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

The Journal of Clinical Endocrinology & Metabolism 89(2):534–543

doi: 10.1210/jc.2003-031287

Measurement of Total Serum Testosterone in Adult Men: Comparison of Current Laboratory Methods *Versus* Liquid Chromatography-Tandem Mass Spectrometry

CHRISTINA WANG, DON H. CATLIN, LAURENCE M. DEMERS, BORISLAV STARCEVIC, AND RONALD S. SWERDLOFF

Division of Endocrinology (C.W., R.S.S.), Department of Medicine, Harbor-UCLA Medical Center and Research and Education Institute, Torrance, California 90502; UCLA-Olympic Analytical Laboratory (D.H.C., B.S.), Los Angeles, California 90025; and Department of Pathology and Medicine (L.M.D.), Pennsylvania State University College of Medicine, H. S. Hershey Medical Center, Hershey, Pennsylvania 17033

The diagnosis of male hypogonadism requires the demonstration of a low serum testosterone (T) level. We examined serum T levels in pedigreed samples taken from 62 eugonadal and 60 hypogonadal males by four commonly used automated immunoassay instruments (Roche Elecsys, Bayer Centaur, Ortho Vitros ECi and DPC Immulite 2000) and two manual immunoassay methods (DPC-RIA, a coated tube commercial kit, and HUMC-RIA, a research laboratory assay) and compared results with measurements performed by liquid chromatography-tandem mass spectrometry (LC-MSMS). Deming's regression analyses comparing each of the test results with LC-MSMS showed slopes that were between 0.881 and 1.217. The interclass correlation coefficients were between 0.92 and 0.97 for all methods. Compared with the serum T concentrations measured by LC-MSMS, the DPC Immulite results were biased toward lower values (mean difference, −90 ± 9 ng/dl) whereas the Bayer Centaur data were biased toward higher values (mean difference, +99 ± 11 ng/dl) over a wide range of serum T levels. At low serum T concentrations (<100 ng/dl or 3.47 nmol/liter), HUMC-RIA overestimated serum T, Ortho Vitros ECi underestimated the serum T concentration, whereas the other two methods (DPC-RIA and Roche Elecsys) showed differences in both directions compared with LC-MSMS. Over 60% of the samples (with T levels within the adult male range) measured by most automated and manual methods were within ± 20% of those reported by LC-MSMS. These immunoassays are capable of distinguishing eugonadal from hypogonadal males if adult male reference ranges have been established in each individual laboratory. The lack of precision and accuracy, together with bias of the immunoassay methods at low serum T concentrations, suggests that the current methods cannot be used to accurately measure T in females or serum from prepubertal subjects. (*J Clin Endocrinol Metab* 89: 534–543, 2004)

THE DIAGNOSIS OF androgen deficiency in men is usually based on clinical features of hypogonadism and the demonstration of a morning serum total testosterone (T) level below the reference range for young male adults. In the past 30 yr, serum T levels have been measured in both research and clinical laboratories using established RIAs that initially employed an extraction and column chromatography purification step before performing the RIA (1–4). Subsequently with the availability of more specific antibodies, the chromatography step and then the extraction step were eliminated in most laboratories. Ready-made commercial kits for RIAs were then introduced and routinely used in most clinical and research laboratories.

More recently, assays for serum T in male and female serum have been performed in many hospital and reference laboratories using rapid automated immunoassay instruments that employ chemiluminescence detection. These assays are performed with proprietary reagents that include analogs of T as standards and reference ranges provided by the instrument manufacturer. While economical and rapid, many of these assays have had limited published validation data, raising questions about the accuracy and/or specificity of these automated immunoassay methods. Furthermore, the approval of these methods by regulatory agencies for clinical use is primarily based on noninferiority comparison against previously approved assays frequently using pooled samples and mostly not from T-free serum spiked with gravimetrically determined standards of authentic T or from individual serum samples independently assayed by other methods such as mass spectrometry methods. A major problem exists when the standard reference texts for physicians (5) describe an adult male reference range that does not correspond to values quoted by many clinical laboratories. Clinicians are being presented with normal male reference ranges for serum T from these automated platforms that have low end clinical limits down to 170–200 ng/dl (5.9–6.9 nmol/liter) and upper range limits of 700–800 ng/dl (24.3–27.7 nmol/liter). These stated reference ranges provided by the manufacturer are significantly lower than the 300-1000 ng/dl (10.4–34.7 nmol/liter) reference range referred to in numerous publications over the past 30 yr based on traditional RIA methods with or without the chromatography step as well as some research techniques employed by internal recovery standards to correct for procedural losses (5).

Abbreviations: CV, Coefficient of variation; GC, gas chromatograph; HRP, horseradish peroxidase; HUMC, Harbor-UCLA Research and Education Institute Endocrine Research Laboratory; LC-MSMS, liquid chromatography-tandem mass spectrometry; LOQ, limit of quantification; MS, mass spectrometry; T, testosterone.

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

External quality control programs such as that provided by the College of American Pathologists allow laboratories to compare results with other laboratories using the same method or kit reagents. As shown in Table 1, the median value of a quality control sample (Y-04, 2002) varied between 215 and 348 ng/dl (7.5 and 12.0 nmol/liter) among methods with coefficients of variation among laboratories using the same method or instrument ranging between 5.1% and 22.7%. The median average for this sample from all methods was 297 ng/dl (10.3 nmol/liter) and results were as low as 160 or as high as 508 ng/dl (5.5 to 17.6 nmol/liter). These results span the hypogonadal to eugonadal range.

A previous study evaluated and compared steroid measurements by RIA and gas chromatography-mass spectrometry using pooled female and male serum samples. They used linear regression analysis and demonstrated that similar results could be obtained for most steroids in serum either by RIA or mass spectrometry (6). This report, however, only tested pooled samples that covered the high, medium, and low range of each steroid standard curve and not pedigreed samples from normal subjects and patients. Moreover the use of least-squares linear regression analysis is not an optimal measure because it does not take into consideration the fact that both the reference and the test methods contain error. In this study, we compared serum T measurements from eugonadal and hypogonadal adult men using liquid chromatography-tandem mass spectrometry (LC-MSMS) (UCLA Olympic Analytical Laboratory) *vs.* two RIAs run in a research laboratory (Harbor-UCLA Research and Education Institute Endocrine Research Laboratory, HUMC-RIA) and a hospital based reference laboratory using a commercially available RIA kit (DPC-RIA, Core Endocrine Laboratory, Penn State University-Hershey Medical Center, Hershey, PA), and compared results with the same specimens run on the most common automated immunoassay instruments used in hospital based laboratories (Penn State University-Hershey Medical Center; University of Pennsylvania, Philadelphia, PA; Mercy Health Laboratories, Philadelphia, PA; and Henry Ford Hospital, Detroit, MI).

