Journal of Steroid Biochemistry & Molecular Biology 121 (2010) 513–519

Contents lists available at ScienceDirect

Journal of Steroid Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/jsbmb

Review

Standardization of testosterone measurements in humans☆,☆☆

Hubert W. Vesper*, Julianne Cook Botelho

Centers for Disease Control and Prevention, Division of Laboratory Sciences, 4770 Buford Highway, MS F25, Atlanta, GA 30341, United States

ARTICLE INFO

Article history:
Received 6 October 2009
Received in revised form 26 January 2010
Accepted 8 March 2010

Keywords:
Testosterone
Standardization
Traceability
CDC

ABSTRACT

Testosterone levels are used primarily for the diagnosis of hypogonadism in men and androgen excess in women. Current studies suggest that serum testosterone measurements may be indicated in a wide range of diseases and conditions. Translation of testosterone levels outside of the reference ranges into clinical treatment, appropriate cut offs for clinical guidelines and epidemiological studies with public health impact pose challenges due to the measurement variability among assays and in assay sensitivity. While introducing mass spectrometry technology can overcome some of these challenges and help to improve measurements, it faces variability issues similar to those observed with immunoassays that need to be addressed. To overcome these problems in testosterone testing, the Centers for Disease Control and Prevention, National Center for Environmental Health, Division of Laboratory Sciences (CDC/NCEH/DLS) started a steroid hormone standardization project. Their objective was to create testosterone measurement results that are traceable to one accuracy basis, thus allowing measurements to be comparable across methods, time, and location. CDC/NCEH/DLS conducts activities to standardize and improve testosterone assays and laboratory measurements by establishing metrological traceability to a higher order reference method and material. In addition, the standardization effort includes pre- and post-analytical challenges, such as test selection, interpretation, and establishing reference ranges to improve the translation of standardized results into clinical guidelines and public health assessments. CDC is conducting these standardization activities in collaboration with the clinical, laboratory, and research communities.

Published by Elsevier Ltd.

Contents

1. Introduction

In clinical settings, testosterone measurements are mainly used for the diagnosis of hypogonadism in men [1] and androgen excess in women [2] with polycystic ovary syndrome being one of the conditions causing androgen excess [3]. Research found that testosterone levels are associated with various diseases and conditions, such as metabolic syndrome [4], diabetes [5], cardiovascular disease [6,7], fractures [8,9], neurodegenerative disorder [10,11], and higher mortality in men with lower testosterone levels [12,13]. These findings stimulated further research as reflected in numerous studies under way (according to the National Institutes of Health [NIH] clinical trials database, more than 170 studies dealing with some aspect of testosterone are in the planning or recruitment phase) [14]. Some research findings were translated into information relevant to patient care and made available through clinical guidelines and recommendations [1,15–18]. These guidelines sug-

☆ Article from special issue on "Steroid profiling and analytics: going towards Sterome".

☆☆ The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention.

* Corresponding author. Tel.: +1 770 488 4191; fax: +1 404 638 5393.
E-mail address: HVesper@cdc.gov (H.W. Vesper).

0960-0760/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.jsbmb.2010.03.032

gest testing for testosterone to aid in diagnosing diseases and disorders or monitoring treatments. The test results would be compared with generally accepted reference ranges, clinical decision levels, or previously obtained measurements on the same patient when monitoring treatments. Thus, these guidelines presume that test results from a patient are comparable with others from clinical and epidemiological studies obtained at an earlier time, a different place, and with a different assay.

Due to reported problems in measurement variability and in sensitivity [19–28], researchers emphasized caution when choosing an assay. Researchers concerns about the quality of testosterone measurements have initiated recommendations to repeat measurements with more reliable assays on already completed epidemiological studies and clinical trials [29]. Others have recommended not using certain assays such as free testosterone analog assays [2,30–32]. However, despite these recommendations, tests are still used that are unsuitable for addressing clinical and public health questions [1].

For research findings to be useful and thus assure consistent and effective patient care and public health activities, test results need to be comparable. Comparability is needed also among tests performed in research studies. This would allow research findings to be verified and combined to facilitate the detection and investigation of potential public health problems or to be compared with tests performed in patients to aid in the diagnosis of diseases. Assuring comparability of measurement results over time and across different assays and laboratories is achieved through standardization efforts.

The need for standardization was formulated by researchers and professional organizations. In 2007, CDC started a steroid hormone standardization project to create testosterone measurement results that are traceable to one accuracy basis and thus are comparable across methods, time, and location. The aim of this paper is to describe the underlying rationale for standardizing testosterone measurements and CDC's plans to perform this effort.

