with T replacement are beyond the scope of the current discussion and will not be reviewed.

Testosterone Physiology

T is the predominant androgen in males and is involved in multiple physiologic processes throughout the body. T production is regulated through the HPG axis, with both T and

estradiol (E2) providing feedback regulation at the level of the hypothalamus and pituitary.[8, 9] Circulating T is predominantly bound to sex hormone binding globulin (SHBG) and albumin, with small percentages freely circulating or associated with corticosteroid binding globulin.[10] The role and importance of free T (fT), bioavailable T, and SHBG is discussed in greater detail later in this manuscript.

Circulating T may exert a direct effect on tissues/cells, or undergo conversion to downstream hormones estradiol or dihydrotestosterone (DHT) via 5-alpha reductase (5AR). Estradiol, DHT, and the DHT product, 3-alpha androstanediol have established roles in reproductive, bone, hepatic, renal, cardiovascular, dermatologic, prostatic, penile, and central nervous system functions, among others.[11–23] T and DHT also likely have differential roles and expression in various tissues, including in the prostate, skin, penis, and testicles, among others.[15, 24–30]

Both T and DHT exert their physiologic effects via androgen receptors (AR). Following binding of androgens with the AR, the complex is translocated to the nucleus, where it interacts with androgen response elements on the DNA. This then serves to express or repress various androgen-dependent genes, which subsequently act on target tissues. Variability in AR size is inversely associated with function, and is determined, in part, by the number of CAG trinucleotide repeats.

Based on these findings, several studies have evaluated associations between CAG repeats and hypogonadal symptoms. Results demonstrated findings similar to those observed in men with low T, including depressed mood, anxiety, impaired spermatogenesis, and loss of bone mineral density (BMD), among others.[31–33] Men with low numbers of CAG repeats (improved AR function) may also experience improved response rates to T supplementation.[34]

Although a complete review of hormonal physiology is beyond the scope of the current text, knowledge of the role and impact of factors including AR, DHT, and 5ARI function is relevant to understanding variable responses to T supplementation. While men with normal T levels and dysfunctional ARs may experience hypogonadal symptoms which are unresponsive to T supplementation, men with low T and increased AR activity may exhibit no findings of T deficiency. Similarly, men with 5ARI abnormalities may fail to achieve benefits with T supplementation on erectile function, despite improvements in other T-related aspects. These issues must be taken into account in interpreting hypogonadism literature and provides some context to interpreting variability in symptoms in men with similar T levels.

Variability in Testosterone Levels (Table 1)

Aging

Numerous studies have consistently identified age-dependent decreases in T.[35–40] Data from a longitudinal cohort of men aged 40–70 years reported an estimated TT decrease of 0.8%/year, with fT decreasing by 2%/year (due to concomitant increases in SHBG).[37] Two additional longitudinal studies reported a 3.6–3.8 ng/dl/year (0.12–0.13 nmol/L/year)

decrease in TT, with progressive increases in the number of men characterized as biochemically hypogonadal (defined as TT<325 ng/dl [11.3 nmol/l] or fT<2.5th percentile): 20% (>60 years), 30% (>70 years), and 50% (>80 years).[39, 40]

Despite the observed decrease in TT and fT levels in aging males, the effect of age as an independent risk factor for decreasing T remains controversial. One study comparing 325 men >age 40 years with self-reported “very good” or “excellent” health demonstrated no significant differences in T levels based on age.[41] A larger study evaluating 1588 men >age 35 years over a 5-year period found that age was not an independent predictor of decreasing T on multivariate analysis, while obesity, smoking status, chronic medical conditions, lifestyle factors, depression, and marital status all likely contributed to the observed decline.[40] These findings are in contrast to a larger analysis (n=3690) of community-dwelling, elderly males (mean age 77 years) which demonstrated similar mean T levels (406 vs 378 ng/dl [14.1 vs 13.1 nmol/L]) among those describing excellent or very good health compared to the overall cohort, suggesting that subjective assessment of overall health may not represent a significant factor for T levels.[35]

Data from the Massachusetts Male Aging Study similarly demonstrated a persistent age-related decrease in T levels even after controlling for obesity, chronic illness, medication use, and excessive alcohol consumption, although an attenuation of rate of decline and 10–15% overall higher T levels were observed among healthier males.[37] The European Male Aging Study (EMAS) also identified age-related increases in symptomatic hypogonadism, with rates of 0.1% among 40–49 year old men rising to 0.6% (50–59), 3.2% (60–69), and 5.1% (70–79).[4] It is noteworthy that the definition utilized for symptomatic hypogonadism in the EMAS was arguably the most strict and required the presence of three sexual symptoms, a TT <317 ng/dl (11 nmol/L) and fT <60 pg/ml (208 pmol/L).

Diurnal Variation

T levels exhibit time, sleep, and age-dependent circadian variations. Early studies comparing young (mean age 25–27 years) versus elderly males (mean age 71 years) demonstrated peak T concentrations in the 3–8 AM time period (extended to 2 PM in elderly males), with trough levels noted in the early afternoon to late evening.[42–45] Despite similar patterns of variation throughout the day, elderly males experienced significant blunting of peak versus trough levels, with differences in high and low values of 61–205 ng/dl (2.1–7.1 nmol/L) compared to 141–354 ng/dl (4.9–12.3 nmol/L) in younger men. In relative terms, T levels obtained at 4 PM are 20–25% lower in men aged 30–40 and 10% lower in 70 year old men compared to those received at 8 AM.[46] SHBG, fT, and bioavailable T also exhibit similar circadian variations, with SHBG peaking in the afternoon hours.[42, 45]

Variations in AM T levels may also reflect abrupt changes occurring following waking from sleep. One study, which obtained saliva-based fT assessments in 783 male twins, demonstrated significant decreases in fT following waking from sleep, with 32–39% of the total decline occurring within 30 minutes of waking.[47] Other studies have confirmed increasing levels of T during sleep, with progressive reductions in T during waking hours.[48] These findings may account, in part, for observed elevations in afternoon T levels (associated with afternoon naps) seen in earlier studies.[42, 43]

Ethnic, Geographic, Genetic, and Seasonal Influences

Several studies have evaluated the impact of genetic and environmental factors on variations in T levels. In comparing T levels among Asian, black, Hispanic, and white men, slight variations in outcomes have been reported. Litman and colleagues demonstrated no significant differences in T, DHT, or SHBG among 1899 black, Hispanic, or white men sampled from a localized community, while Rohrmann and colleagues evaluated men presenting for a national exam and noted mild significant increases in T (mean 548 ng/dl [19 nmol/L], $p<0.05$) among Hispanics compared to blacks (mean 529 ng/dl [18.4 nmol/L]) or whites (mean 511 ng/dl [17.7 nmol/L]).[49, 50] A broader sampling of 1127 African American, Chinese American, Japanese American, or white men from Hawaii, California, and Canada, demonstrated the highest T levels in Asian Americans (mean 512–521 ng/dl [17.8–18.1 nmol/L], $p<0.05$ compared to white males), intermediate in African Americans (495 ng/dl [17.2 nmol/L]), and lowest in whites (471 ng/dl [16.3 nmol/L]).[51]

In contrast to the minimal variations noted among different ethnicities, geographical location appears to have a significant impact on T levels. Orwoll and colleagues reported on 5003, community-dwelling men from Japan, Hong Kong, Sweden, Tobago, and the United States (US).[52] With adjustment for age and body-mass index, T levels were 16% higher in men living in Hong Kong and Japan, while Asian men living in the US were found to have similar levels to other US residents. No differences were noted in T levels based on ethnicity alone. These combined results suggest that ethnicity alone is not likely a significant factor resulting in T variability.

Genetic heritability likely accounts for a significant portion of mean T levels in a select population as well as observed diurnal variations. Studies of male twins have attempted to identify the extent of interindividual T variability attributable to genetic factors. Pre-adolescent and adolescent males demonstrated genetic contribution rates of 52–66%, with the remaining percentage due to various environmental factors.[53, 54] Results of adult males have confirmed similar findings, with genetic contribution towards T variability of 42–65%.[47, 55–57] These results confirm prior genetic studies noting an association of T values within brothers in the same family.[58]

The effect of seasonal variation on T levels remains controversial, with no definitive evidence available at the present time. Several studies using varied study designs, populations, and analytical methods have identified seasonal variations, with some suggesting higher T levels in winter months.[59–61] In contrast, other authors have reported no significant findings, with one well-designed study of 120 men aged 30–79 demonstrating larger variations in intra-individual measurements than seasonal variability.[62, 63] Given the discrepant and inconsistent findings, there are insufficient data to support or refute seasonal variability in T levels.

Intra-individual Variation

Intra-individual variability is another significant factor, which must be taken into account when interpreting T values. Several studies have noted fluctuations in T values obtained from the same subject at similar time points.[64–67] One study of eight healthy males

undergoing T assessments on two separate days noted a 32% absolute variation in reported results, while a second study assessing T levels over a 12-month period reported a 10.9% probability of distribution.[65, 66] A similar study of 16 men found that 50% of men who were identified as hypogonadal (defined as <300 ng/dl [10.4 nmol/L]) on one given measurement were noted to be eugonadal on repeat testing.[64] The largest study specifically evaluating intra-individual variability reported on 121 men, aged 30–79 randomly selected from the Boston Area Community Health Survey.[67] Participants had six samples obtained on separate visits, within four hours of waking. Results demonstrated greater variation between serial intra-individual measurements than from differences in assays themselves. Based on one sample obtained, 95% confidence limits were calculated at 65–153% of the value obtained, while the average of two and three measurements reduced the limit by 30% and 43%, respectively. Of interest, among men found to have T < 250 ng/dl (8.7 nmol/L) on the first assay, only 40% were confirmed to have a mean T < 250 ng/dl (8.7 nmol/L) over the six visits, with 20% averaging > 300 ng/dl (10.4 nmol/L).

Lifestyle Factors and Disease States

Lifestyle factors are independently associated with variations in T levels and account for a percentage of the observed age-associated decline in T.[68] Obesity and T are inversely correlated, with increasing obesity resulting in progressive impairments in gonadotropins and T.[69, 70] A 4–5 point increase in BMI is roughly equivalent to a 10-year decline in T, while weight loss directly correlates with increasing T.[68, 71] A meta-analysis of studies reviewing the effect of weight loss on T demonstrated mean T increases of 83 ng/dl (diet alone [2.9 nmol/L]) and 252 ng/dl (bariatric surgery [8.7 nmol/L]), with greater improvements in bariatric surgery attributed to the more extensive weight loss achieved.[71]

In addition to weight loss, exercise is independently associated with increasing T levels. The degree of T increase is related to several factors including exercise duration, extent of resistance provided, and participant age. Men performing 236 minutes of moderate-intensity exercise experienced greater increases in T (59 ng/dl versus 23 ng/dl [2 vs 0.8 nmol/L]) over those performing 105 minutes.[72] Compared to aerobic activity, resistance exercise likely results in greater improvements in T, with untrained men experiencing larger increases compared to those routinely performing resistance exercises.[73–75] Findings suggest that regular training may result in physiologic adaptation and necessitate greater stimuli to achieve similar hormonal elevations with subsequent exercises.