Subjects and Methods

Subjects

Serum samples were collected from normal (n = 62) and hypogonadal men (n = 60) from June 1995 to September 1999. The 62 normal healthy volunteers were 18–60 yr of age. Serum was collected between 0800 and 1000 h from healthy volunteers in the basal state without any research protocol interventions. These subjects were recruited at Harbor-UCLA Center of Men's Health for other research studies on androgen metabolism. They had no significant medical history and were not taking medications. They had a normal physical examination, normal clinical chemistry values, normal semen analyses, and normal serum gonadotropin levels. Sera were also obtained from 25 hypogonadal men (age range from 19–68 yr) who had serum T levels less than 300 ng/dl (10.4 nmol/liter, as previously determined by RIA at HUMC) before T therapy. In addition, sera were collected from 35 hypogonadal men after transdermal T replacement therapy. Of the samples from T-replaced hypogonadal men, 20 were within the normal range and 15 were above the normal range as previously determined by an RIA at HUMC.

Samples

The serum was stored at −20 C at HUMC. Since their original collection and aliquoting, the samples were thawed only once before the current study. Aliquots from each serum sample were pooled and mixed thoroughly by the laboratory supervisor before being aliquoted into portions for each of the laboratories participating in the study. Samples were bar-coded at HUMC and sent to the UCLA Olympic Analytical Laboratory for LC-MSMS assay and to the Penn State-Hershey Medical Center Core Endocrine Laboratory for RIA and for assay on four different automated instruments. The bar codes were linked to a database that contained demographics including the origin of the sample, the date of the sample collection, and the original T concentration assayed at the HUMC. This database was maintained by the laboratory supervisor at HUMC and was not made available to the investigators or the different technicians performing the assays. To maintain blinding of the samples at the HUMC, an aliquot of each sample was sent to the Penn State-Hershey Medical Center Core Endocrine Laboratory where each sample was recoded and sent back to the HUMC for assay. The listing of the recoded samples were not made available to the HUMC until all T assays were performed and entered into a database by an independent data manager. Thus, all samples were assayed in the different laboratories without prior knowledge of the serum T concentrations of the samples.

Methods

All assays used appropriate quality control material and standards either as steroid-free serum samples spiked with T or samples provided, by the manufacturer as defined by the standard operating procedures established and validated in each laboratory. Steroid-free sera were charcoal stripped sera prepared in the laboratory, newborn bovine serum, or steroid free sera obtained commercially. These steroid-free sera were tested in each individual laboratory to ensure that they did not show any T at the limit of detection of the assay used in each laboratory. All samples were measured similarly to other test samples run in each laboratory. For LC-MSMS, each sample was extracted and injected into the LC-MSMS once because of inadequate serum volume for replicates for most test samples. As routinely done at the laboratories performing the RIAs, the serum T result for each sample was determined from the

TABLE 1. Examples of serum total testosterone (ng/dl) external quality control program (College of American Pathologists, sample Y-04)

Instrument/assay	No. of labs	Mean (ng/dl)	SD	CV	Median	Range	
						Low	High
Abbott Architect	11	243.5	13.8	5.7	243	219	262
Bayer ACS:180	83	317.6	39.0	12.3	314	227	410
Bayer Centaur	231	324.0	41.5	12.8	319	234	454
Bayer Immuno-1	43	300.6	16.7	5.6	300	254	335
Beckman Access/2	98	297.8	15.3	5.1	298	239	330
Diagnostic Systems solid	10	352.7	80.1	22.7	375	177	440
DPC Coat-a-Count	76	277.8	34.2	12.3	281	196	363
DPC Immulite	86	232.0	32.9	14.2	228	160	330
DPC Immulite 2000	83	210.8	33.5	15.9	215	130	299
Roche Elecsys/E170	87	349.9	23.0	6.6	348	299	408
Ortho Vitros ECi	54	282.3	15.8	5.6	280	254	324
All instruments	891	293.6	56.2	19.1	297	160	508

average of two duplicates. Samples were run in singlicate on all four automated immunoassay instruments as specified by the procedure manuals of each laboratory. Data from all laboratories were sent to the HUMC and data entry validated before statistical analyses. The characteristics of the various methods are listed in Table 2 and detailed below.

LC-MSMS

The UCLA Olympic Analytical Laboratory used LC-MSMS to quantitate serum T levels. Advantages of the LC-MSMS method include easy and simple sample preparation (nonderivatized steroids can be analyzed directly), high recovery with improved signal to noise ratio, enhanced specificity, and low interference due to MSMS technology (7–9). A 2.0-ml sample was used for analyses and trideuterated T was used as the internal standard to monitor recovery. A LC-10A Shimadzu binary pump LC equipped with a PE-Applied Biosystem (Foster City, CA) PE Series 200 autosampler was used for LC and an Applied Biosystem-Sciex API-300 triple quadruple mass spectrometer equipped with an API interface was used to perform the T analysis.

The LC-MSMS method was validated using protocols specified by the Federal Drug Administration. This included determining the limit of detection (10), the limit of quantitation (LOQ), the characteristics of the calibration curve, and the within- and between-day reproducibility at three different concentrations of serum T. The standard curve for T was linear between 0 and 2000 ng/dl (0–69 nmol/liter) and the calibration plots over four days showed a slope 0.752–0.787, intercept 0.068–0.139, regression coefficient 0.997 to 0.999. The LOQ was 20 ng/dl (0.69 nmol/liter) and the accuracy for that level was 84.6% of the nominal value with %CV (coefficient of variation) of 9.4%. The between-day %CV was 7.4, 6.1, and 6.5 at 50, 750, and 1500 ng/dl, respectively. The dynamic range of the assay is 20 to 2000 ng/dl or 0.7–69.4 nmol/liter. Bovine newborn serum (determined by LC-MSMS to contain less than 20 ng/dl of T, LOQ of assay) was spiked with T (Sigma, St. Louis, MO) determined to be 99.8% pure by LC-MSMS and gas chromatograph (GC)-MS. The accuracy was 100.7, 93.6, 100.4, 100.3,103.5, and 97.8 for samples known to contain 20, 50, 250, 100, 500, 1000, and 2000 ng/dl, respectively. The corresponding precision values were: 10.5, 10.4, 7.2, 4.8, 1.7, and 5.9%. Recovery (% recovery of the analyte during analysis) was 77.0% at 50 ng/dl, 76.9% at 750 ng/dl, and 71.4% at 1500 ng/dl. Only a single extraction and injection were performed for each sample due to inadequate serum volume for replicate assays for most samples.