1.1. Current variability in T measurements

Data from the College of American Pathologists (CAP) proficiency testing program, which surveys over 1000 clinical laboratories, showed profound variability between laboratories [32–36]. The magnitude of this variability, as expressed in the ratio between the highest and lowest reported value for a single sample, is greater than two and has not changed over the years (Table 1). High variability is observed between laboratories performing the same assay (within peer-group) as well as between laboratories performing different assays (among peer-groups) [32]. The high coefficients of variation (CVs) within and among peer-groups indicate problems in assay accuracy and precision. One limitation of the CAP survey is that the commutability of the materials used is unknown. Commutability is a characteristic that describes whether a material behaves like an authentic patient sample [37]. Frequently, materials used in external quality assessment schemes are pooled and otherwise modified to an extent that the material itself introduces a measurement bias. Therefore, it is unclear whether the observed variability among assays can be solely contributed to the assay performance or whether some variability is contributed to non-commutable materials.

In other studies, the performance of selected immunoassays using sera from individual donors and mass spectrometry-based (MS) assays as a comparative method was investigated [30,35,38–40]. For most assays, the investigators observed high correlations between the MS assay and the immunoassays. However, all studies showed substantial differences in absolute values. These studies were performed at different times and in different countries indicating that the problem of assay performance is ongoing and international in its scope. In the study by Wang et al. [35], the average percent differences between the immunoassays and the MS assay ranged between −18% and +15.9%; for concentrations of less than 100 ng/dL (3.47 nmol/L), they ranged between −40% and +40%. Thus, the reported variability among assays is similar to the variability found with the CAP surveys, especially at concentrations commonly observed in women and hypogonadal men. Some researchers concluded that immunoassays can distinguish eugonadal from hypogonadal males using laboratory specific reference ranges. However, due to the lack of precision and accuracy, they do not seem suitable for measuring testosterone in females or prepubertal subjects [35,41].

Because MS assays can provide highly specific and sensitive measurements, they are frequently called the "gold standard"

Table 1
Highest and lowest reported total testosterone value from individual samples reported from several surveys from the College of American Pathologists.

Reference	Year	Sample	Total testosterone ng/dL (nmol/L)		Ratio high/low
			Low	High	
Steinberger et al. [20]	1991	Y-01	300 [10.41]	588 [20.4]	2.0
		Y-02	59 [2.05]	164 [5.69]	2.8
		Y-05	9 [0.31]	54 [1.87]	6.0
		Y-06	228 [7.91]	522 [18.1]	2.3
		Y-09	508 [17.6]	1148 [39.8]	2.3
		Y-10	27 [0.94]	91 [3.16]	3.4
		Y-13	2 [0.07]	66 [2.29]	33.0
		Y-14	226 [7.84]	600 [20.8]	2.7
		Y-94	38 [1.32]	141 [4.89]	3.7
		Y-95	1001 [34.7]	1707 [59.2]	1.7
Steinberger et al. [20]	1995	Y-01	0 [0]	78 [2.71]	78.0
		Y-02	538 [18.7]	1140 [39.6]	2.1
		Y-05	597 [20.7]	1144 [5.00]	1.9
		Y-06	276 [9.58]	600 [20.8]	2.2
		Y-09	72 [2.50]	264 [9.16]	3.7
		Y-10	33 [1.15]	170 [5.90]	5.2
Wang et al. [35]	2004	Y-04	160 [5.55]	580 [20.1]	3.6
Rosner et al. [32]	2007	N/A[a]	7 [0.25]	100 [3.47]	14.3
		N/A	45 [1.56]	365 [12.7]	8.1
		N/A	276 [9.58]	744 [25.8]	2.7

[a] Information not available.

Table 2
Within-run and between-run imprecision determined with three serum pools using the same high-performance liquid chromatography tandem mass spectrometry conditions and three different sample preparation procedures.

	Within-run CV[a] (%) N = 3			Between-run CV (%) N = 5		
	Pool 1 (14.5 ng/dL)	Pool 2 (104 ng/dL)	Pool 3 (973 ng/dL)	Pool 1 (14.5 ng/dL)	Pool 2 (104 ng/dL)	Pool 3 (973 ng/dL)
	[0.5 nmol/L]	[3.61 nmol/L]	[33.8 nmol/L]	[0.5 nmol/L]	[3.61 nmol/L]	[33.8 nmol/L]
Method A[b]	5.0	3.5	5.2	4.1	3.3	8.9
Method B[c]	6.8	3.8	5.0	7.5	11.5	8.4
Method C[d]	2.7	1.8	2.4	3.7	1.9	5.0

[a] Coefficient of variation.
[b] An on-line solid phase extraction method with protein precipitation.
[c] An on-line solid phase extraction method.
[d] An offline solid phase extraction method with liquid–liquid extraction.