Contradictory data are available on the effect of smoking on T levels. Some studies have associated smoking with increasing T, while others report no or inhibiting effects.[68, 76–80] The duration of smoking may account for some of the variability, as one recent study demonstrated decreased T levels among those with >20 pack year histories.[79] In contrast to smoking, moderate alcohol use has not been associated with altered T levels.[76–78]

Both acute and chronic disease states are associated with decreasing T levels. Muehlenbein reported on 25 young men with upper respiratory infections who underwent serial T assessments.[81] Results demonstrated a transient mean 10% decline in T during the acute phase of the illness, with reductions of up to 30% in a select cohort of patients. Similarly,

development of additional chronic illnesses or increasing medication use results in a more rapid age-associated decline in T.[68]

Testosterone Assays

Assay Techniques

The accurate and precise measurement of T has remained a challenge since its initial discovery. Contemporary assay techniques to assess TT include immunoassays and mass spectrometry (MS). See Table 2 for a summary of advantages and disadvantages of various assay techniques. In immunoassays, as a category, tracer-linked T competes with T present in the sample for binding to T antibody. The tracer may be a radioisotope (radioimmunoassay, RIA), enzyme (enzyme immunoassay, EIA), or a fluorescent (fluoroimmunoassay, FIA) or chemiluminescent compound. In contrast to immunoassays, MS ionizes serum compounds and measures their subsequent mass to charge ratios. To enhance assay sensitivity and specificity, samples to be tested may undergo pre-analysis extraction or chromatography (gas or liquid) to separate proteins and hormones, which might otherwise impair the accurate measurement of T.

Given their simplicity, ease of use, and high-throughput, immunoassays (IA) were widely adopted in clinical practice, with the majority of TT reference ranges established using these techniques.[82, 83] However, over the past 10 years, MS, and in particular liquid chromatography tandem mass spectrometry (LC-MS/MS), has become increasingly adopted due to its high throughput, limited requirement for sample preparation, and high sensitivity/specificity at low and high T concentrations.[82, 84, 85]

Assay Variability

One of the challenges with T interpretation is the significant variability existing among laboratories and various assay techniques. Several studies have examined the extent of variation using standardized reference samples. In comparing IA to the gold-standard MS, reported variability ranges from −14.1% to +19.2% in samples overall and −40% to +40% among samples <100 ng/dl (3.5 nmol/L).[86, 87] At lower concentrations, IA demonstrate particularly significant variations among techniques, with equivalent samples resulting in 2.7–14.3 fold variations in reported results.[88] And despite the recognition of MS as a reference technique, the reliability of results depends upon regular calibration maintenance, which is labor intensive and limits the ability to achieve consistently high throughputs without deterioration.

Additional factors, which may account for variability of results include specimen handling and preparation, calibration methods utilized, specimen commutability, and interference from the matrix material used to store and transport sample T preparations.[83, 89, 90] Few studies discussing T supplementation report details on specifics of the testing modality itself. Ideally, information on lower limits of T detection, quantitation, and the method of extraction should be included.

To aid in the standardization of reported TT levels, the Centers for Disease Control (CDC) has initiated a hormone standardization program, with minimum guidelines established to

receive certification.[91] Currently, the CDC requires that a certifying facility report TT values within ±6.4% of samples tested ranging from 2.50–1000 ng/dl (0.9–35 nmol/L). For the benefit of practitioners, laboratories meeting these criteria are available on the CDC website, with dates of most recent certification reported.[91]

Reference Values

Currently, there is no consensus as to the accepted lower T limits or established reference intervals.[92, 93] A guideline statement endorsed by the International Society of Andrology (ISA), International Society for the Study of Aging Male (ISSAM), European Association of Urology (EAU), European Academy of Andrology (EAA), and the American Society of Andrology (ASA) establishes TT values of <230 ng/dl (8 nmol/L) in young men as benefiting from treatment, while >350 ng/dl (12.1 nmol/L) does not require therapy.[92] Similarly, guidelines by the Endocrine Society have agreed upon a lower TT threshold for therapeutic consideration at 280–300 ng/dl (9.7–10.4 nmol/L).[1]

Given the increasingly wide-spread adoption of MS for measuring T levels, reference ranges using LC-MS/MS have recently been described.[35, 36] Bhasin and colleagues reported on a cohort of 456 men, aged 19–40 years from the Framingham Heart Study Generation 3.[36] All patients were healthy with no obesity, cardiovascular disease, DM, hypertension, dyslipidemia, or tobacco users. Mean TT was 724 ng/dl (25.1 nmol/L), with upper (97.5%) and lower (2.5%) intervals noted to be 1197 and 348 ng/dl (41.5 and 12.1 nmol/L), respectively. A second study evaluating 3690 elderly (mean age 77), community-dwelling men identified a mean TT of 378 ng/dl (13.1 nmol/L), with upper (97.5%) and lower (2.5%) reference ranges of 693 and 145 ng/dl (24 and 5 nmol/L), respectively.[35] Of interest, a subset of patients describing themselves as in excellent or very good health had similar mean, 97.5%, and 2.5% TT levels compared to the entire cohort (mean-406, 97.5%–739, 2.5%–184 ng/dl [14.1, 25.6, and 6.4 nmol/L]). These combined findings highlight the difficulty in establishing standardized reference ranges, given the declining T associated with aging and the lack of defined, age-specific cut-points for various symptomatology.

Free Testosterone

Physiology

Testosterone circulates in the plasma as either a free molecule (fT - 2%) or complexed with varying affinities to proteins including albumin (loosely bound - 50%), SHBG (tightly bound - 44%), or corticosteroid-binding globulin (loosely bound - 4%).[94] In addition to fT, biologically-active T is commonly reported, which represents TT minus the percentage of T bound to SHBG. As with TT, fT is influenced by diurnal variations and is inversely associated with age and BMI.[45, 68]

Thresholds for Low Free Testosterone

Two notable studies have performed population-based assessments to identify the 2.5th percentile for low T. In a community-based sample of 3,690 elderly men (mean age 77), calculated fT levels were 24.2 pg/ml (83.9 pmol/L), with a separate cohort identifying themselves as being in excellent or very good health reported at 29.9 pg/ml (103.7 pmol/

L).[35] A second study of non-obese healthy men aged 19–40 years from the Framingham Heart Study reported calculated fT levels of 70 pg/ml (242.7 pmol/L).[36]

Current Endocrine Society guidelines recommend measuring fT in men with low-normal TT levels in whom SHBG alterations are suspected with low values established within each laboratory.[1] The combined guidelines from the ASA, EAA, EAU, ISA, and ISSAM similarly recommend obtaining fT levels in men with TT 231–346ng/dl (8–12 nmol/L), with a suggested threshold of 65 pg/ml (225.4 pmol/L) established as the lower range of normal.

Clinical Relevance of Free Versus Total Testosterone

The clinical relevance of TT compared to fT is controversial, with many suggesting that fT is a more appropriate measure of hypogonadism due to SHBG variations occurring with aging, acute/chronic diseases, hormonal alterations, obesity, and normal diurnal rhythms.[95]

Limited data comparing the association of TT or fT to clinical variables in elderly males have demonstrated a stronger relationship between calculated bioavailable T and muscle strength, bone mineral density, and fat mass compared to TT.[96] Other studies have also demonstrated increased associations between fT and other comorbid conditions, including depression and hypogonadal symptoms.[2, 97] Both fT and TT are equally associated with BMI, ED, and decreased libido.[70, 76]

However, many of these studies are hindered by heterogeneity in the definition for low fT (50–65 pg/ml [173.4–225.4 pmol/L]), with increases in the threshold for low fT resulting in higher sensitivity and reduced specificity when compared to TT.[2] This alone may account for the stronger associations reported with fT over TT. Free T is also better indicator in young males as compared to elderly men, suggesting that the age-associated increase in SHBG may further reduce the specificity of fT.[2, 98]

Similar to findings with the AR gene, SHBG polymorphisms also directly impact the expression and variability of SHBG itself as well as systemic fT and E2.[99, 100] Despite an increasing body of literature on the impact of SHBG polymorphisms on hypogonadal-related conditions, limited data are currently available on their impact in hypogonadal men undergoing T supplementation.

The relevance of fT versus TT is also debatable, as recent data have identified cellular uptake of SHBG-bound T with resultant physiologic hormonal activity.[101] Given this observation, the optimal method for diagnosing hypogonadism is unclear with the purported advantages of fT over TT currently in question. Similarly, it is important to recognize that findings of studies which utilize TT to diagnose hypogonadism should not necessarily be extrapolated to suggest that similar results can be expected using specific fT thresholds, and vice-versa.

Measurement of Free Testosterone

Free T is measured either through direct assays or indirectly via several different published calculations. See Table 2 for a summary of advantages and disadvantages of various measurement techniques. Currently, the gold-standard method to measure fT is equilibrium

dialysis, which achieves differential passage of fT via a low molecular weight semi-permeable membrane. The percentage of fT is then calculated via displacement of tracer-labeled T and multiplied by TT. Limitations with this technique include increased expense, dependence on TT accuracy, and variations created by radiotracer impurities, temperature control, and sample dilution, among others.[83]

Numerous estimating equations have been described in the literature including bioavailable T (non-SHBG bound T), free androgen index (FAI; 100T/SHBG), and free T index (Vermeulen method), among others, with results suggesting high correlations to direct assays.[102–105] Several studies have evaluated the predictive accuracy of calculated fT, with one study of 1072 men comparing estimates to assay-determined bioavailable T.[106] Results demonstrated high predictability using TT (r^2=0.68), with TT being better than other tested modalities for determining biochemical hypogonadism (area under receiver operative curve; TT=0.93, FAI=0.72, Nanjee and Wheeler calculation=0.91, Vermeulen calculation=0.88). At lower TT levels (216–346ng/dl [7.5–12 nmol/L]), fT was found to be superior to TT alone. These findings suggest that TT may be a better indicator of hypogonadism, except in cases of borderline-low TT, where calculated fT or bioavailable T may help to confirm a diagnosis. Current guidelines from the ASA, EAA, EAU, ISA, and ISSAM have adopted this strategy of using fT as a confirmatory marker in cases of borderline low TT.[1]

In the largest study evaluating predictive accuracy of fT calculations, Sartorius and colleagues concluded that empirical methods (Ly, Sartorius) of calculated fT are most concordant, with other estimations (Vermeulen, Nanjee and Wheeler, Sodergard) resulting in overestimation of the true value.[102–105, 107] Of interest, overall variability was only minimally influenced by the calculated fT algorithm utilized (14% of variability), while the different TT and SHBG assays accounted for 82% and 4% of observed variance, respectively.[107] A more recent, multi-step model of estimating fT using a dichotomized analysis of SHBG reported estimates which were not statistically different from those obtained via equilibrium dialysis.[108] To date, no study has compared this new technique with more established empirical methods (Ly, Sartorius).