During the study, the standard curve was linear between 0 and 2000 ng/dl (0–69 nmol/liter) of T concentrations and the calibration lines for 4 d showed a slope 0.789–0.833, intercept 0.072–0.301, regression coefficient 0.997–0.999. The LOQ was 20 ng/dl (0.69 nmol/liter) and the accuracy for that level was 85.2% of the nominal value with %CV of 17.9%. The interday %CV was 10.5, 8.6, and 8.4 at 50, 750, and 1500 ng/dl. The accuracy was 110.4, 98.1, 98.5, 98.3, 96.6, and 102.4% for samples known to contain 20, 50, 250, 100, 500, 1000, and 2000 ng/dl, respectively. The corresponding values for precision were: 10.4, 8.3, 5.7, 9.5, 6.5, and 3.2%.

RIAs

RIA at HUMC. Serum T was measured by a T RIA using reagents including the iodinated tracer obtained from ICN (Costa Mesa, CA). The cross reactivity of the ICN antibody used in the T RIA were 2.0% for 5α-dihydrotestosterone, 2.3% for androstenedione, 0.8% for 3α-androstanediol, 0.6% for etiocholanolone, and less than 0.01% for all other steroids tested (from 0.1–1000 ng/ml, up to 200-fold of the highest T standard). Before analysis, the samples (0.1 ml) were extracted with 2.0 ml of ethyl acetate:hexane, 3:2 (vol:vol). Initially tritiated T was used as an internal standard for each sample. The average recovery of the internal standard was 102 ± 1% (range 99.6–105.1%). Because of the proven minimal procedural loss, subsequently no internal standard was used to correct for the extraction. The extract was then dissolved in the assay buffer and two aliquots were assayed in sequence in the RIA. The average of the T levels in each of the two aliquots were reported. This RIA was validated using the guidelines published by Shah *et al.* (11). The following were data from the validation studies. The lower limit of quantitation of serum T measured by this assay was 0.87 nmol/liter (25 ng/dl). This was the lowest concentration of T measured in serum that can be accurately distinguished from steroid-free serum with a 12% CV. The accuracy of the T assay, determined by spiking steroid-free serum (ICN) with 25, 50, 100, 500, 1000, and 1500 ng/dl of T was 114, 118, 109, 94, 92, and 92%, respectively (mean 104%). The T was obtained from Sigma and was 99.8% as determined by celite column chromatography. The within-run precision (CV) at a serum T concentration of 646 ng/ml (22.4 nmol/liter) was 5.9%. The between-run precision (CV) for low, medium, and high serum T concentrations of 136, 531, and 1477 ng/dl (4.7, 18.4, and 51.2 nmol/liter) was 12.4, 9.3, and 12.5%, respectively. The adult male reference range in this laboratory was 298-1043 ng/dl (10.33 to 36.17 nmol/liter) determined from samples in young men (18–50 yr) with normal physical examination, serum gonadotropin and semen analyses (12, 13). This RIA was developed and validated primarily for research studies in men. Although not used in this study, a separate protocol was available using more serum for extraction of samples suspected of containing very low T levels such as that seen in women and children. All the samples for this study were done in three assays on three different days where two sets of quality control samples were run with each assay. The interassay CV for serum T levels of 101, 518, and 1201 ng/dl were 15.4, 14.0, and 9.1%, respectively. The HUMC-RIA protocol required repeating the analyses if the CV for the duplicate counts exceeds 10%; however, in this study all CV were less than 10%.

RIA at Penn State-Hershey Medical Center. Serum T was measured using the DPC coat-a-tube RIA method (Diagnostic Products Corp., Los Angeles, CA). This method used an iodinated tracer and a T-specific antibody immobilized to the wall of a polypropylene tube. Duplicates samples were run in sequence in the assay and the average serum T levels were reported. Antibody cross-reactivity against androstenedione, 3β-androstanediol, dehydroepiandrosterone, and other possible interfering steroids was less than 1%. Cross-reactivity with 5α-dihydrotestosterone was 2.8%. Accuracy studies averaged 101% with steroid-stripped serum samples spiked with T (purity ascertained by celite

TABLE 2. Characteristics of the methods

Assay	LLOQ (ng/dl)	Accuracy (%)	Interassay Precision (CV%)	Reference[a] range for adult men (ng/dl)
LC-MSMS	20	84.6–110.4	8.0 at 750 ng/dl	
HUMC-RIA	25	92–118%	9.3 at 530 ng/dl	298–1043
DPC-RIA	14	101%	5.3 at 602 ng/dl	250–900
Roche Elecsys	11.5	NA	4.3 at 271 ng/dl	210–810
Bayer Centaur	34.6	NA	7.3 at 671 ng/dl	241–827
Ortho Vitros ECi	14	NA	2.8 at 271 ng/dl	132–813
DPC Immulite 2000	49	NA	13.7 at 427 ng/dl	286–1510

LLOQ, Lower limit of quantitation.

[a] Reference ranges for HUMC-RIA and DPC-RIA were determined from serum obtained in healthy men between the ages of 18 and 50 yr with normal physical examination, serum gonadotropins, and normal gonadal semen analyses. The ranges for automatic immunoassays were based on reference ranges quoted by manufacturer. Each individual laboratory then verified the reference range with samples from normal men with normal gonadotropin levels and normal physical examination.

column chromatography) at a concentration of 250 ng/dl (8.7 nmol/liter). The within-run precision (CV) at a serum T concentration of 545 ng/dl (18.9 nmol/liter) was 3.9%. The between-run precision (CV) for samples with low, medium and high serum T concentrations of 83.6, 602, and 1229 ng/dl (2.9, 20.9, and 42.6 nmol/liter) was 11.4, 5.3, and 4.5%, respectively. The assay reportable range extends from 14–1600 ng/dl (0.5–55.5 nmol/liter). The adult male reference range for this assay was 250–900 ng/dl (8.7–31.2 nmol/liter). During the study the between run CV averaged 4.8%.

Automated platform assays

The measurement of T on the different automated immunoassay systems was carried out at four institutions including The Penn State-Hershey Medical Center, Hershey, PA; The University of Pennsylvania; Mercy Health Laboratories; and Henry Ford Hospital. The automated systems included the Roche Elecsys, the Bayer Centaur, the Ortho Vitros ECi, and the DPC-Immulite 2000. The references range quoted in Table 2 are based on those provided by the manufacturer. These reference ranges were verified by the individual laboratories using serum samples obtained from men with normal physical examination and normal gonadotropins.

Roche Elecsys. The Elecsys 2010 automated analyzer (Roche Diagnostics GmbH, Mannheim, Germany) measures T in serum using electrochemiluminescence. This assay uses a highly specific antibody to measure T. Briefly, 50 μl of serum and a biotinylated antibody against T are incubated together. A second antibody labeled with a ruthenium complex is then added together with streptavidin-coated microparticles. A sandwich complex is formed that is bound to the solid phase (the microparticles) via biotin-streptavidin interaction. The microparticles are then magnetically captured onto the surface of an electrode. Application of voltage on this electrode induces a chemiluminescence emission, which is detected by a photomultiplier and the signal compared with a T calibration curve, which is instrument-specific. This instrument uses a two-point calibration curve for day-to-day analysis, and a master curve provided by the manufacturer for each lot of reagents. A three-level assay control provided by the manufacturer was used with each assay run. The LOQ of the Elecsys T assay is 11.5 ng/dl (0.4 nmol/liter) and between-run precision averaged 4.3% at a concentration of 271 ng/dl (9.4 nmol/liter). The reference range for adult males for this method was 210–810 ng/dl (7.3–28.1 nmol/liter). During the study the between run CV averaged 4.6%.