for measuring testosterone at low concentrations such as those observed in women and children [2,31,42]. In contrast to commercial immunoassays, MS assays are developed and validated by the laboratory, and thus are commonly referred to as "in-house" assays. Though they use the same detection principle (MS), they frequently differ in sample handling and preparation and in calibration. Two studies investigated the performance of MS assays against a MS reference method [43,44]. One study investigating 7 MS assays reported mean biases ranging between −14.1% and 19.2%; at concentrations less than 100 ng/dL (3.47 nmol/L), the biases were as high as 25.3%. The coefficient of variations measured on two samples with total testosterone values of 296 ng/dL (10.3 nmol/L) and 8.47 ng/dL (0.29 nmol/L) ranged between 2.19% and 11.36% and 2.67% and 25.58%, respectively [44]. The other study investigated 4 MS assays found that between 7% and 26% of the results reported by these assays were outside a 14% total error limit [43]. The precision at concentrations commonly observed in female samples particularly was smaller for most MS assays than those seen with immunoassays. Further, the MS assays differed in their detection limit and reportable range. The results from these two studies demonstrate that MS assays are a heterogeneous group of assays with the same measurement principles but with differences in assay performance.

While the differences in accuracy among MS methods seem attributable to differences in calibration, the differences in assay precision could be explained in part with variations in sample preparation. To test this hypothesis, we investigated the effect of sample preparation on assay precision [45]. In this study, we used three different serum sample preparation procedures (protein precipitation using acetonitrile followed by automated on-line solid phase extraction on a C18 reversed phase column [Method A], automated on-line solid phase extraction on a C18 reverse-phase column only [Method B] and off line solid phase extraction on a C18 reversed phase column followed by liquid–liquid extraction with hexane [Method C]) to analyze quality control pools and 53 male and female patient samples with the same HPLC-isotope dilution mass spectrometry method. These different procedures gave different imprecision (Table 2), which resulted in different population data distributions (Fig. 1) but not statistically significant differences in assay accuracy. The reasons for these differences in assay imprecision are not fully understood and could be related to ion suppression effects as suggested by other researchers [46]. Ion suppression results from the presence of less volatile compounds in the sample that can change the efficiency of droplet formation or droplet evaporation in the ion source of the MS, which in turn affects the amount of charged ion in the gas phase that ultimately reaches the detector. Compounds shown to cause ion suppression are salts, ion-pairing agents, drugs, metabolites, and proteins [47]. It needs to be noted that small changes to the sample preparation procedures tested in this study may lead to different outcomes. Therefore, none of the described sample preparation principles can be considered superior. Nonetheless, this study points out that special attention should be given to the sample preparation portion of the MS assays to assure appropriate performance.

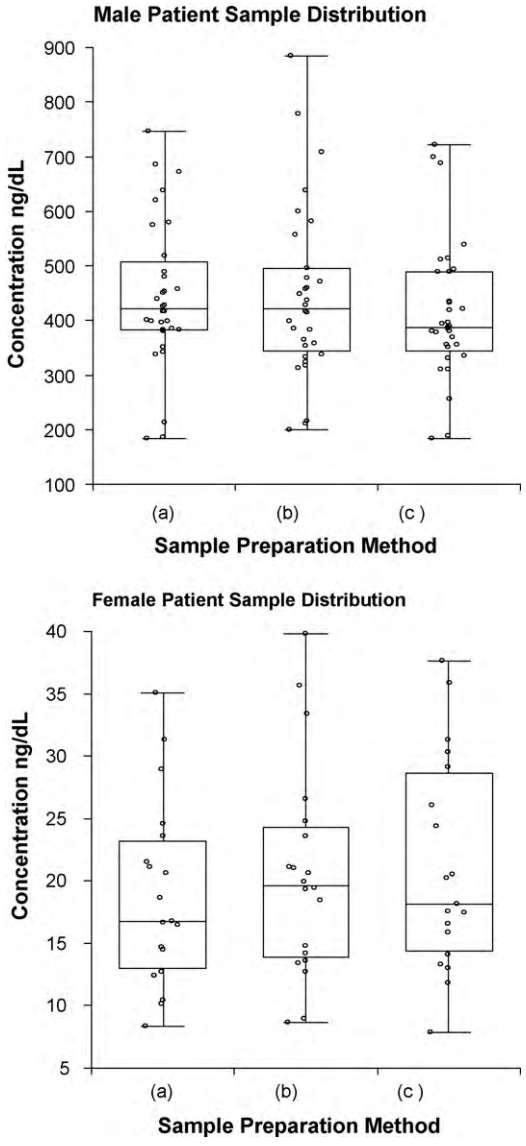

Fig. 1. Box–Whisker plot of total testosterone values obtained with 53 male and female patient samples using three different sample preparation procedures with the same high-performance liquid chromatography/mass spectrometry/mass spectrometry. (HPLC/MS/MS) method.