Objective Assessments

Several questionnaires have been developed to aid providers in screening men for hypogonadism or to follow symptomatic improvements with T supplementation. Currently published questionnaires specific for hypogonadism include the Aging Male's Symptoms Scale (AMS – 1999), Androgen Deficiency in Aging Males (ADAM – 2000), Massachusetts Male Aging Study (MMAS – 2000; also referred to as Smith's screener), Age-related Hormone Deficiency-dependent Quality of Life Questionnaire (A-RHDQoL - 2003), and Hypogonadism-related Symptom Scale (HRS – 2009).[109–113] The ANDROTEST (2009) has also been published as a structured interview to provide scoring relevant to identifying hypogonadism-related signs and symptoms.[114]

The sensitivity and specificity of the various scales have been reported, with overall findings demonstrating the highest sensitivities with the AMS and ADAM questionnaires (81–97%) with concomitant poor specificity (19–39%).[115–117] The MMAS and ANDROTEST

questionnaires exhibit improved specificity (53–65%), with loss of overall sensitivity (60–71%).[111, 114, 115] Relatively limited data are currently available on the A-RHDQoL and HRS questionnaires.

Available questionnaires demonstrate correlations with aging and DM with variable associations noted with TT and fT levels.[115, 118, 119] In comparing the AMS, ADAM, and MMAS questionnaires, Heinemann and colleagues noted significant similarities and compatibility among instruments, with no one scale demonstrating superiority.[120] In comparing the AMS, ADAM, and MMAS scales on ability to identify low bioavailable T levels, one study noted inferiority of the MMAS scale as a screening modality due to a relatively lower sensitivity (60%).[115]

The role for questionnaires in identifying improvements with T supplentation is unclear. Although select RCTs have demonstrated modest improvements in scales following T administration, others have failed to identify consistent associations.[119, 121]

Given the variable sensitivities, specificities, inconsistent correlations with biochemical hormonal parameters, and inability to reliably track outcomes with T supplementation, currently available questionnaires are not suitable as screening modalities or as surrogates for hormonal testing.[1, 92] Due to these limitations, studies reporting beneficial outcomes using questionnaires alone should be interpreted with caution.

Critical Evaluation of Testosterone Literature

With the preceding information as a foundation, the following sections are outlined to review and critically appraise T literature, with particular emphasis on common errors in data interpretation. See Table 3 for a sample checklist to critically evaluate T literature.

Study Design

Recognizing the strengths and limitations of study designs is critical to accurately assessing T literature. Although several guideline bodies have established criteria for determining literary quality, one commonly utilized method is assigning levels of evidence based on the Oxford Criteria.[122]

Among other factors, Oxford criteria include study design, population homogeneity, and data confidence intervals to better estimate the overall reliability of reported results. Using these criteria, greater weight should be given to studies of higher levels of evidence when compared to lesser quality studies. As such, systematic reviews and meta-analyses of RCTs (level Ia) are of significantly greater value and reliability than retrospective series (level 2b).

The majority of studies evaluating comorbid conditions and hypogonadism are retrospective in nature. As such, they are only able to identify correlations between low T and comorbid conditions. In these settings, it is important to recognize that associated conditions do not necessarily represent a causative link. Only RCTs are able to isolate causative factors.

One example highlighting the difference between associations and causation as well as overall quality of evidence is with T supplementation and mortality. A commonly cited observational series involved 1031 male veterans with hypogonadism receiving T supplementation compared against those not receiving treatment.[123] Results demonstrated a significant association between low T and overall mortality, while those who received T supplementation were shown to have improved survival. In contrast, a systematic review and meta-analysis of T trials, including RCTs demonstrated no significant effect of T supplementation on overall survival.[124] These contrasting studies highlight the need for caution in interpreting results from retrospective series reporting on observed associations rather than relying on RCT data.

Study Duration

Study duration is particularly relevant in interventional trials of T supplementation. Potential beneficial and harmful effects of T supplementation are time-dependent and may not be observed in trials of insufficient duration. Two RCTs demonstrating the importance of study duration evaluated the effect of the aromatase inhibitor anastrazole on bone health. Leder and colleagues evaluated 68 men treated with anastrozole 1mg/day over a period of 3 months, while Burnett-Bowie's group performed a similar trial for 12 months.[125, 126] Results from the Leder trial reported no change in markers of bone turnover, despite greater mean reductions in estradiol, while Burnett-Bowie's group identified significant reductions in bone mineral density with therapy. Inversely, a 3-year prospective, controlled trial of T supplementation in hypogonadal men demonstrated a 5% per year improvement in BMD.[127] In addition to changes in bone density, studies of extended duration are likely required to assess other factors including changes in cognition, muscle mass, adiposity, bone fracture risks, cardiovascular risk, and comorbid condition sequelae. Duration of therapy is also significant in regards to trials evaluating T supplementation and prostate cancer, particularly given the extended natural history of prostate cancer development, progression, and recurrence.

Population

The population studied has a significant impact on outcomes, including the study setting, participant age and comorbidities, biochemical definition of hypogonadism, and the use of specific populations.

Community-based, military conscript, or similar trials will likely yield significant differences in sampling when compared to populations common to urologic practices, including those with sexual dysfunction, infertility, or other urologic conditions. Geographical differences are also associated with differential baseline T levels.

Similarly, the date of record sampling is relevant, particularly given that the majority of laboratories assessing T prior to 2005 would have likely used EIA or RIA rather than MS. This may result in reduced accuracy, particularly at lower T levels.

Age and Comorbid Conditions

Participant age is a common confounder in the hypogonadism literature and must be controlled in observational or retrospective studies of hypogonadism-associated conditions. Although controversial, the preponderance of data have demonstrated an association between aging and low T. In addition to the potentially confounding effect of age on associated conditions, T supplementation in elderly males may exhibit a physiologically blunted effect. Similarly, the use of alternative T therapies, such as selective estrogen receptor modulators (SERMs) may result in reduced efficacy in elderly compared to younger males. Tenover and colleagues prospectively treated young and elderly (mean age 29 years vs. 73 years) men with clomiphene citrate 50mg/day for two months.[128] Results demonstrated an approximately 870ng/dl (30.2 nmol/L) increase in the youthful group compared to 489ng/dl (17 nmol/L) in elderly males. Findings suggested reduced responsiveness of the hypothalamic-pituitary-gonadal axis in elderly males, an effect, which must be accounted for in comparing studies of men with differing ages. Similarly, although data are lacking, potential benefits of T supplementation may be reduced in elderly males due to age-associated impairments in T responsiveness.

In addition to age, comorbid conditions must be accounted for in interpreting hypogonadism data. Obesity, DM, MetS, depression and other comorbid conditions have marked effects on T levels and impact of supplementation.[40] As such, comparisons between hypogonadism studies evaluating populations of differential comorbid status is of questionable reliability and benefit. This is particularly relevant in performing meta-analyses and systematic reviews of heterogenous populations. Additionally, comparisons between treatment and control groups must be appropriately matched to reduce errors in data interpretation.

One recent example of poorly matched groupings was the Vigen et al's retrospective analysis of a VA population, which compared hypogonadal men receiving or not receiving T supplementation.[129] Baseline comorbid states were significantly different between groupings, including obesity, obstructed coronary artery status, hypertension, hyperlipidemia, myocardial infarction, and cerebrovascular disease, among others. Additionally, groupings were not matched for age or T levels. Although the authors identified a higher rate of AEs in the T supplementation group (after statistical manipulation of the raw data), given the wide discrepancy in groupings, it is unclear if this group represented men with a higher baseline risk for subsequent events and that a lower T in this group was an indirect indicator of their overall comorbid burdens (despite lower ages). This example highlights the need for cautious interpretation of studies with inappropriately matched groupings and avoidance of accepting reported conclusions at face value.

Criteria for Low Testosterone

The definition utilized for TT and fT may contribute to outcomes reported. As different T thresholds may exist for different symptoms including libido, erectile function, muscle anabolism, bone turnover, and adipogenesis, supplementation beyond a set threshold likely has declining efficacy for that variable.[130, 131] Given these findings, studies utilizing lower T values for inclusion criteria are more likely to demonstrate improved outcomes with T supplementation, while those including men with higher baseline T may falsely conclude no

effect of therapy. This observation was highlighted by Corona and colleagues who performed a meta-analysis of the effects of T supplementation on erectile function and noted improvements in men with lower baseline T levels and no benefits for treatment of eugonadal men.[132]

The use of fT over TT is also associated with similar limitations based on the threshold value utilized. Araujo and colleagues performed a population-based observational survey of 1,475 men, aged 30–79 years to evaluate the association between hypogonadal symptoms and TT/fT levels.[2] The authors noted that increasing the lower fT threshold (50–60, or 60–70pg/ml [173.4–208 and 208–242.7 pmol/L]) resulted in reduced specificity for symptoms such as low libido (28.1%, 23.9%, and 19.8% for 50, 60, and 70pg/ml [242.7 pmol/L], respectively). This observation suggests that the reported increased specificity of fT over TT may reflect inappropriately matched threshold values rather than an intrinsic advantage of fT. This concept is also highlighted in a recent study demonstrating high correlation of hypogonadal symptoms between TT and fT when select cutpoints were used: TT 235 ng/dl (10.5 nmol/L), fT 98 pg/ml (220 pmol/L).[133]

Methodology for T sampling is also an important aspect of study design, as significant day-to-day, intra-individual and diurnal variability occurs. Studies reporting consistent early morning sampling, with one or more confirmatory tests obtained are more likely representative of a low T population rather than those with heterogenous sampling. The type of assay utilized for T assessment should be reported, with greater variability noted among non-MS TT samples (particularly at lower levels) and systematic overestimation resulting from non-empirically calculated fT measurements.[107]

Special Populations

Many hypogonadism publications are performed in specialized populations, with outcomes reported specifically for that cohort. Outcomes are not necessarily generalizable to other populations, and caution should be observed in extrapolating findings. Commonly evaluated groups include men with benign prostatic hyperplasia, baseline or men “at-risk” for cardiovascular disease, ED, sexual dysfunction, prostate cancer, or human immunodeficiency virus (HIV), among others.