Bayer (Centaur). The Bayer ACS Centaur (Bayer Diagnostics, Tarrytown, NY) is a fully automated random access immunoassay analyzer that used paramagnetic solid-phase particles and an acridinium ester-based direct chemiluminescence tracer that is coupled to T antibodies in a second reagent. After magnetic separation and washing of the particles, luminescence is initiated by the addition of an acid and base reagent. Individual assays are calibrated using a two-point calibration curve and a three level assay control is used with each run. A master curve is provided for each lot of reagents. The functional sensitivity of the Centaur T assay was 34.6 ng/dl (1.2 nmol/liter) and between run precision at a concentration of 671 ng/dl (23.3 nmol/liter) averaged 7.3%. The reference range for adult males was 241–827 ng/dl (8.36–28.7 nmol/liter). During the study, the between run CV averaged 6.8%.

Ortho Vitros Eci. The Vitros T assay is performed using the Vitros T Reagent Pack and Vitros Immunodiagnostic Product T calibrators on a fully automated random access immunoassay system that used enhanced chemiluminescence technology with horseradish peroxidase (HRP) as a label and a luminol substrate for signal detection (Ortho Clinical Diagnostics, Rochester, NY). The assay depends on competition between T present in a serum sample with an HRP-labeled T conjugate for binding sites on a biotinylated mouse anti-T antibody. The antigen-antibody complex is then captured by streptavidin in the incubation wells. Following a wash step, the bound HRP conjugate is determined by a luminescence reaction with a luminol derivative and a peracid salt. The HRP in the bound conjugate catalyzes the oxidation of the luminal derivative, producing a flash of light. An electron transfer reagent is present to enhance the level of light produced prolonging its emission spectra. The amount of HRP conjugate bound is in direct proportion to the concentration of T present in the sample. Calibration is lot specific, and the T calibrators are supplied by the manufacturer ready for use. On board calibration stability is 28 d. A three-level control was run with each assay run. The calibration range of the Vitros T assay is 0–2163 ng/dl (0–75 nmol/liter) (calibrated against samples measured by isotope dilution-gas chromatography/mass spectrometry, ID-GC/MS). The functional sensitivity of the Vitros T assay was 14 ng/dl (0.5 nmol/liter) with a between run precision of 2.8% at a concentration of 271 ng/dl (9.4 nmol/liter). The reported adult male range was 132–813 ng/dl (4.6–28.2 nmol/liter). During the study, the between run CV averaged 3.6%.

DPC Immulite 2000. The Immulite 2000 is an automated, random-access immunoassay analyzer with a solid-phase washing process and a chemiluminescence detection system. The solid phase is made up of a polystyrene bead enclosed within the Immulite test unit that is coated with a polyclonal rabbit antibody specific for T. The patient's serum sample and an alkaline phosphatase-conjugated T reagent are simultaneously introduced into the test unit. During a 60-min incubation period at 37 C with intermittent shaking, the T in the serum sample competes with the enzyme-labeled T for a limited number of antibody binding sites on the bead. Unbound enzyme conjugate is then removed by a patented five-spin-wash technique. The chemiluminescence substrate, a phosphate ester of adamantyl dioxetane, is added and the test unit incubated for 10 min. The substrate is hydrolyzed by the alkaline phosphatase to an unstable anion. The decomposition of the anion yields a sustained emission of light. The bound complex, corresponding to the photon output, is inversely proportional to the concentration of T in the sample. A single determination uses 25 μl of serum, and the dynamic range of the Immulite T assay is 14 to 1586 ng/dl (0.5–55 nmol/liter). The functional sensitivity for the T assay on this system is 49 ng/dl (1.7 nmol/liter) and the average between run imprecision was 13.7% at a concentration of 427 ng/dl (14.8 nmol/liter). The normal range for adult male between 20 and 49 yr is reported to be 286-1510 ng/dl (9.9–52.4 nmol/liter). During the study, the between run CV averaged 11.5%.

Data analyses

Because serum T concentrations were not normally distributed, we estimated the median and the 10th, 25th, 75th, and 90th percentiles of the values obtained from the different methods. The serum T results obtained from the four automated immunoassay systems and the two RIAs (test methods) were compared with values obtained with the LC-MSMS method to determine the extent of agreement among methods (14). Deming regression was used to estimate the slope and intercept (15). We computed the interclass correlation coefficient (16). Plots of the percent differences of the values between two methods (test *vs.* LC-MSMS) *vs.* the mean of the values generated by the two methods as initially described by Bland and Altman were used (17–20) to identify other types of systematic bias.

Of the 122 samples that were distributed, seven were below the LOQ in one or more assays, 13 were not analyzed in all assays (inadequate volume of serum) and one sample was excluded from the analysis because the result from one method were one third that of the others (outlier). The data analyses were based on 101 samples. Because the serum T values spanned a large range (<50–1500 ng/dl), our sample size of 101 samples should provide stable estimates for the measures of agreement, should not be influenced by individual variables, and should be reproducible in other studies (21). The use of samples from hypogonadal men as well as normal men assured that our results would cover the widest range of possible T values seen in clinical practice in adolescent and adult men.

Results

Comparison of median and range

Figure 1 shows the median and the 10th, 25th, 75th, and 90th percentiles of the serum T levels measured by the seven different methods. Compared with the median serum T value obtained by LC-MSMS (462 ng/dl), the median value determined by the DPC Immulite was lower (318 ng/dl), whereas the median T result obtained from the Bayer Centaur was higher (514 ng/ml). The median serum T levels

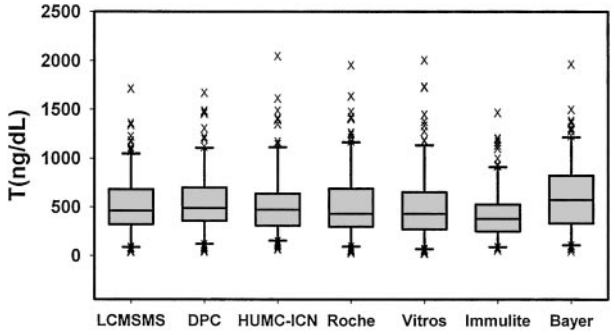

FIG. 1. Median levels of serum T measured by the seven different methods. *Line within the box* represents the median, l*ower boundary of box* indicates the 25th percentile, and the *upper boundary of box* indicates the 75th percentile. Whiskers above and below indicate the 90th and 10th percentiles. x, Outlying points.

determined by DPC-RIA, HUMC-RIA, Roche Elecsys, and OrthoVitros ECi were similar to LC-MSMS at 490, 473, 431, and 431 ng/dl, respectively.