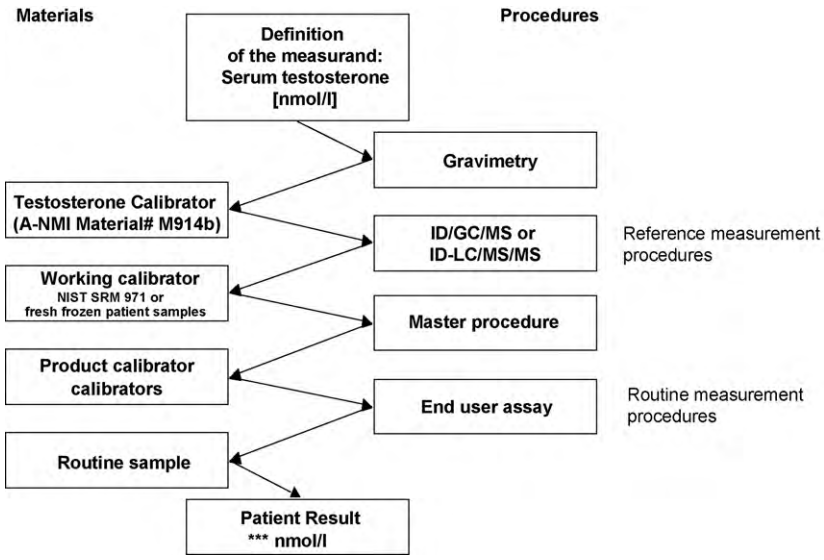

Fig. 2. Metrological traceability chain for testosterone (arrows to the right: calibration activities; arrows to the left: value assignment activities). Activities 1 and 2 are performed by the reference laboratory at the CDC, steps 3–6 are performed by the assay manufacturer or, for in-house assays, by the laboratory. A-NMI: Australian National Metrological Institute. NIST: U.S. National Institute of Standards and Technology. SRM: Standard Reference Material. ID/GC/MS: Isotope dilution gas chromatography mass spectrometry. ID-LC/MS/MS: Isotope dilution high-performance liquid chromatography tandem mass spectrometry.

1.2. Testosterone standardization

In March 2008, CDC conducted a workshop to discuss problems and needs concerning steroid hormone testing with the clinical, research, and diagnostic community [48]. The needs can be summarized as:

- comparability of data across studies and measurement systems,
- appropriate performance of the assays, especially at low concentrations of hormones such as those observed among women and children for testosterone,
- generally accepted reference ranges for people of all ages and ethnicities and both sexes,
- consensus on the use of testing for different forms of steroids such as free, bioavailable, and total testosterone, and
- greater awareness within the clinical and research community about the problems and limitations of different steroid assays.

Based on these findings, CDC's effort to standardize testosterone was structured into a reference laboratory component and a translational component. The first component includes activities to standardize and improve the testosterone assays and laboratory measurements, thus addressing the analytical component of the measurement process. The second component comprises activities that address issues relevant to selecting a test and interpreting results, which is the pre- and post-analytical component in the measurement process.

1.3. Reference laboratory component

The activities related to this component aim for assuring calibration of assays and laboratories to a common reference basis, and for assuring consistency of this calibration across assays and laboratories over time. Thus, the aim is to assure the measurement results from a sample are the same, independent of the methodology or technology used.

Assuring calibration to a common standard can be achieved through establishing metrological traceability as described in International Organization for Standardization (ISO) document 17511 [49] and further explained in a Clinical and Laboratory Standards Institute (CLSI) document [50] and a recent review [51]. Establishing traceability is an alternating process of assigning values to a material that is then used to calibrate a method (Fig. 2). The process starts with gravimetrically preparing a pure compound standard (commonly called "primary standard," such as the certified reference material M914b for testosterone from the Australian National Metrology Institute) that is then used to calibrate a reference method. Because most clinical immunoassays require matrix-based calibrators, the reference method is then used to assign values to matrix-based materials (i.e., fresh-frozen patient samples). These matrix-based materials are then used by the assay manufacturers to calibrate their immunoassays, which then measure testosterone in a patient sample (as indicated in Fig. 2, assay manufacturers may insert additional steps by calibrating an "in-house reference method" (master assay) first that is used to assign values to the manufacturer's specific calibrators, which are then used to calibrate the assays that measure patient samples).

Pure compound reference materials exist. However, there is a need for commutable, matrix-based materials that can be used to calibrate clinical and research assays. To assure commutability, sets of single-donor (non-pooled) sera are used for calibrating clinical assays. Because the volume of single-donor serum is limited, many sera are needed over time. This will require frequent value assignments by reference laboratories. More reference methods and laboratories are needed to provide sufficient matrix-based calibrators that help assay manufacturers and laboratories in their calibration. To satisfy this need, CDC is developing a reference method for testosterone in serum and is using this method to assign total testosterone values to sera that are available to laboratories and assay manufacturers in its laboratory standardization program. Further, CDC is collaborating with other organizations such as CAP that will offer products for assessing assay calibration [52].

Reference methods are intended to provide highly accurate and precise measurements (low measurement uncertainty). Because the measurement uncertainty increases with each step in the traceability chain, the methods used at the top of this chain should have a low measurement uncertainty to assure that measurement results at the bottom of the chain are still small enough to be meaningful for patient care and public health. To achieve the high accuracy and precision of reference methods, these methods are designed and operated differently than the regular routine methods as outlined in Table 3.