One example of this occurs with cognitive symptoms, including mood and depression. Several placebo-controlled RCTs have evaluated the efficacy of T supplementation in men with depression, with seemingly contradictory results reported. A meta-analysis of studies noted significant benefits in select sub-populations, particularly among those with baseline low T and HIV.[134] The authors hypothesized that the greater benefits among HIV men may be secondary to additional improvements in energy in this select cohort. Similar or lesser improvements have been noted in other subgroups. This example highlights the need for qualitative statements identifying specific sub-populations among whom beneficial effects have been demonstrated, rather than generalized statements of improvements noted with T supplementation.

Therapeutic Agent Utilized

Various forms of T supplementation have been utilized including exogenous T (intramuscular, topical), SERMs, aromatase inhibitors, human chorionic gonadotropin, and varicocele ligation surgery, among others. Outcomes from studies using one form over another should not be used interchangeably, as each treatment results in varying effects on gonadotropins, estradiol, hormonal ratios, and potential unknown factors. This is evidenced by the significant variation on bone mineral density among the reported therapies.[126, 127, 135]

Similarly, varying formulations of T supplementation result in different peaks, troughs, and duration of therapeutic levels. The overall impact of the varying formulations on hypogonadal symptoms and AEs is poorly described, with limited data demonstrating higher rates of erythrocytosis noted among therapies achieving higher peak T levels.[136, 137]

Objective Measures and End-points

Study end-points and objective methods for data acquisition are important aspects of study design. Direct measures are always preferable to indirect, as they often provide more conclusive evidence of effect. On example is seen with the impact of aromatase inhibitors on bone mineral density. Although one publication of men undergoing anastrozole therapy over a 3-month period demonstrated no significant change in markers of bone resorption, a subsequent 12-month study demonstrated direct radiologic evidence of bone mineral density loss despite unchanged markers of bone resorption.[125, 126] These studies highlight limitations of indirect markers and suggest a need for reliance on objective measures which are able to definitively address the research question.

Study design and selected objective measures should coincide with defined study endpoints. Additional information obtained outside of study end-points should be interpreted with caution and require confirmation to support findings. The use of hypogonadism questionnaires as study endpoints is of questionable clinical utility and validity and should not be relied upon in lieu of objective laboratory, physiologic, or radiologic testing. More specific measures of individual symptoms such as the International Index of Erectile Function or the Becks Depression Inventory should be used to confirm improvements in hypogonadal symptoms rather than hypogonadism-specific questionnaires.

In addition to study end-points, data from interventional trials should be analyzed to assess the absolute and relative changes in T levels achieved with therapy. Common T therapies result in normalization of TT to eugonadal levels in approximately 75–85% of men, indicating a 15–25% rate of suboptimal response.[138, 139] Outcomes should be differentiated between responders and non-responders to more accurately assess the impact of T supplementation. To address this limitation, interventional trials may permit variable dosing.

Critical Evaluation of Meta-analyses and Systematic Reviews

The importance of study design and end-points is particularly relevant in regards to systematic reviews and meta-analyses, as often they extract data from studies of varying end-points, with resultant decreased methodological quality and reliability. Given the lack of a universally accepted standard for conducting meta-analyses, there is significant

heterogeneity in the quality and reliability of findings. Therefore, meta-analytic outcomes must also be scrutinized, particularly in regards to study inclusion and exclusion criteria. Meta-analyses including varying study designs are more likely to highlight erroneous findings, while highly restrictive meta-analyses may exclude a large portion of available data and thus are at risk of failing to identify significant outcomes. A comparative analysis between paper-based journal meta-analyses and the more rigorous Cochrane reviews demonstrated significant heterogeneity in the quality of published systematic reviews with relatively poor adherence to sound methodological principles with the majority of non-Cochrane reviews.[140]

Clinical Relevance versus Statistical Significance

Interpretation of study outcomes must also be analyzed from a broader clinical perspective. Statistically significant differences do not necessarily correlate with clinically relevant findings. As an example, Rosen and colleagues noted that clinically significant changes in IIEF scores varied depending on baseline erectile function status.[141] Men with mild ED would report a clinically relevant improvement if the IIEF score increased by two points, while those with severe ED required a seven point improvement to note equally significant changes. Similarly, statistically significant improvements in various indices (cardiovascular, diabetic, BMD, MetS, sleep apnea, depression) must be measured in terms of changes in actual rates of meeting criteria for diagnosis, changes in need for medications or clinical sequelae resulting from the condition.

Adverse Events

The methodology for AE reporting is an important aspect of study design and may impact the rate of AEs identified. Studies relying on unsolicited self-reports or physician reporting likely under-represent the true rate of symptoms when compared to patient-completed questionnaires. Objective AEs assessed through laboratory or radiological testing should be analyzed in regards to the duration of therapy prior to testing, percentage of patients completing testing, and relative changes compared to baseline status. Definitions as to what defines an AE (erythrocytosis, PSA elevation, PSA recurrence) should also be established *a priori* to reduce potential reporting bias.

Statistical Interpretation

A complete discussion on statistical interpretation is beyond the scope of the current manuscript, however, the clinical translatability of statistical outcomes will be reviewed. One of the most significant aspects of study design is statistical power. A measure of statistical power determines the ability of the study to detect significant differences based on anticipated deviations in a measured variable. Power analyses should be performed prior to study initiation and be based on previously published studies using similar measures. Power analyses are presented as a study's ability to detect a pre-determined percentage of change within a pre-selected percentage of confidence. Outcome measures with high variability (larger standard deviations among or within a population) and small changes following intervention require much larger patient populations to detect statistically significant differences. As an example, one meta-analyses evaluating prostate-specific events associated with T supplementation estimated that 85,862 men would be required in each arm of a 1-

year experimental study to identify a 20% increased risk of prostate cancer with 80% power at a 5% significance level.[142] Therefore, studies evaluating the risk of T supplementation on subsequent development of prostate cancer which have fewer than the minimum numbers are unlikely to have sufficient power to detect a significant difference, if present.

A second statistical concept, which is relevant to interpreting study outcomes is regarding statistical errors. Type I errors define the likelihood of falsely concluding an effect of therapy, which is otherwise not true. The most commonly accepted p-value for significance is defined as statistically likely to occur less than 5% of the time. By definition, a researcher evaluating 20 variables is likely to falsely identify at least one statistically significant outcome, which occurred by chance alone. Type II statistical errors occur when an anticipated effect is not observed despite its presence. This leads to concluding that a given therapy did not exhibit a selected effect, which would have been observed had sufficient patient numbers been evaluated.

An example of a type-II error can be given with the concept of smoking and lung cancer. If a study evaluating the effect of smoking in 100 healthy young men over three years concludes that smoking did not result in an increased rate of lung cancer in this group, it is committing a type-II error (assuming the accepted link between smoking and lung cancer). Similarly, studies of insufficient power, which conclude the absence of a particular AE with T supplementation are at risk of type-II errors. This is particularly relevant in regards to T supplementation in men with active or treated prostate cancer. The ability to detect the impact of T supplementation on prostate cancer progression, recurrence, cancer-specific mortality, and overall mortality requires large patient numbers prospectively evaluated in a RCT design over an extended period of time. As such, insufficiently powered studies which conclude safety of T supplementation in men with prostate cancer are at risk of committing type-II errors and are not statistically supportable.[143, 144]

Odds-ratios (OR) and confidence intervals (CI) are commonly reported and are methods of reporting the concordance of associated variables. For example, T supplementation is associated with an increased risk of hematocrits > 50%, with an odds-ratio of 3.7 and a 95% confidence interval of 1.82–7.51 in one study.[142] This suggests that men undergoing T supplementation are 3.7 times more likely to experience hematocrits >50% compared to those not receiving T supplementation. The confidence interval suggests that the authors are 95% confident that the actual odds-ratio falls somewhere between 1.82 and 7.51. This is significant because if the odds-ratio includes 1.0, by definition the finding cannot be statistically significant. Wider confidence intervals suggest greater variability with reduced likelihood that the OR reported represents the true value. As such, outcomes with wide CIs are less reliable compared to those with narrower intervals. To account for the variability in statistically significant outcomes, the Oxford criteria include CIs in grading the quality of evidence presented.[122]

Another relevant factor in interpreting study outcomes is the use of the coefficient of determination (R^2) to express the degree to which data fit a line or curve. These are often used to determine the extent to which predicted results match observed findings. Results are frequently reported between zero and one, with one representing a perfect fit, zero

suggesting no association and a value ≥0.5 representing a significant correlation. As with OR, R^2 does not imply a causative relationship between variables.

SUMMARY AND CONCLUSIONS

Critically appraising hypogonadism literature requires a thorough understanding of T physiology, natural history, and variability as well as limitations with current objective assessment techniques. Additionally, knowledge of various aspects of study design and methodology and their impact on outcomes is required to accurately interpret data in an unbiased manner.

Variability in T levels is dependent on several factors including advancing age, cultural/geographic location, life-style choices (exercise, obesity), and comorbid conditions, among others. Levels also vary in a diurnal manner and within individuals over time. Accurate assessment of T remains an ongoing challenge, with MS and equilibrium dialysis representing the gold-standard techniques for TT and fT, respectively. The role for fT over TT remains poorly defined, with no consensus available as to hormonal thresholds for hypogonadism. The use of hypogonadism-specific questionnaires for screening is not routinely performed, with their utility in the management of hypogonadism of debatable clinical value.

Study design and methodology contribute significantly to the quality of evidence presented and may directly impact reported outcomes. From an evidence-based perspective, retrospective and uncontrolled prospective studies are unable to establish causative relationships and are of lesser value than RCTs and meta-analyses of RCTs. The study duration and population included also are significant factors, which may result in inadvertent under or over-generalization of study findings. Results from trials using different T supplementation medications should not be used interchangeably, particularly in regards to specific outcomes.

End-points should be specifically defined prior to study onset, with objective measures selected to best directly assess the end-point selected. Findings unrelated to defined study end-points should be interpreted with caution and require subsequent confirmation. Data on T supplementation-associated AEs are frequently limited due to a lack of pre-defined AEs, exclusion of AEs as separate study end-points, and method of AE reporting.

In addition to study design and methodology, accurate interpretation of statistical analyses is essential to reduce misinterpretation of data. Conclusions may only be based on outcomes presented and are limited by the statistical power of the study design. Studies evaluating multiple factors are at an increased risk of performing type-I errors, while those with inadequate power are more likely to report type-II errors. Through a better understanding of T and its interpretation, distinguishing clinicians are better equipped to critically evaluate published literature and more accurately understand presented data.