Comparison using regression analyses and correlation coefficient

Figure 2 shows the Deming regression analyses for the RIAs and platform analog assays *vs.* LC-MSMS. Table 3 gives the slope and intercept of the Deming regression the interclass correlation coefficient and the 95% confidence interval for all parameters. The slope was closest to one between the DPC-RIA and LC-MSMS (1.098), whereas the other assays ranged from 0.881 (DPC Immulite) to 1.217 (Ortho Vitros ECi). The intercepts for DPC-RIA and Beyer Centaur are not significantly different from zero. The Vitros ECi intercept was the largest. The interclass correlation coefficient for all methods was between 0.92 and 0.97. The 95% confidence intervals for this correlation were 0.63–0.97 and 0.71–0.96 for DPC Immulite and Bayer Centaur, respectively, and exceeded 0.92 for the other four assays.

Assessment of agreement and bias between methods

Figure 3 shows the plots of the percent difference between each method and LC-MSMS against the means of serum T concentrations obtained by LC-MSMS and the values obtained by each immunoassay. The plots also showed percent difference ± 2 SD (95% limits of agreement). In the quoted adult male range (between 300-1000 ng/dl or 10.4–34.7 nmol/liter), agreement of serum T concentrations among the two RIAs, Roche Elecsys, Ortho Vitros ECi were within ± 20% in over 60% of the samples of that measured by LC-MSMS (Fig. 3, A–D, and Table 4). As shown in Fig. 3, the average percent difference in serum T levels between DPC-RIA, HUMC-RIA, Roche Elecsys, Ortho Vitros ECi, DPC Immulite and Bayer Centaur and LC-MSMS were +9.7, +9.7, −3.4, −11.2, −18.7, and +15.9%, respectively. The mean differences in measured serum T levels between DPC-RIA, HUMC-RIA, Roche Elecsys, Ortho Vitros ECi and LC-MSMS were +48.1 ± 7.5, +33.8 ± 11.1, 10.8 ± 9.6, and −3.5 ± 11.2 ng/dl, respectively. At serum T levels above the adult reference range, the values obtained by LC-MSMS were lower than all the other methods except the results obtained with the DPC Immulite. It is evident from Fig. 3 that compared with LC-MSMS in the adult male reference range, the DPC Immulite assay generally underestimates the serum T values (mean difference −90 ± 8.7 ng/dl; Fig. 3E). In contrast, the Bayer Centaur overestimates serum T levels (mean difference +99 ± 11 ng/dl; Fig. 3F).

The left side of each graph shows more clearly the differences between the methods when serum T levels were considerably below the adult male reference range. At values less than 100 ng/dl (3.47 nmol/liter), the percent difference between DPC-RIA and LC-MSMS varied between −40% and +40% (Fig. 3A). Similarly, the percent difference between T values estimated by Roche Elecsys and LC-MSMS ranged from −80 to +40% (Fig. 3C). At low serum T concentrations (<100 ng/dl), the HUMC-RIA was biased in the high direction (+20 to 80%; Fig. 3B) and the Ortho Vitros ECi in the low direction (0 to −100%; Fig. 3D). Figure 3E shows that the serum T values at low serum T levels obtained by the DPC Immulite is again systematically biased in the low direction for serum T values and those measured by the Bayer Centaur is systematically biased in the high direction for samples at all T concentrations (Fig. 3F).

For the 102 samples analyzed by all seven methods, Table 4 shows the percent of the T values obtained by the various test methods that fell outside ± 20% of the LC-MSMS values. It can be seen from Table 4 that 19.8, 25.7, 39.6, 39.6, 48.5, and 50.4% of the samples fell outside the ± 20% range of the LC-MSMS generated serum T value by DPC-RIA, Roche-Elecsys, Ortho Vitros-Eci, HUMC-RIA, Immulite and Bayer, respectively. This difference was especially noted in the samples with T values less than 100 ng/dl (3.47 nmol/liter) obtained by the six different immunoassays, the majority (55.5–90.0% of the samples) fell outside the ± 20% range of those obtained by LC-MSMS.

Lower limit of quantitation

The LOQ of each assay is listed in Table 2. Seven samples were excluded because the serum T values measured by one or more of the assays were below the LOQ. One sample was below the LOQ of LC-MSMS, HUMC-RIA, Ortho Vitros ECi, and Immulite. Another sample was below the LOQ of all the platform methods. All seven samples were below the LOQ of DPC Immulite, whereas none were below the LOQ by DPC-RIA.

Discussion

In this study, we have compared serum total T levels using two RIAs and four automated analog platform assays against LC-MSMS as the reference method using the standard operating procedures for measuring clinical samples particular to each laboratory. The results indicate that despite an apparent good correlation as evidenced by the slope (between 0.88 and 1.23) and the interclass correlation coefficients (0.92–0.97) between the immunoassays and LC-MSMS method, there were systematic biases detected in some of the methods. Using Deming's regression, the DPC-RIA has a slope that was closest to one as well as a small intercept that was not significantly different from zero when compared with LC-MSMS. Others like the DPC-Immulite and the Bayer Cen-