Table 3
Typical characteristics of reference and routine measurement procedures.

Routine method	Reference method
Ensure adequate accuracy and precision for purpose of method (i.e., screening, diagnosis)	Optimized for trueness and precision to minimize uncertainty
Same calibration curve (widest range possible) and internal standard concentration for all samples	Adjust calibration and IS concentration for each sample to best match analyte concentration (bracketing) and use weight instead of volume
Balance sample preparation and chromatographic separation with throughput	Optimize chromatographic separation for compound of interest for full resolution
Typically single measurement	Typically replicate measurements
Frequently multi-analyte methods	Typically single-analyte methods
Generally high throughput	Generally low throughput

Calibration with single-donor sera is currently the best approach for calibrating clinical and research assays because the calibration material is as similar to a patient sample as possible. However, it is also desirable to have large volumes of materials (pooled sera) that can be used as calibrators or in accuracy-based, external quality of assessment (EQA) schemes. Pooled and otherwise modified sera that are intended as a calibrator, trueness control, or in accuracy-based EQA schemes should be commutable to fit the intended use. CDC is working with material manufacturers, such as the National Institute of Standards and Technology and EQA providers to assess the commutability of such pooled materials [53].

Establishing metrological traceability focuses on an individual assay or laboratory and involves procedures performed by reference laboratories, assay manufactures, and end-user laboratories as described above. Inter-laboratory comparison studies or monitoring of end-user performance through EQA schemes should be conducted to verify that these procedures are performed in a manner that produces consistent results across assays, laboratories, and time. CDC is performing studies with assay manufacturers or with laboratories in the case of "in-house assays," is collaborating with EQA providers on monitoring end-user performance and on supporting accuracy-based surveys to better assess consistency of assay calibration.

The described approach for standardizing testosterone assays is based on other successful approaches such as the CDC standardization program for cholesterol and blood lipids [54] and in the National Glycohemoglobin Standardization Program (NGSP) [55]. Similar approaches are currently under way for other analytes such as serum creatinine [56]. Glycated hemoglobin A1c (HbA1c) measurements showed in 1993 similar measurement variability as is seen today for total testosterone measurements. The efforts of the NGSP achieved a profound reduction in assay imprecision and bias in 2002 with most certified methods having between-laboratory CVs of less than 5% and all certified methods having mean HbA1c values within 0.8% HbA1c from the NGSP target [55,57]. This report shows that improvements in measurement performance can be achieved through assay standardization. It also indicates that it takes many years to reach such achievements and continuous efforts by all parties involved in laboratory testing to maintain them.

1.4. Translational activities

According to the CDC workshop on steroid hormone testing, improvements are needed on both the analytical and pre- and post-analytical portions. Understanding the strengths and limitations of current testosterone tests, such as assay sensitivity and precision, is important when choosing an assay to answer specific clinical or research questions. This requires knowledge about the biological variability and factors affecting the assay to select tests that are precise and accurate enough to distinguish between a normal biological variation and a true physiological change. Ricos et al. have suggested assay performance criteria that are based on biological variability [58], and databases with method performance criteria have been generated [59]. These criteria suggest a desirable imprecision of 4.7%, bias of 6.4%, and total error of 14% for total testosterone in serum. These criteria were derived from data obtained mainly from men using non-standardized immunoassays. Further studies are needed to verify these data and the derived performance criteria.

To be able to distinguish normal testosterone values from impaired testosterone levels in a subject or population, generally accepted reference ranges are needed. High variability in reference ranges has been described [20,32,60], with limits as low as 84 ng/dL (2.91 nmol/L) and as high as 1727 ng/dL (59.9 nmol/L) reported for men, and as low as 2 and as high as 95 ng/dL (3.30 nmol/L) reported for women. The high variability in these reference ranges can be contributed to the variability in measurements, but also seems to be caused by differences in the characteristics of the population used to determine the reference range [1]. Frequently, only limited information about the subjects used to establish reference ranges is available. Reference ranges should be derived from well-characterized, adequate-sized populations using standardized procedures such as those formulated by CLSI [61].

2. Summary/conclusions

Testosterone is an analyte commonly used for many years in research and patient care. Problems in comparability of measurement results and assay performance, especially at low serum concentrations, prevent using new research findings in patient care and public health and impede its effectiveness in current clinical applications. These problems can be addressed through standardization efforts that focus not only on the analytical measurement process but also on pre- and post-analytical issues such as test selection and reference ranges. CDC is performing such a standardization effort for testosterone in collaboration with the clinical, laboratory, and research community.