LEGEND OF ABBREVIATIONS

5AR 5-alpha reductase

A-RHDQoL	Age-related hormone deficiency-dependent quality of life questionnaire
ADAM	Androgen deficiency in aging males
AE	Adverse event
AMS	Aging male's symptoms scale
AR	Androgen receptor
ASA	American society of andrology
CDC	Centers for disease control
CI	Confidence interval
DHT	Dihydrotestosterone
DM	Diabetes mellitus
EAA	European academy of andrology
EAU	European association of urology
EMAS	European male aging study
E2	Estradiol
EIA	Enzyme immunoassay
FAI	Free androgen index
fT	Free testosterone
GC-MS	Gas chromatography mass spectrometry
HIV	Human immunodeficiency virus
HRS	Hypogonadism-related symptom scale
IA	Immunoassay
ISA	International society of andrology
ISSAM	International society for the study of aging male
LC-MS/MS	Liquid chromatography tandem mass spectrometry
MMAS	Massachusetts male aging study
MS	Mass spectrometry
OR	Odds ratio
R^2	Coefficient of determination
RIA	Radioimmunoassay

SERM	Selective estrogen receptor modulator
SHBG	Sex hormone binding globulin
T	Testosterone
TT	Total testosterone
US	United States

References

1. Bhasin S, Cunningham GR, Hayes FJ, et al. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. The Journal of clinical endocrinology and metabolism. 2010; 95:2536–59. [PubMed: 20525905]
2. Araujo AB, Esche GR, Kupelian V, et al. Prevalence of symptomatic androgen deficiency in men. The Journal of clinical endocrinology and metabolism. 2007; 92:4241–7. [PubMed: 17698901]
3. Liu CC, Wu WJ, Lee YC, et al. The prevalence of and risk factors for androgen deficiency in aging Taiwanese men. The journal of sexual medicine. 2009; 6:936–46. [PubMed: 19210712]
4. Wu FC, Tajar A, Beynon JM, et al. Identification of late-onset hypogonadism in middle-aged and elderly men. The New England journal of medicine. 2010; 363:123–35. [PubMed: 20554979]
5. Spitzer M, Huang G, Basaria S, Travison TG, Bhasin S. Risks and benefits of testosterone therapy in older men. Nature reviews Endocrinology. 2013; 9:414–24.
6. Baillargeon J, Urban RJ, Ottenbacher KJ, Pierson KS, Goodwin JS. Trends in androgen prescribing in the United States, 2001 to 2011. JAMA Intern Med. 2013; 173:1465–6. [PubMed: 23939517]
7. Handelsman DJ. Global trends in testosterone prescribing, 2000–2011: expanding the spectrum of prescription drug misuse. The Medical journal of Australia. 2013; 199:548–51. [PubMed: 24138381]
8. Hayes FJ, DeCruz S, Seminara SB, Boepple PA, Crowley WF Jr. Differential regulation of gonadotropin secretion by testosterone in the human male: absence of a negative feedback effect of testosterone on follicle-stimulating hormone secretion. The Journal of clinical endocrinology and metabolism. 2001; 86:53–8. [PubMed: 11231978]
9. Schnorr JA, Bray MJ, Veldhuis JD. Aromatization mediates testosterone's short-term feedback restraint of 24-hour endogenously driven and acute exogenous gonadotropin-releasing hormone-stimulated luteinizing hormone and follicle-stimulating hormone secretion in young men. The Journal of clinical endocrinology and metabolism. 2001; 86:2600–6. [PubMed: 11397860]
10. Mazer NA. A novel spreadsheet method for calculating the free serum concentrations of testosterone, dihydrotestosterone, estradiol, estrone and cortisol: with illustrative examples from male and female populations. Steroids. 2009; 74:512–9. [PubMed: 19321131]
11. Carani C, Qin K, Simoni M, et al. Effect of testosterone and estradiol in a man with aromatase deficiency. The New England journal of medicine. 1997; 337:91–5. [PubMed: 9211678]
12. Pentikainen V, Erkkila K, Suomalainen L, Parvinen M, Dunkel L. Estradiol acts as a germ cell survival factor in the human testis in vitro. The Journal of clinical endocrinology and metabolism. 2000; 85:2057–67. [PubMed: 10843196]
13. Melcangi RC, Magnaghi V, Martini L. Steroid metabolism and effects in central and peripheral glial cells. Journal of neurobiology. 1999; 40:471–83. [PubMed: 10453050]
14. Andersson S, Russell DW. Structural and biochemical properties of cloned and expressed human and rat steroid 5 alpha-reductases. Proceedings of the National Academy of Sciences of the United States of America. 1990; 87:3640–4. [PubMed: 2339109]
15. Schwartz JI, Tanaka WK, Wang DZ, et al. MK-386, an inhibitor of 5alpha-reductase type 1, reduces dihydrotestosterone concentrations in serum and sebum without affecting dihydrotestosterone concentrations in semen. The Journal of clinical endocrinology and metabolism. 1997; 82:1373–7. [PubMed: 9141518]

16. Thigpen AE, Silver RI, Guileyardo JM, Casey ML, McConnell JD, Russell DW. Tissue distribution and ontogeny of steroid 5 alpha-reductase isozyme expression. The Journal of clinical investigation. 1993; 92:903–10. [PubMed: 7688765]
17. Kim KS, Liu W, Cunha GR, et al. Expression of the androgen receptor and 5 alpha-reductase type 2 in the developing human fetal penis and urethra. Cell and tissue research. 2002; 307:145–53. [PubMed: 11845321]
18. Amory JK, Wang C, Swerdloff RS, et al. The effect of 5alpha-reductase inhibition with dutasteride and finasteride on semen parameters and serum hormones in healthy men. The Journal of clinical endocrinology and metabolism. 2007; 92:1659–65. [PubMed: 17299062]
19. Traish AM, Goldstein I, Kim NN. Testosterone and erectile function: from basic research to a new clinical paradigm for managing men with androgen insufficiency and erectile dysfunction. European urology. 2007; 52:54–70. [PubMed: 17329016]
20. Traish AM, Guay AT. Are androgens critical for penile erections in humans? Examining the clinical and preclinical evidence. The journal of sexual medicine. 2006; 3:382–404. discussion 04–7. [PubMed: 16681465]
21. Yassin AA, Saad F, Traish A. Testosterone undecanoate restores erectile function in a subset of patients with venous leakage: a series of case reports. The journal of sexual medicine. 2006; 3:727–35. [PubMed: 16839330]
22. Agis-Balboa RC, Pinna G, Zhubi A, et al. Characterization of brain neurons that express enzymes mediating neurosteroid biosynthesis. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:14602–7. [PubMed: 16984997]
23. Majewska MD, Harrison NL, Schwartz RD, Barker JL, Paul SM. Steroid hormone metabolites are barbiturate-like modulators of the GABA receptor. Science. 1986; 232:1004–7. [PubMed: 2422758]
24. Borst SE, Yarrow JF, Conover CF, et al. Musculoskeletal and prostate effects of combined testosterone and finasteride administration in older hypogonadal men: a randomized, controlled trial. American journal of physiology Endocrinology and metabolism. 2014; 306:E433–42. [PubMed: 24326421]
25. Bang HJ, Yang YJ, Lho DS, Lee WY, Sim WY, Chung BC. Comparative studies on level of androgens in hair and plasma with premature male-pattern baldness. Journal of dermatological science. 2004; 34:11–6. [PubMed: 14757277]
26. Zhang MG, Wu W, Zhang CM, et al. Effects of oral finasteride on erectile function in a rat model. The journal of sexual medicine. 2012; 9:1328–36. [PubMed: 22375859]
27. Pinsky MR, Gur S, Tracey AJ, Harbin A, Hellstrom WJ. The effects of chronic 5-alpha-reductase inhibitor (dutasteride) treatment on rat erectile function. The journal of sexual medicine. 2011; 8:3066–74. [PubMed: 21834872]
28. Imperato-McGinley J, Guerrero L, Gautier T, Peterson RE. Steroid 5-alpha-reductase deficiency in man: an inherited form of male pseudohermaphroditism. Science. 1974; 186:1213–5. [PubMed: 4432067]
29. Katz MD, Kligman I, Cai LQ, et al. Paternity by intrauterine insemination with sperm from a man with 5alpha-reductase-2 deficiency. The New England journal of medicine. 1997; 336:994–7. [PubMed: 9077378]
30. Ivarsson SA, Nielsen MD, Lindberg T. Male pseudohermaphroditism due to 5 alpha-reductase deficiency in a Swedish family. European journal of pediatrics. 1988; 147:532–5. [PubMed: 3409930]
31. Mifsud A, Sim CK, Boettger-Tong H, et al. Trinucleotide (CAG) repeat polymorphisms in the androgen receptor gene: molecular markers of risk for male infertility. Fertility and sterility. 2001; 75:275–81. [PubMed: 11172827]
32. Zitzmann M, Brune M, Kornmann B, Gromoll J, Junker R, Nieschlag E. The CAG repeat polymorphism in the androgen receptor gene affects bone density and bone metabolism in healthy males. Clinical endocrinology. 2001; 55:649–57. [PubMed: 11894977]
33. Schneider G, Nienhaus K, Gromoll J, Heuft G, Nieschlag E, Zitzmann M. Sex hormone levels, genetic androgen receptor polymorphism, and anxiety in >/=50-year-old males. The journal of sexual medicine. 2011; 8:3452–64. [PubMed: 21883946]