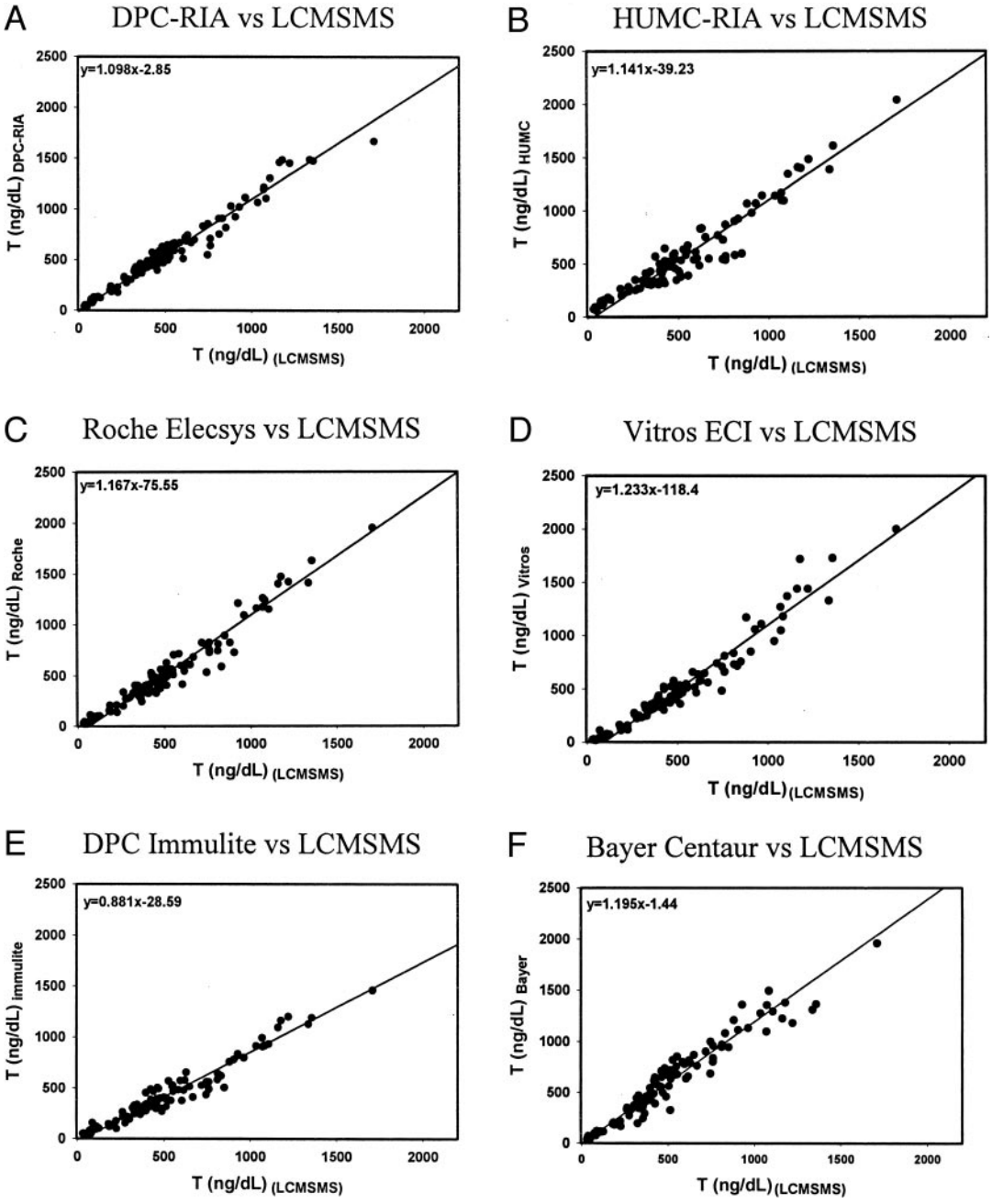

FIG. 2. Deming regression plots of serum T concentrations measured by the six different immunoassays (y-axis) against LC-MSMS (x-axis).

taur methods showed lower agreement with LC-MSMS with a lower 95% confidence interval of the correlation coefficient of 0.63 and 0.71, respectively. Our results corroborate those recently reported by Taieb *et al.* (22) who demonstrated that the serum T measured by GC-MS and 10 immunoassays showed correlation coefficients between 0.92 and 0.97 in male sera. They also indicated that only DPC-RIA and three other platform immunoassays not examined in our present study gave serum T levels that were not significantly different from GS-MS. It should be noted that the GC-MS method reported required extraction purification by ethylene-glycol impregnated celite chromatography and derivatization of the steroid before quantitation of T from the sample, which is more time consuming and complicated than our LC-MSMS assay.

Using the method described by Bland and Altman (17–20), which shows the relationship between the mean of LC-MSMS and various values of serum T on the x-axis and the percent difference the various assays from LC-MSMS value on the y-axis, the DPC-RIA, HUMC-RIA, Roche Elecsys and Bayer Centaur showed that all these methods gave T values higher than LC-MSMS, whereas the DPC Immulite and Ortho Vitros ECi gave lower values. When the individual graphs were examined, it was shown that values obtained by the Bayer Centaur showed a bias in the high direction. In

TABLE 3. The slope and intercept of Deming regression and interclass correlation coefficient for LC-MSMS *vs.* immunoassays

	Slope	Intercept	Interclass correlation coefficient
DPC-RIA	1.098 (1.032–1.165)	−2.9 (−30.9 to 25.2)	0.968 (0.918–0.984)
HUMC-RIA	1.141 (1.076–1.206)	−39.2 (−73.7 to −4.2[a])	0.948 (0.910–0.967)
Roche Elecsys	1.167 (1.112–1.222)	−75.5 (−102 to −49.1[a])	0.965 (0.939–0.978)
Vitros ECi	1.233 (1.136–1.330)	−118.4 (−160.5 to −76.4[a])	0.954 (0.921–0.971)
DCP Immulite	0.881 (0.838–0.924)	−28.6 (−49.8 to −7.4[a])	0.925 (0.628–0.969[b])
Bayer Centaur	1.195 (1.112–1.277)	−1.4 (−36.8 to 33.9)	0.919 (0.711–0.963[b])

Numbers in *parentheses* are 95% confidence intervals.
[a] Significantly different from zero.
[b] Data not exchangeable with LC-MSMS (see Ref. 16).