Acknowledgements

Solvay Pharmaceuticals provided funding for this project through the CDC Foundation. The Division of Laboratory Sciences at the National Center for Environmental Health and the Division of Cancer Prevention and Control at the National Center for Chronic Disease Prevention and Health Promotion also contributed to this project. Professional societies and organizations, such as the Endocrine Society, the American Association of Clinical Endocrinologist, and the American Association of Clinical Chemistry have provided additional support. The authors also acknowledge the involvement of past and present members of the hormone standardization research team at CDC: Brittany Butler, Gabrielle Gay, Raj Razdan, Christopher Shacklady, Antoinette R. Smith, and Jamie White.

References

[1] S. Bhasin, G.R. Cunningham, F.J. Hayes, A.M. Matsumoto, P.J. Snyder, R.S. Swerdloff, V.M. Montori, Testosterone therapy in adult men with androgen deficiency syndromes: an endocrine society clinical practice guideline, J. Clin. Endocrinol. Metab. 91 (2006) 1995–2010.

[2] M.E. Wierman, R. Basson, S.R. Davis, S. Khosla, K.K. Miller, W. Rosner, N. Santoro, Androgen therapy in women: an Endocrine Society Clinical Practice guideline, J. Clin. Endocrinol. Metab. 91 (2006) 3697–3710.

[3] V. Nisenblat, R.J. Norman, Androgens and polycystic ovary syndrome, Curr. Opin. Endocrinol. Diabetes Obes. 16 (2009) 224–231.

[4] F. Saad, L. Gooren, The role of testosterone in the metabolic syndrome: a review, J. Steroid Biochem. Mol. Biol. 114 (2009) 40–43.

[5] E. Selvin, M. Feinleib, L. Zhang, S. Rohrmann, N. Rifai, W.G. Nelson, A. Dobs, S. Basaria, S.H. Golden, E.A. Platz, Androgens and diabetes in men: results from the Third National Health and Nutrition Examination Survey (NHANES III), Diabetes Care 30 (2007) 234–238.

[6] P. Hakimian, M. Blute Jr., J. Kashanian, S. Chan, D. Silver, R. Shabsigh, Metabolic and cardiovascular effects of androgen deprivation therapy, BJU Int. 102 (2008) 1509–1514.

[7] G.M. Rosano, A. Cornoldi, M. Fini, Effects of androgens on the cardiovascular system, J. Endocrinol. Invest. 28 (2005) 32–38.

[8] S.P. Tuck, R.M. Francis, Testosterone, bone and osteoporosis, Front. Horm. Res. 37 (2009) 123–132.

[9] M.I. Hu, R.F. Gagel, C. Jimenez, Bone loss in patients with breast or prostate cancer, Curr. Osteoporos. Rep. 5 (2007) 170–178.

[10] C.J. Pike, J.C. Carroll, E.R. Rosario, A.M. Barron, Protective actions of sex steroid hormones in Alzheimer's disease, Front. Neuroendocrinol. 30 (2009) 239–258.

[11] M.M. Cherrier, Testosterone effects on cognition in health and disease, Front. Horm. Res. 37 (2009) 150–162.

[12] G.A. Laughlin, E. Barrett-Connor, J. Bergstrom, Low serum testosterone and mortality in older men, J. Clin. Endocrinol. Metab. 93 (2008) 68–75.

[13] M.M. Shores, A.M. Matsumoto, K.L. Sloan, D.R. Kivlahan, Low serum testosterone and mortality in male veterans, Arch. Intern. Med. 166 (2006) 1660–1665.

[14] National Institute of Health. www.clinicaltrials.gov, 2009 (accessed 09/15/2009).

[15] American Association of Clinical Endocrinologists Polycystic Ovary Syndrome Writing Committee, American Association of Clinical Endocrinologists position statement on metabolic and cardiovascular consequences of polycystic ovary syndrome, Endocr. Pract. 11 (2005) 126–134.

[16] N.F. Goodman, M.B. Bledsoe, R.H. Cobin, W. Futterweit, J.W. Goldzieher, S.M. Petak, K.D. Smith, E. Steinberger, American Association of Clinical Endocrinologists Hyperandrogenic Disorders Task Force, American Association of Clinical Endocrinologists Medical Guidelines for the clinical practice for the diagnosis and treatment of hyperandrogenic disorders, Endocr. Pract. 7 (2001) 120–134.

[17] S.M. Petak, H.R. Nankin, R.F. Spark, R.S. Swerdloff, L.J. Rodriguez-Rigau, American Association of Clinical Endocrinologists, American Association of Clinical Endocrinologists Medical Guidelines for clinical practice for the evaluation and treatment of hypogonadism in adult male patients—2002 update, Endocr. Pract. 8 (2002) 440–456.

[18] C. Wang, E. Neischlag, R. Swerfloff, H.M. Behre, W.J. Hellstrom, L.J. Gooren, J.M. Kaufman, J.-J. Legros, B. Lunenfeld, A. Morales, J.E. Morley, C. Shulman, I.M. Thompson, W. Weidner, F.C.W. Wu, Review: ISA, ISSAM, EAU, EAA, and ASA recommendations: investigation, treatment and monitoring of late-onset hypogonadism in males, Int. J. Impot. Res. 21 (2009) 1–8.