34. Dhiman P, Bhansali A, Prasad R, Dutta P, Walia R, Ravikiran M. Predictors of pilosebaceous unit responsiveness to testosterone therapy in patients with hypogonadotrophic hypogonadism. Andrologia. 2011; 43:422–7. [PubMed: 21486418]
35. Yeap BB, Alfonso H, Chubb SA, et al. Reference ranges and determinants of testosterone, dihydrotestosterone, and estradiol levels measured using liquid chromatography-tandem mass spectrometry in a population-based cohort of older men. The Journal of clinical endocrinology and metabolism. 2012; 97:4030–9. [PubMed: 22977273]
36. Bhasin S, Pencina M, Jasuja GK, et al. Reference ranges for testosterone in men generated using liquid chromatography tandem mass spectrometry in a community-based sample of healthy nonobese young men in the Framingham Heart Study and applied to three geographically distinct cohorts. The Journal of clinical endocrinology and metabolism. 2011; 96:2430–9. [PubMed: 21697255]
37. Feldman HA, Longcope C, Derby CA, et al. Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. The Journal of clinical endocrinology and metabolism. 2002; 87:589–98. [PubMed: 11836290]
38. Ferrini RL, Barrett-Connor E. Sex hormones and age: a cross-sectional study of testosterone and estradiol and their bioavailable fractions in community-dwelling men. American journal of epidemiology. 1998; 147:750–4. [PubMed: 9554416]
39. Harman SM, Metter EJ, Tobin JD, Pearson J, Blackman MR. Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. The Journal of clinical endocrinology and metabolism. 2001; 86:724–31. [PubMed: 11158037]
40. Shi Z, Araujo AB, Martin S, O'Loughlin P, Wittert GA. Longitudinal changes in testosterone over five years in community-dwelling men. The Journal of clinical endocrinology and metabolism. 2013; 98:3289–97. [PubMed: 23775354]
41. Sartorius G, Spasevska S, Idan A, et al. Serum testosterone, dihydrotestosterone and estradiol concentrations in older men self-reporting very good health: the healthy man study. Clinical endocrinology. 2012; 77:755–63. [PubMed: 22563890]
42. Plymate SR, Tenover JS, Bremner WJ. Circadian variation in testosterone, sex hormone-binding globulin, and calculated non-sex hormone-binding globulin bound testosterone in healthy young and elderly men. Journal of andrology. 1989; 10:366–71. [PubMed: 2592266]
43. Bremner WJ, Vitiello MV, Prinz PN. Loss of circadian rhythmicity in blood testosterone levels with aging in normal men. The Journal of clinical endocrinology and metabolism. 1983; 56:1278–81. [PubMed: 6841562]
44. Crawford ED, Barqawi AB, O'Donnell C, Morgentaler A. The association of time of day and serum testosterone concentration in a large screening population. BJU international. 2007; 100:509–13. [PubMed: 17555474]
45. Diver MJ, Imtiaz KE, Ahmad AM, Vora JP, Fraser WD. Diurnal rhythms of serum total, free and bioavailable testosterone and of SHBG in middle-aged men compared with those in young men. Clinical endocrinology. 2003; 58:710–7. [PubMed: 12780747]
46. Brambilla DJ, Matsumoto AM, Araujo AB, McKinlay JB. The effect of diurnal variation on clinical measurement of serum testosterone and other sex hormone levels in men. The Journal of clinical endocrinology and metabolism. 2009; 94:907–13. [PubMed: 19088162]
47. Panizzon MS, Hauger R, Jacobson KC, et al. Genetic and environmental influences of daily and intra-individual variation in testosterone levels in middle-aged men. Psychoneuroendocrinology. 2013
48. Axelsson J, Ingre M, Akerstedt T, Holmback U. Effects of acutely displaced sleep on testosterone. The Journal of clinical endocrinology and metabolism. 2005; 90:4530–5. [PubMed: 15914523]
49. Rohrmann S, Nelson WG, Rifai N, et al. Serum estrogen, but not testosterone, levels differ between black and white men in a nationally representative sample of Americans. The Journal of clinical endocrinology and metabolism. 2007; 92:2519–25. [PubMed: 17456570]
50. Litman HJ, Bhasin S, Link CL, Araujo AB, McKinlay JB. Serum androgen levels in black, Hispanic, and white men. The Journal of clinical endocrinology and metabolism. 2006; 91:4326–34. [PubMed: 16912139]

51. Wu AH, Whittemore AS, Kolonel LN, et al. Serum androgens and sex hormone-binding globulins in relation to lifestyle factors in older African-American, white, and Asian men in the United States and Canada. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 1995; 4:735–41.
52. Orwoll ES, Nielson CM, Labrie F, et al. Evidence for geographical and racial variation in serum sex steroid levels in older men. The Journal of clinical endocrinology and metabolism. 2010; 95:E151–60. [PubMed: 20668046]
53. Harris JA, Vernon PA, Boomsma DI. The heritability of testosterone: a study of Dutch adolescent twins and their parents. Behavior genetics. 1998; 28:165–71. [PubMed: 9670592]
54. Hoekstra RA, Bartels M, Boomsma DI. Heritability of testosterone levels in 12-year-old twins and its relation to pubertal development. Twin research and human genetics: the official journal of the International Society for Twin Studies. 2006; 9:558–65. [PubMed: 16899163]
55. Bogaert V, Taes Y, Konings P, et al. Heritability of blood concentrations of sex-steroids in relation to body composition in young adult male siblings. Clinical endocrinology. 2008; 69:129–35. [PubMed: 18598274]
56. Kuijper EA, Lambalk CB, Boomsma DI, et al. Heritability of reproductive hormones in adult male twins. Hum Reprod. 2007; 22:2153–9. [PubMed: 17569675]
57. Ring HZ, Lessov CN, Reed T, et al. Heritability of plasma sex hormones and hormone binding globulin in adult male twins. The Journal of clinical endocrinology and metabolism. 2005; 90:3653–8. [PubMed: 15755867]
58. Meikle AW, Smith JA, West DW. Familial factors affecting prostatic cancer risk and plasma sex-steroid levels. The Prostate. 1985; 6:121–8. [PubMed: 3975174]
59. Valero-Politi J, Fuentes-Arderiu X. Annual rhythmic variations of follitropin, lutropin, testosterone and sex-hormone-binding globulin in men. Clinica chimica acta; international journal of clinical chemistry. 1998; 271:57–71. [PubMed: 9564557]
60. Ruhayel Y, Malm G, Haugen TB, et al. Seasonal variation in serum concentrations of reproductive hormones and urinary excretion of 6-sulfatoxymelatonin in men living north and south of the Arctic Circle: a longitudinal study. Clinical endocrinology. 2007; 67:85–92. [PubMed: 17547693]
61. Nicolau GY, Lakatua D, Sackett-Lundeen L, Haus E. Circadian and circannual rhythms of hormonal variables in elderly men and women. Chronobiology international. 1984; 1:301–19. [PubMed: 6600031]
62. Brambilla DJ, O'Donnell AB, Matsumoto AM, McKinlay JB. Lack of seasonal variation in serum sex hormone levels in middle-aged to older men in the Boston area. The Journal of clinical endocrinology and metabolism. 2007; 92:4224–9. [PubMed: 17684044]
63. Martikainen H, Tapanainen J, Vakkuri O, Leppaluoto J, Huhtaniemi I. Circannual concentrations of melatonin, gonadotrophins, prolactin and gonadal steroids in males in a geographical area with a large annual variation in daylight. Acta endocrinologica. 1985; 109:446–50. [PubMed: 3929512]
64. Morley JE, Patrick P, Perry HM 3rd. Evaluation of assays available to measure free testosterone. Metabolism: clinical and experimental. 2002; 51:554–9. [PubMed: 11979385]
65. Ahokoski O, Virtanen A, Huupponen R, et al. Biological day-to-day variation and daytime changes of testosterone, follitropin, lutropin and oestradiol-17beta in healthy men. Clinical chemistry and laboratory medicine: CCLM/FESCC. 1998; 36:485–91.
66. Valero-Politi J, Fuentes-Arderiu X. Within- and between-subject biological variations of follitropin, lutropin, testosterone, and sex-hormone-binding globulin in men. Clinical chemistry. 1993; 39:1723–5. [PubMed: 8353962]
67. Brambilla DJ, O'Donnell AB, Matsumoto AM, McKinlay JB. Intraindividual variation in levels of serum testosterone and other reproductive and adrenal hormones in men. Clinical endocrinology. 2007; 67:853–62. [PubMed: 18052942]
68. Travison TG, Araujo AB, Kupelian V, O'Donnell AB, McKinlay JB. The relative contributions of aging, health, and lifestyle factors to serum testosterone decline in men. The Journal of clinical endocrinology and metabolism. 2007; 92:549–55. [PubMed: 17148559]
69. Lima N, Cavaliere H, Knobel M, Halpern A, Medeiros-Neto G. Decreased androgen levels in massively obese men may be associated with impaired function of the gonadostat. International

journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity. 2000; 24:1433–7.

70. Hammoud A, Gibson M, Hunt SC, et al. Effect of Roux-en-Y gastric bypass surgery on the sex steroids and quality of life in obese men. The Journal of clinical endocrinology and metabolism. 2009; 94:1329–32. [PubMed: 19174499]
71. Corona G, Rastrelli G, Monami M, et al. Body weight loss reverts obesity-associated hypogonadotropic hypogonadism: a systematic review and meta-analysis. European journal of endocrinology/European Federation of Endocrine Societies. 2013; 168:829–43.
72. Khoo J, Tian HH, Tan B, et al. Comparing effects of low- and high-volume moderate-intensity exercise on sexual function and testosterone in obese men. The journal of sexual medicine. 2013; 10:1823–32. [PubMed: 23635309]
73. Cadore EL, Lhullier FL, Alberton CL, et al. Salivary hormonal responses to different water-based exercise protocols in young and elderly men. Journal of strength and conditioning research/National Strength & Conditioning Association. 2009; 23:2695–701.
74. Ahtiainen JP, Pakarinen A, Kraemer WJ, Hakkinen K. Acute hormonal responses to heavy resistance exercise in strength athletes versus nonathletes. Canadian journal of applied physiology = Revue canadienne de physiologie appliquee. 2004; 29:527–43. [PubMed: 15507691]
75. Cadore EL, Lhullier FL, Brentano MA, et al. Hormonal responses to resistance exercise in long-term trained and untrained middle-aged men. Journal of strength and conditioning research/National Strength & Conditioning Association. 2008; 22:1617–24.
76. Svartberg J, Midtby M, Bonaa KH, Sundsfjord J, Joakimsen RM, Jorde R. The associations of age, lifestyle factors and chronic disease with testosterone in men: the Tromso Study. European journal of endocrinology/European Federation of Endocrine Societies. 2003; 149:145–52.
77. Ponholzer A, Plas E, Schatzl G, et al. Relationship between testosterone serum levels and lifestyle in aging men. The aging male: the official journal of the International Society for the Study of the Aging Male. 2005; 8:190–3.
78. Field AE, Colditz GA, Willett WC, Longcope C, McKinlay JB. The relation of smoking, age, relative weight, and dietary intake to serum adrenal steroids, sex hormones, and sex hormone-binding globulin in middle-aged men. The Journal of clinical endocrinology and metabolism. 1994; 79:1310–6. [PubMed: 7962322]
79. Jeng HA, Chen YL, Kantaria KN. Association of cigarette smoking with reproductive hormone levels and semen quality in healthy adult men in Taiwan. Journal of environmental science and health Part A, Toxic/hazardous substances & environmental engineering. 2014; 49:262–8.
80. Halmenschlager G, Rossetto S, Lara GM, Rhoden EL. Evaluation of the effects of cigarette smoking on testosterone levels in adult men. The journal of sexual medicine. 2009; 6:1763–72. [PubMed: 19473474]
81. Muehlenbein MP, Hirschtick JL, Bonner JZ, Swartz AM. Toward quantifying the usage costs of human immunity: Altered metabolic rates and hormone levels during acute immune activation in men. American journal of human biology: the official journal of the Human Biology Council. 2010; 22:546–56. [PubMed: 20309883]
82. Carvalho VM. The coming of age of liquid chromatography coupled to tandem mass spectrometry in the endocrinology laboratory. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2012; 883–884:50–8.
83. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. The Journal of clinical endocrinology and metabolism. 2007; 92:405–13. [PubMed: 17090633]
84. Albrecht L, Styne D. Laboratory testing of gonadal steroids in children. Pediatric endocrinology reviews: PER. 2007; 5(Suppl 1):599–607. [PubMed: 18167469]
85. Wierman ME, Basson R, Davis SR, et al. Androgen therapy in women: an Endocrine Society Clinical Practice guideline. The Journal of clinical endocrinology and metabolism. 2006; 91:3697–710. [PubMed: 17018650]
86. Wang C, Catlin DH, Demers LM, Starcevic B, Swerdloff RS. Measurement of total serum testosterone in adult men: comparison of current laboratory methods versus liquid

chromatography-tandem mass spectrometry. The Journal of clinical endocrinology and metabolism. 2004; 89:534–43. [PubMed: 14764758]