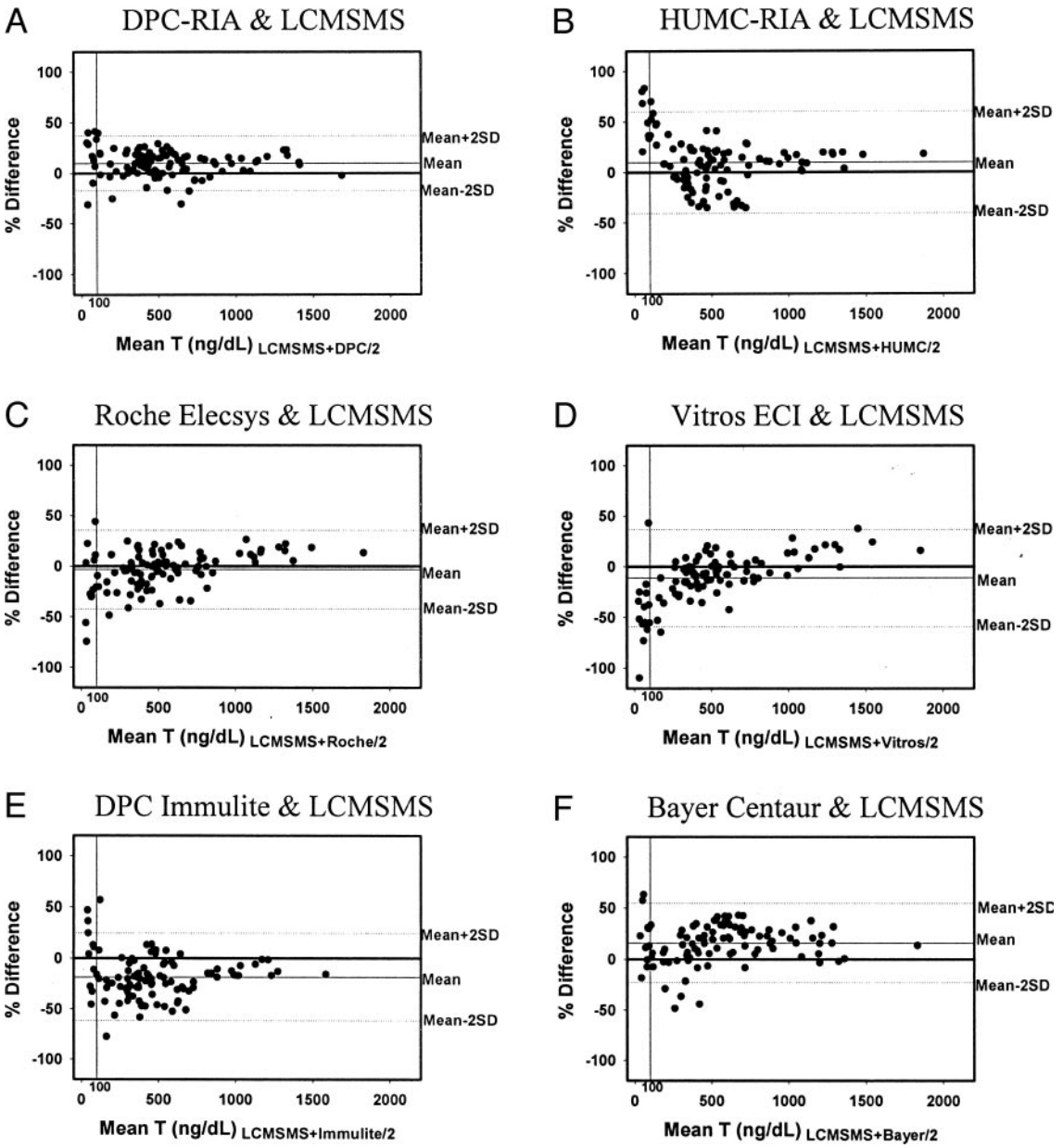

FIG. 3. Plots of percentage differences in serum T levels (test minus LC-MSMS) against the average of the two methods. The *bold solid line* represents 0%, the *light solid line* the mean percentage difference between the methods, and the *dashed lines* 2 SD of the mean percentage difference.

contrast, serum T values obtained by the DPC-Immulite were biased in the low direction. For both the DPC-RIA and HUMC-RIA the mean serum T was higher by 48 and 34 ng/dl, respectively, when compared with LC-MSMS. The comparison of mean serum T results obtained by Roche-Elecsys (+10.8 ng/dl) and Ortho Vitros ECi (−3.5 ng/dl)

TABLE 4. Samples with serum T values determined by the six assays outside of ±20% range of LC-MSMS values

	DPC-RIA	HUMC-RIA	Roche Elecsys	Ortho-Vitros ECi	DPC-Immulite	Bayer Centaur
Number of samples						
< −20% of LC-MSMS	3	12	19	25	45	5
> +20% of LC-MSMS	17	28	7	6	4	46
Samples outside ± 20% of LC MSMS values (%)						
All T values	20/101 (19.8%)	40/101 (39.6%)	26/101 (25.7%)	40/101 (39.7%)	49/101 (48.5%)	51/101 (50.4%)
T value <100 ng/dl	7/11 (63.6)	10/11 (90.9%)	6/11 (55.5%)	10/11 (90.9%)	7/11 (63.6%)	6/11 (55.5%)

were less different from those obtained by LC-MSMS. These differences in serum T levels are not clinically relevant in the adult male reference range. Using GC-MS as the standard method and the Bland-Altman analyses, Taieb *et al.* (22) also reported that Roche Elecsys underestimated serum T levels that was not demonstrated in our study, whereas their results demonstrating that Bayer Centaur displayed a positive and DPC-Immulite a negative bias for male sera concurred with our data. They also reported the DPC-RIA displayed no bias in male range but overestimated serum T in the female range which was quite similar with our findings. When the percent differences were plotted against the means, using LC-MSMS as the reference method, the largest difference was observed in the serum T concentrations less than 100 ng/dl (3.47 nmol/liter). Again, the values of serum T obtained by DPC Immulite were systematically lower and those by the Bayer Centaur higher than LC-MSMS. At very low serum T values compared with the LC-MSMS method, the HUMC-RIA was biased toward the high direction, whereas the Ortho Vitros ECi was biased in the low direction. The DPC-RIA and Roche Elecsys showed large percent difference both in the high and low directions. The results indicate that none of the assays as performed are of sufficient accuracy at low serum T levels using LC-MSMS as the gold standard. Our data are similar to the previous findings comparing immunoassays with GC-MS demonstrating that none of the immunoassays tested was sufficiently reliable for investigation from children and women (22). However, from a clinical use perspective, the RIA and some automated methods would be acceptable for use in adult males even at the very low range (<100 ng/dl, 3.47 nmol/liter) as these males would be diagnosed to be hypogonadal who would be investigated and treated with T. The RIAs and some of the automated methods may also be acceptable for discerning abnormal elevations in T (above 100 ng/ml, 3.47 nmol/liter) in females and prepubertal children. The dose-response curve of RIAs, immunoradiometric assays, and enzyme-linked immunosorbent assay are nonlinear and various curve-fitting methods have been used. The most common data reduction method in use is the four-parameter logistics model (23–25). Despite use of these curve-fitting techniques, only a segment of the standard curve is linear with relatively low variance. For many immunoassays, low concentrations of the hormone are measured at a portion of the calibration (standard) curve where the variance is larger than that at the more linear portion of the calibration (standard) curve. This is not the case for LC-MSMS where the calibration curve is linear. The RIAs designed for serum T assays are standardized for use in male serum and optimized for lower variance in the adult male range (*e.g.* HUMC-RIA and DPC-RIA). Because of the high variance of the immunoassays at low concentrations as illustrated by the data from this study, a high proportion of samples with serum T values less than 100 ng/dl when measured by various immunoassays were outside of ± 20% range of the LC-MSMS values (55.5% for Roche Elecsys and Bayer Centaur, 63.6% for DPC-RIA and DPC Immulite, and 90.9% for HUMC-RIA and Ortho Vitros ECi). Based on these data, we conclude that these assays should be modified to increase their sensitivity and accuracy at low serum T levels less than 100 ng/dl (3.47 nmol/liter) to improve their applicability to serum T measurements in prepubertal children and female serum. For the RIAs, increased sensitivity can be achieved by adjusting the antibody titer, selecting more specific antibodies, preincubation of the antibodies with the test serum (nonequilibrium), and changing methods for the separation of bound from free hormone. For the automated platform assays, the reagents, the time of reaction, and the capture antibody may be adjusted by the manufacturer to produce more accurate and precise results in ranges capable of measuring low serum T levels expected for normal women and children.