[19] S.E. Hankinson, J.E. Manson, S.J. London, W.C. Willett, F.E. Speizer, Laboratory reproducibility of endogenous hormone levels in postmenopausal women, Cancer Epidemiol. Biomarkers Prev. 3 (1994) 51–56.

[20] E. Steinberger, C. Ayala, B. Hsi, K.D. Smith, L.J. Rodriguez-Rigau, E.R. Weidman, G.G. Reimondo, Utilization of commercial laboratory results in management of hyperandrogenism in women, Endocr. Pract. 4 (1998) 1–10.

[21] F.Z. Stanczyk, J.S. Lee, R.J. Santen, Standardization of steroid hormone assays: why, how, and when? Cancer Epidemiol. Biomarkers Prev. 16 (2007) 1713–1719.

[22] S.S. Sacks, Are routine testosterone assays good enough? Clin. Biochem. Rev. 26 (2005) 43–45.

[23] A.M. Matsumoto, W.J. Bremner, Serum testosterone assays—accuracy matters, J. Clin. Endocrinol. Metab. 89 (2004) 520–524.

[24] D.A. Herold, R.L. Fitzgerald, Immunoassays for testosterone in women: better than a guess? Clin. Chem. 49 (2003) 1250–1251.

[25] R.S. Swerdloff, C. Wang, Free testosterone measurements by analog displacement direct assay: old concerns and new evidence, Clin. Chem. 54 (2008) 458–459.

[26] S. Rinaldi, H. Déchaud, C. Biessy, V. Morin-Raverot, P. Toniolo, A. Zeleniuch-Jacquotte, A. Akhmedkhanov, R.E. Shore, G. Secreto, A. Ciampi, E. Riboli, R. Kaaks, Reliability and validity of commercially available, direct radioimmunoassays for measurement of blood androgens and estrogens in postmenopausal women, Cancer Epidemiol. Biomarkers Prev. 10 (2001) 757–765.

[27] K.S. Fritz, A.J. McKean, J.C. Nelson, R.B. Wilcox, Analog-based free testosterone test results linked to total testosterone concentrations, not free testosterone concentrations, Clin. Chem. 54 (2008) 512–516.

[28] K. Van Uytfanghe, D. Stöckl, J.M. Kaufman, T. Fiers, A. De Leenheer, L.M. Thienpont, Validation of 5 routine assays for serum free testosterone with a candidate reference measurement procedure based on ultrafiltration and isotope dilution-gas chromatography–mass spectrometry, Clin. Biochem. 38 (2005) 253–261.

[29] J.H. Barth, E. Yasmin, A.H. Balen, The diagnosis of polycystic ovary syndrome: the criteria are insufficiently robust for clinical rese, Clin. Endocrinol. (Oxf) 67 (2007) 811–815.

[30] J. Kane, J. Middle, M. Cawood, Measurement of serum testosterone in women; what should we do? Ann. Clin. Biochem. 44 (2007) 5–15.

[31] F.Z. Stanczyk, Measurement of androgens in women, Semin. Reprod. Med. 24 (2006) 78–85.

[32] W. Rosner, R.J. Auchus, R. Azziz, P.M. Sluss, H. Raff, Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement, J. Clin. Endocrinol. Metab. 92 (2007) 405–413.

[33] S.J. Soldin, O.P. Soldin, Steroid hormone analysis by tandem mass spectrometry, Clin. Chem. 55 (2009) 1061–1066.

[34] T. Guo, M. Chan, S.J. Soldin, Steroid profiles using liquid chromatography–tandem mass spectrometry with atmospheric pressure photoionization source, Arch. Pathol. Lab. Med. 128 (2004) 469–475.

[35] C. Wang, D.H. Catlin, L.M. Demers, B. Starcevic, R.S. Swerdloff, Measurement of total serum testosterone in adult men: comparison of current laboratory methods versus liquid chromatography–tandem mass spectrometry, J. Clin. Endocrinol. Metab. 89 (2004) 534–543.

[36] R.J. Elin, S.J. Winters, Current controversies in testosterone testing: aging and obesity, Clin. Lab. Med. 24 (2004) 119–139.

[37] H.W. Vesper, W.G. Miller, G.L. Myers, Reference materials and commutability, Clin. Biochem. Rev. 28 (2007) 139–147.

[38] J. Taieb, B. Mathian, F. Millot, M.-C. Patricot, E. Mathieu, N. Queyrel, I. Lacroix, C. Somma-Delpero, P. Boudou, Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography–mass spectrometry in sera from 116 men, women, and children, Clin. Chem. 49 (2003) 1381–1395.

[39] A.W. Hsing, F.Z. Stanczyk, A. Bélanger, P. Schroeder, L. Chang, R.T. Falk, T.R. Fears, Reproducibility of serum sex steroid assays in men by RIA and mass spectrometry, Cancer Epidemiol. Biomarkers Prev. 16 (2007) 1004–1008.