87. Vesper HW, Bhasin S, Wang C, et al. Interlaboratory comparison study of serum total testosterone [corrected] measurements performed by mass spectrometry methods. Steroids. 2009; 74:498–503. [PubMed: 19428438]
88. Steinberger E, Ayala C, Hsi B, et al. Utilization of commercial laboratory results in management of hyperandrogenism in women. Endocrine practice: official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists. 1998; 4:1–10.
89. Vesper HW, Miller WG, Myers GL. Reference materials and commutability. The Clinical biochemist Reviews/Australian Association of Clinical Biochemists. 2007; 28:139–47.
90. Vesper HW, Botelho JC. Standardization of testosterone measurements in humans. The Journal of steroid biochemistry and molecular biology. 2010; 121:513–9. [PubMed: 20302935]
91. Control CfD.
92. 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. European journal of endocrinology/European Federation of Endocrine Societies. 2008; 159:507–14.
93. Bhasin S, Zhang A, Coviello A, et al. The impact of assay quality and reference ranges on clinical decision making in the diagnosis of androgen disorders. Steroids. 2008; 73:1311–7. [PubMed: 18687348]
94. Dunn JF, Nisula BC, Rodbard D. Transport of steroid hormones: binding of 21 endogenous steroids to both testosterone-binding globulin and corticosteroid-binding globulin in human plasma. The Journal of clinical endocrinology and metabolism. 1981; 53:58–68. [PubMed: 7195404]
95. Hammond GL, Wu TS, Simard M. Evolving utility of sex hormone-binding globulin measurements in clinical medicine. Current opinion in endocrinology, diabetes, and obesity. 2012; 19:183–9.
96. van den Beld AW, de Jong FH, Grobbee DE, Pols HA, Lamberts SW. Measures of bioavailable serum testosterone and estradiol and their relationships with muscle strength, bone density, and body composition in elderly men. The Journal of clinical endocrinology and metabolism. 2000; 85:3276–82. [PubMed: 10999822]
97. Almeida OP, Yeap BB, Hankey GJ, Jamrozik K, Flicker L. Low free testosterone concentration as a potentially treatable cause of depressive symptoms in older men. Archives of general psychiatry. 2008; 65:283–9. [PubMed: 18316674]
98. Brand JS, van der Tweel I, Grobbee DE, Emmelot-Vonk MH, van der Schouw YT. Testosterone, sex hormone-binding globulin and the metabolic syndrome: a systematic review and meta-analysis of observational studies. International journal of epidemiology. 2011; 40:189–207. [PubMed: 20870782]
99. Coviello AD, Haring R, Wellons M, et al. A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. PLoS Genet. 2012; 8:e1002805. [PubMed: 22829776]
100. Vanbillemont G, Bogaert V, De Bacquer D, et al. Polymorphisms of the SHBG gene contribute to the interindividual variation of sex steroid hormone blood levels in young, middle-aged and elderly men. Clinical endocrinology. 2009; 70:303–10. [PubMed: 18681858]
101. Hammes A, Andreassen TK, Spoelgen R, et al. Role of endocytosis in cellular uptake of sex steroids. Cell. 2005; 122:751–62. [PubMed: 16143106]
102. Sodergard R, Backstrom T, Shanbhag V, Carstensen H. Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins at body temperature. Journal of steroid biochemistry. 1982; 16:801–10. [PubMed: 7202083]
103. Nanjee MN, Wheeler MJ. Plasma free testosterone--is an index sufficient? Annals of clinical biochemistry. 1985; 22(Pt 4):387–90. [PubMed: 4041156]
104. Ly LP, Handelsman DJ. Empirical estimation of free testosterone from testosterone and sex hormone-binding globulin immunoassays. European journal of endocrinology/European Federation of Endocrine Societies. 2005; 152:471–8.

105. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. The Journal of clinical endocrinology and metabolism. 1999; 84:3666–72. [PubMed: 10523012]

106. Morris PD, Malkin CJ, Channer KS, Jones TH. A mathematical comparison of techniques to predict biologically available testosterone in a cohort of 1072 men. European journal of endocrinology/European Federation of Endocrine Societies. 2004; 151:241–9.

107. Sartorius G, Ly LP, Sikaris K, McLachlan R, Handelsman DJ. Predictive accuracy and sources of variability in calculated free testosterone estimates. Annals of clinical biochemistry. 2009; 46:137–43. [PubMed: 19225026]

108. Zakharov MN, Bhasin S, Travison TG, et al. A multi-step, dynamic allosteric model of testosterone's binding to sex hormone binding globulin. Molecular and cellular endocrinology. 2015; 399:190–200. [PubMed: 25240469]

109. Heinemann LA, Zimmermann T, Vermeulen A, Thiel C. A new 'Aging Male's Symptom' rating scale. The aging male: the official journal of the International Society for the Study of the Aging Male. 1999; 2:105–14.

110. Morley JE, Charlton E, Patrick P, et al. Validation of a screening questionnaire for androgen deficiency in aging males. Metabolism: clinical and experimental. 2000; 49:1239–42. [PubMed: 11016912]

111. Smith KW, Feldman HA, McKinlay JB. Construction and field validation of a self-administered screener for testosterone deficiency (hypogonadism) in ageing men. Clinical endocrinology. 2000; 53:703–11. [PubMed: 11155092]

112. Wiltink J, Beutel ME, Brahler E, Weidner W. Hypogonadism-related symptoms: development and evaluation of an empirically derived self-rating instrument (HRS 'Hypogonadism Related Symptom Scale'). Andrologia. 2009; 41:297–304. [PubMed: 19737277]

113. McMillan CV, Bradley C, Giannoulis M, Martin F, Sonksen PH. Preliminary development of a new individualised questionnaire measuring quality of life in older men with age-related hormonal decline: the A-RHDQoL. Health and quality of life outcomes. 2003; 1:51. [PubMed: 14613571]

114. Corona G, Mannucci E, Petrone L, et al. ANDROTEST: a structured interview for the screening of hypogonadism in patients with sexual dysfunction. The journal of sexual medicine. 2006; 3:706–15. [PubMed: 16839327]

115. Morley JE, Perry HM 3rd, Kevorkian RT, Patrick P. Comparison of screening questionnaires for the diagnosis of hypogonadism. Maturitas. 2006; 53:424–9. [PubMed: 16140484]

116. Blumel JE, Chedraui P, Gili SA, Navarro A, Valenzuela K, Vallejo S. Is the Androgen Deficiency of Aging Men (ADAM) questionnaire useful for the screening of partial androgenic deficiency of aging men? Maturitas. 2009; 63:365–8. [PubMed: 19481382]

117. Tancredi A, Reginster JY, Schleich F, et al. Interest of the androgen deficiency in aging males (ADAM) questionnaire for the identification of hypogonadism in elderly community-dwelling male volunteers. European journal of endocrinology/European Federation of Endocrine Societies. 2004; 151:355–60.

118. Spetz AC, Palmefors L, Skobe RS, et al. Testosterone correlated to symptoms of partial androgen deficiency in aging men (PADAM) in an elderly Swedish population. Menopause. 2007; 14:999–1005. [PubMed: 17529900]

119. Emmelot-Vonk MH, Verhaar HJ, Nakhai-Pour HR, Grobbee DE, van der Schouw YT. Low testosterone concentrations and the symptoms of testosterone deficiency according to the Androgen Deficiency in Ageing Males (ADAM) and Ageing Males' Symptoms rating scale (AMS) questionnaires. Clinical endocrinology. 2011; 74:488–94. [PubMed: 21138462]

120. Heinemann LA, Saad F, Heinemann K, Thai DM. Can results of the Aging Males' Symptoms (AMS) scale predict those of screening scales for androgen deficiency? The aging male: the official journal of the International Society for the Study of the Aging Male. 2004; 7:211–8.

121. Ho CC, Tong SF, Low WY, et al. A randomized, double-blind, placebo-controlled trial on the effect of long-acting testosterone treatment as assessed by the Aging Male Symptoms scale. BJU international. 2012; 110:260–5. [PubMed: 22093057]

122. Medicine CfEB. Oxford Centre for Evidence-based Medicine - Levels of Evidence. 2009.

123. Shores MM, Smith NL, Forsberg CW, Anawalt BD, Matsumoto AM. Testosterone treatment and mortality in men with low testosterone levels. The Journal of clinical endocrinology and metabolism. 2012; 97:2050–8. [PubMed: 22496507]

124. Fernandez-Balsells MM, Murad MH, Lane M, et al. Clinical review 1: Adverse effects of testosterone therapy in adult men: a systematic review and meta-analysis. The Journal of clinical endocrinology and metabolism. 2010; 95:2560–75. [PubMed: 20525906]

125. Leder BZ, Finkelstein JS. Effect of aromatase inhibition on bone metabolism in elderly hypogonadal men. Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2005; 16:1487–94.

126. Burnett-Bowie SA, McKay EA, Lee H, Leder BZ. Effects of aromatase inhibition on bone mineral density and bone turnover in older men with low testosterone levels. The Journal of clinical endocrinology and metabolism. 2009; 94:4785–92. [PubMed: 19820017]

127. Aversa A, Bruzziches R, Francomano D, et al. Effects of long-acting testosterone undecanoate on bone mineral density in middle-aged men with late-onset hypogonadism and metabolic syndrome: results from a 36 months controlled study. The aging male: the official journal of the International Society for the Study of the Aging Male. 2012; 15:96–102.

128. Tenover JS, Bremner WJ. The effects of normal aging on the response of the pituitary-gonadal axis to chronic clomiphene administration in men. Journal of andrology. 1991; 12:258–63. [PubMed: 1917692]

129. Vigen R, O'Donnell CI, Baron AE, et al. Association of testosterone therapy with mortality, myocardial infarction, and stroke in men with low testosterone levels. JAMA: the journal of the American Medical Association. 2013; 310:1829–36. [PubMed: 24193080]

130. Finkelstein JS, Lee H, Burnett-Bowie SA, et al. Gonadal steroids and body composition, strength, and sexual function in men. The New England journal of medicine. 2013; 369:1011–22. [PubMed: 24024838]

131. Snyder PJ, Peachey H, Hannoush P, et al. Effect of testosterone treatment on bone mineral density in men over 65 years of age. The Journal of clinical endocrinology and metabolism. 1999; 84:1966–72. [PubMed: 10372695]

132. Corona G, Isidori AM, Buvat J, et al. Testosterone supplementation and sexual function: a meta-analysis study. The journal of sexual medicine. 2014; 11:1577–92. [PubMed: 24697970]

133. Antonio L, Wu FC, O'Neill TW, et al. Low Free Testosterone is Associated with Hypogonadal Signs and Symptoms in Men with Normal Total Testosterone. The Journal of clinical endocrinology and metabolism. 2016:jc20154106.

134. Zarrouf FA, Artz S, Griffith J, Sirbu C, Kommor M. Testosterone and depression: systematic review and meta-analysis. Journal of psychiatric practice. 2009; 15:289–305. [PubMed: 19625884]

135. Moskovic DJ, Katz DJ, Akhavan A, Park K, Mulhall JP. Clomiphene citrate is safe and effective for long-term management of hypogonadism. BJU international. 2012; 110:1524–8. [PubMed: 22458540]

136. Ip FF, di Pierro I, Brown R, Cunningham I, Handelsman DJ, Liu PY. Trough serum testosterone predicts the development of polycythemia in hypogonadal men treated for up to 21 years with subcutaneous testosterone pellets. European journal of endocrinology/European Federation of Endocrine Societies. 2010; 162:385–90.

137. Vorkas CK, Vaamonde CM, Glesby MJ. Testosterone replacement therapy and polycythemia in HIV-infected patients. AIDS. 2012; 26:243–5. [PubMed: 22008652]

138. Kaufman JM, Miller MG, Garwin JL, Fitzpatrick S, McWhirter C, Brennan JJ. Efficacy and safety study of 1.62% testosterone gel for the treatment of hypogonadal men. The journal of sexual medicine. 2011; 8:2079–89. [PubMed: 21492400]

139. Mazer N, Bell D, Wu J, Fischer J, Cosgrove M, Eilers B. Comparison of the steady-state pharmacokinetics, metabolism, and variability of a transdermal testosterone patch versus a transdermal testosterone gel in hypogonadal men. The journal of sexual medicine. 2005; 2:213–26. [PubMed: 16422889]

140. Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. JAMA: the journal of the American Medical Association. 1998; 280:278–80. [PubMed: 9676681]

141. Rosen RC, Allen KR, Ni X, Araujo AB. Minimal clinically important differences in the erectile function domain of the International Index of Erectile Function scale. European urology. 2011; 60:1010–6. [PubMed: 21855209]

142. Calof OM, Singh AB, Lee ML, et al. Adverse events associated with testosterone replacement in middle-aged and older men: a meta-analysis of randomized, placebo-controlled trials. The journals of gerontology Series A, Biological sciences and medical sciences. 2005; 60:1451–7.

143. Kaplan AL, Hu JC. Use of testosterone replacement therapy in the United States and its effect on subsequent prostate cancer outcomes. Urology. 2013; 82:321–6. [PubMed: 23706552]

144. Pastuszak AW, Pearlman AM, Lai WS, et al. Testosterone Replacement Therapy in Patients with Prostate Cancer after Radical Prostatectomy. The Journal of urology. 2013

TAKE HOME MESSAGE

An accurate diagnosis of male hypogonadism depends on reliable assessments of testosterone values. The current manuscript reviews factors which impact testosterone levels in situ and the accuracy of laboratory assessments and assists readers in interpreting outcomes of testosterone literature.

J Sex Med. Author manuscript; available in PMC 2017 July 19.

Table 1

Summary of Factors Impacting Testosterone Variability

Variable/Impact on Testosterone Measurements
Acute and Chronic Disease • 10–30% decline with acute respiratory illness in young men[81] • Chronic illness and increasing medication usage associated with more rapid age-related T decline[68]
Age • T levels decline with age[35–39] • Factors contributing to age-related decline include T include obesity, chronic disease, comorbid conditions, lifestyle choices, and medication use[40] • Age as an independent factor resulting in T variation is debatable[37, 40, 41]
Assay Techniques • IA vs MS results in −14.1% to 19.2% variability and ±40% at T < 100 ng/dl[86, 87] • At low T, IA varies by 2.7 to 14.3 fold[88] • Specimen handling, preparation, and commutability, calibration methods, and matrix interference introduce variability • MS is gold standard for TT, equilibrium dialysis for fT • Variability between calculated fT methods is ~14% with empiric methods most concordant[107] • Non-empiric fT calculations overestimate the true value[107]
Diurnal Variation • Peak concentration in AM • Rapidly decreases after waking • 4 PM vs 8 AM values 20–25% lower in 30–40 year old males and 10% lower in 70 year olds[46] • SHBG, fT, and bioavailable T also vary diurnally
Ethnicity • Likely minimal to no clinically-relevant impact
Genetics • Accounts for 42–65% of T variability[47, 55–57]
Geography

Variable/Impact on Testosterone Measurements
• May result in variations in T levels • Hong Kong and Japan with ~20% higher T compared to Sweden, Tobago, and the US[52]
Intra-individual
• Repeated measures vary from 65–153%[67] • Up to 50% of men with T < 300 ng/dl will be > 300 ng/dl on repeat testing[64] • Averaging 2–3 tests reduces range variability by 30% and 43%, respectively[67]
Lifestyle Factors
• Obesity inversely associated with T • 4–5 point BMI increase associated with 10-year equivalent T decline[68] • Loss of body fat increases T[71] • Exercise increases T in an intensity, duration, and age-dependent manner[72–75] • Smoking has unclear effect on T with contradictory studies available[68, 76–80] • Moderate alcohol intake does not likely significantly impact T[76–78]
Seasonal
• Conflicting data on seasonal variability in T levels[59–63]

BMI - Body mass index; fT - Free testosterone; IA - Immunoassay; MS - Mass spectrometry; SHBG - Sex hormone binding globulin; T - Testosterone; TT - Total testosterone; VS - versus

Table 2

Summary of Advantages and Disadvantages of Various Testosterone Assays

Assay/Description	Advantages	Disadvantages
Total Testosterone		
Immunoassay		
• Serum mixed with T antibodies and tracer element • Tracer can be a radioisotope (RIA), enzyme (EIA), fluorescent, or chemiluminescent compound	• Rapid and simple • Inexpensive • Commonly Utilized • Numerous reference intervals in varied populations • Acceptable performance at normal T levels • High throughput	• T extraction/chromatography requires additional time, technical expertise, and generates organic waste • Reduced accuracy at low/high T • RIA generates radioactive waste
Mass Spectrometry		
• Ionizes molecules and measures mass-to-charge ratios	• Gold-standard technique • Excellent sensitivity and specificity • Accurate at high/low T • Simultaneous measurement of multiple steroids • High throughput (LC-MS/MS)	• Expensive • Requires calibration • GC-MS requires extensive sample preparation and expertise • Extraction may generate organic waste
Free Testosterone		
Ammonium sulfate precipitation (bioavailable testosterone)		
• Tracer-labeled T added to serum • SHBG precipitated via addition of ammonium sulfate • Remaining tracer multiplied by TT	• Correlates well with equilibrium dialysis	• Time/labor intensive • Tracer impurities may compromise results • Relies on accuracy of TT assay
Calculated FT		
• Law of mass action (Nanjee and Wheeler, Sodergard, Vermeulen)	• Correlates well with equilibrium dialysis • Rapid and simple	• Relies on equilibrium dissociation constants for binding of SHBG and albumin to T • Relies on TT and SHBG assay accuracy • Tends to overestimate true value

Assay/Description	Advantages	Disadvantages
• Empiric formulas (Ly, Sartorius)	• Correlates well with equilibrium dialysis • Rapid and simple • More accurate than calculations based on law of mass action	• Lab specific • Requires large numbers of samples initially
Equilibrium Dialysis		
• Serum placed in dialysis chamber • Tracer-labeled T added to serum • Equilibrium achieved • Low molecular weight permeable membrane restricts passage of small molecules • Proportion of bound and free-labeled T assessed	• Gold-standard method for FT • Excellent sensitivity • Reproducible	• Technically challenging • Expensive • Time/labor intensive • Tracer impurities may compromise results • Relies on accuracy of TT assay
Free Androgen Index		
• Ratio of TT to SHBG	• Rapid and simple	• Poor correlation with FT at low levels • Relies on accuracy of TT and SHBG
Immunoassay		
• T analogue competes with FT for binding to antibody	• Rapid • Does not require extraction step • Inexpensive • Automated	• Inaccurate • Not recommended

EIA-Enzyme-linked immunosorbent assay; FT-Free testosterone; GC-MS-Gas chromatography mass spectrometry; LC-MS/MS-Liquid chromatography tandem mass spectrometry; MS-Mass spectrometry; RIA-Radioimmunoassay; SHBG-sex hormone binding globulin; T-testosterone; TT-total testosterone

Table 3

Sample Checklist for the Critical Analysis of Testosterone Literature

Category/Sample Questions
Authors
• What are potential biases of the authors? • What is the source of funding? • What conflicts of interest do the authors report?
Study Design
• Is the study prospective, retrospective, randomized, controlled? • What is the level of evidence? • Is the study design appropriate for the study objective? I.e. retrospective studies are unable to prove causation • Is the study duration sufficient to demonstrate an effect (including adverse effects)? • What are inclusion/exclusion criteria? • What are the primary and secondary endpoints for the study and/or study objective?
Population
• What are the population demographics: age, geography, BMI, comorbid conditions, chronic disease status, tobacco use? • Are concurrent medications a part of the inclusion/exclusion criteria? • What is the clinical setting for the population (i.e. randomized sample of college students, men attending a sexual dysfunction clinic, etc.)? • For comparative trials, are the populations well matched?
Objective Testing
• Is TT or fT used? • Which assay techniques/formulas for calculation are used? • When were samples obtained (AM vs PM) • Were multiple samples obtained? • What threshold is used for diagnosing low T? • Are the objective measures utilized appropriate to answer the study question? • Are questionnaires used as a screening modality? • How are adverse events reported?
Intervention

Category/Sample Questions
• Which therapy is used? • How are questionnaires utilized (if at all)? • Is a placebo utilized? • What is the treatment protocol? Is variable dosing permitted?
Results
• Are the presented results able to be shown by the study design (i.e. retrospective studies are unable to prove causation)? • Do the results represent direct or indirect evidence for the reported findings? • For interventional trials, are absolute changes in T reported? • Are responders and non-responders differentiated? • Are statistically significant changes clinically relevant? • Are adverse events reported?
Statistics
• Is the analysis sufficiently powered to address the study question? • Are groups appropriately matched? • Are the correct statistical tests employed? • Which type of error is the study at risk of reporting? • With the number of variables assessed, how many would you expect to be significant by chance alone? Are univariate and multivariate analyses utilized and reported appropriately and sufficiently powered?
• **Conclusions**
• Are the stated conclusions consistent with the data presented? • Are additional statements made beyond data presented?

fT – Free testosterone; T – Testosterone; TT – Total testosterone