[40] K. Sikaris, R.I. McLachlan, R. Kazlauskas, D. de Krester, C.A. Holden, D.J. Handelsman, Reproductive hormone reference intervals for healthy fertile young men: evaluation of automated platform assays, J. Clin. Endocrinol. Metab. 90 (2005) 5928–5936.

[41] V. Moal, E. Mathieu, P. Reynier, Y. Malthièry, Y. Gallois, Low serum testosterone assayed by liquid chromatography–tandem mass spectrometry: comparison with five immunoassay techniques, Clin. Chim. Acta 386 (2007) 12–19.

[42] L. Albrecht, D. Styne, Laboratory testing of gonadal steroids in children, Ped. Endocrinol. Rev. 5 (Suppl. 1) (2007) 599–607.

[43] L.M. Thienpont, K. Van Uytfanghe, S. Blincko, C.S. Ramsay, H. Xie, R.C. Doss, B.G. Keevil, L.J. Owen, A.L. Rockwood, M.M. Kushnir, K.Y. Chun, D.W. Chandler, H.P. Field, P.M. Sluss, State-of-the-art of serum testosterone measurement by isotope dilution-liquid chromatography–tandem mass spectrometry, Clin. Chem. 54 (2008) 1290–1297.

[44] H.W. Vesper, S. Bhasin, C. Wang, S.S. Tai, L.A. Dodge, R.J. Singh, J. Neslon, S. Ohorodnik, N.J. Clarke, C.R. Parker Jr., R. Razdan, E.A. Monsell, G.L. Myers, Interlaboratory comparison study of serum total testosterone measurements performed by mass spectrometry methods, Steroids 74 (2009) 498–503.

[45] J.C. Botelho, C. Shacklady, R. Razdan, H.W. Vesper, Impact of sample preparation procedures of testosterone measurements by isotope-dilution liquid chromatography–tandem mass spectrometry (abstract), Clin. Chem. 54 (2008) A122.

[46] M.M. Kushnir, A.L. Rockwood, W.L. Roberts, E.G. Pattison, A.M. Bunker, R.L. Fitzgerald, A.W. Meikle, Performance characteristics of a novel tandem mass spectrometry assay for serum testosterone, Clin. Chem. 52 (2006) 120–128.

[47] T.M. Annesley, Ion suppression in mass spectrometry, Clin. Chem. 49 (2003) 1041–1044.

[48] W. Rosner, H. Vesper, CDC workshop report improving steroid hormone measurements in patient care and research translation, Steroids 73 (2008) 1285.

[49] ISO, In Vitro Diagnostic Medical Devices—Measurement of Quantities in Biological Samples—Metrological Traceability of Values Assigned to Calibrators and Control Materials. ISO 17511:2003, International Organization for Standardization, Geneva, 2003.

[50] CLSI, Metrological Traceability and its Implementation: A Report. CLSI Document X5-R, NCCLS, Wayne (PA), 2006.

[51] H.W. Vesper, L.M. Thienpont, Traceability in laboratory medicine, Clin. Chem. 55 (2009) 1067–1075.

[52] College of American Pathologist. www.cap.org, 2010 (accessed 01/20/10).

[53] J.C. Botelho, C. Shacklady, R. Razdan, A. Smith, H.W. Vesper, Commutability of serum-based reference and proficiency testing materials for total testosterone with selected immunoassays (abstract), Clin. Chem. 55 (Suppl. 6) (2009) A173.

[54] Lipid Standardization Program. www.cdc.gov/labstandards, 2010 (accessed 01/20/10).

[55] National Glycohemoglobin Standardization Program. www.ngsp.org, 2010 (accessed 01/20/10).

[56] National Kidney Disease Education Program: Creatinine Standardization. www.nkdep.nih.gov/labprofessionals/creatinine_standardization, 2010 (accessed 01/20/10).

[57] R.R. Little, Glycated hemoglobin standardization—National Glycohemoglobin Standardization Program (NGSP) perspective, Clin. Chem. Lab. Med. 41 (2003) 1191–1198.

[58] C. Ricos, V. Alvarez, F. Cava, J.V. Garcia-Lario, A. Hernandez, C.V. Jimenez, J. Minchinela, C. Perich, M. Simon, Current databases on biologic variation: pros, cons and progress, Scand. J. Clin. Lab. Invest. 59 (1999) 491–500.
[59] J. Westgard. http://www.westgard.com/biodatabase1.htm, 2009 (accessed 09/15/2009).
[60] S. Bhasin, A. Zhang, A. Coviell, R. Jasuja, J. Ulloor, R. Singh, H. Vesper, R.S. Vasan, The impact of assay quality and reference ranges on clinical decision making in the diagnosis of androgen disorders, Steroids 73 (2008) 1311–1317.
[61] CLSI, Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory, CLSI Document C28, CLSI, Wayne (PA), 2008.