From our results, all assays without a relatively large systematic bias for the adult male range (*i.e.* DPC-RIA, HUMC-RIA, Roche Elecsys and Ortho Vitros ECi) would be acceptable assays for measuring adult male sera. These assays could also be used for the diagnosis for male hypogonadism usually defined as serum T values less than 300 ng/dl (10.4 nmol/liter). For a serum sample in a male with a T concentration at or less than 200 ng/dl (6.9 nmol/liter), a method that measures serum T above +40% of LC-MSMS values, would give a T value of 280 ng/dl (9.7 nmol/liter) that would be below the normal adult male range of 300 ng/dl. It is however essential that each laboratory using their own method establish a reference range specific for subjects of interest, for example young adult males, women, prepubertal children.

The lower LOQ was 0.69 nmol/liter (20 ng/dl) for the LC-MSMS method when 2 ml of sera was used. This LOQ was similar to a prior report using LC-MSMS in bovine sera (26) and could be lowered by using more sera and revalidated for female samples. For the DPC Immulite, seven of 122 samples were below the LOQ. DPC-RIA gave readings above the LOQ for all these seven samples and LC-MSMS and HUMC-RIA each reported one sample below the LOQ. It should be noted that in this comparison study a standard volume of serum was used as routinely performed for each assay. In laboratory practice, more serum could be used in some of these assays to bring the LOQ to a lower threshold.

If more serum were used in the assays, validation studies would need to be done to ensure that increasing the amount of serum would not affect the characteristics of the assay.

Because of the limitation of the volume of serum available for this study, the values obtained by LC-MSMS were based on a single sample that was taken through extraction, LC followed by mass spectroscopy. Despite this limitation, the LC-MSMS assay underwent vigorous validation with a linear calibration curve spanning 20–2000 ng/dl, accuracy between 96.6 and 110.4% and precision of less than 10% at all points except for the LOQ results (8). The range of serum T values obtained in 17 normal men ages 18–50 yr in this study was 302–905 ng/dl by the LC-MSMS T method.

As shown in the College of American Pathologists quality control program, the four instrument-based assays we evaluated were some of the commonest used by laboratories participating in this program. The DPC-RIA (DPC-Coat-a-Count) is the most common RIA used in hospital or reference laboratories and appears to show the best agreement with serum T values measured in male serum by LC-MSMS. The RIAs used by the Penn State-Hershey Medical Center (DPC-RIA) and the HUMC-RIA were both fully validated according to standard procedures recommended (11). The HUMC-RIA uses an extraction step. An internal standard was not used to monitor procedural losses because during initial validation this was found not to improve assay performance. Possibly because of this reason, the HUMC-RIA had a higher LOQ and higher interassay and intraassay variability than the DPC-RIA. The medians for all the evaluable serum T values were 490 and 473 ng/dl for DPC-RIA and HUMC-RIA, respectively. The correlation coefficient between the two RIAs was 0.964 and Deming's regression with T values measured by HUMC-RIA on the vertical axis showed a slope of 1.05 and an intercept of −85.6 ng/ml (data not shown). There was no systematic bias between the two RIAs, and these two assays also gave similar adult male range.

The automated assay instruments are widely used in clinical and reference laboratories. Our comparison results indicate that the DPC Immulite gives T values that are biased in the low direction. This assay also had a high LOQ (49 ng/dl). The normal range given by the manufacturer (286–1510 ng/dl) had a similar low male reference range as other methods but with an extremely high upper limit. This suggests that the adult male range might not have been generated by each laboratory and both the lower and the upper limit of the reference range might have to be adjusted. The Bayer Centaur assay on the other hand showed a systematic bias toward higher serum T levels when compared with LC-MSMS. Despite this bias toward higher values, the reference range for adult men with this instrument is reported as 241–827 ng/dl. This range obtained from the manufacturer should be validated in each laboratory that uses this instrument with an adequate number of adult healthy male samples as suggested by Shah *et al.* (11). Our study suggests that the reference range quoted by the manufacturer may be inappropriate for individual laboratories and the determination of reference ranges for male, female, and children's serum should be determined by each laboratory using this method.

We conclude that using LC-MSMS as our gold standard for estimating serum T levels in male serum, the DPC-RIA, the Roche Elecsys, the Ortho Vitros ECi, and HUMC-RIA gave results that are within the clinically acceptable limits of ± 20% of the reference method in over 60% of the samples. At low T concentrations (≤100 ng/dl), HUMC-RIA is biased toward higher values, whereas the Ortho Vitros ECi results are biased toward lower values. The DPC Immulite method showed a systematic bias in the low direction, whereas the Bayer Centaur was biased in the high direction for serum T levels at all concentrations. In this study, the DPC-RIA and Roche Elecsys methods for determining serum T levels show the closest correlation with values determined by LC-MSMS. Without modification, none of the automated methods are currently acceptable for the measurement of T in the serum of normal females or children. These methods lack adequate precision, accuracy, and have a sufficiently low limit of quantitation to preclude their use in these populations. Because free T measurements either directly by equilibrium dialysis, from bioavailable T calculations or from a total T to sex hormone binding globulin ratio are dependent on an accurate T measurement, the results of this study has significant implications on free T determinations as well (27).

Acknowledgments

The authors thank Nancy Berman, Ph.D., for her advice with the statistical analyses. This study would not have been possible without the effort of Andrew Leung, HTC, who coordinated all the samples and the assays at the HUMC. We thank Alfred De Leon from the UCLA Olympic Analytical Laboratory for the excellence in analytical chemistry. Chris Hamilton from the Penn State Core Endocrine Laboratory at Hershey kindly blinded all of the samples for the different assay methods and shipped samples to the individual clinical laboratories for analysis. We also thank those laboratories who performed T analysis on the automated platforms: Drs. Peter Wilding and Marilyn Senior at the University of Pennsylvania, William Pepper Laboratories, Dr. Carolyn Feldkamp at Henry Ford Hospital, and Dr. Bette Seamonds at Mercy Health Systems. We thank Laura Hull, B.A., who managed the database and was responsible for the graphic presentations, and Sally Avancena, M.A., for preparing the manuscript.

Received July 24, 2003. Accepted September 29, 2003.

Address all correspondence and requests for reprints to: Christina Wang, M.D., UCLA School of Medicine, General Clinical Research Center, Box 16, 1000 West Carson Street, Torrance, California 90502.

This work was supported by the Core Endocrine Laboratory at Penn State-Hershey Medical Center; National Institutes of Health (NIH) Grant MO1 RR00543 to the GCRC at Harbor-UCLA Medical Center; NIH Grants RO1 CA 71053 and RO1 DK 61006 (to C.W., D.H.C., and R.S.S.); and United States Anti-Doping Agency (to D.H.C.). The samples were collected by the nurses of the Harbor-UCLA GCRC, supported by NIH Grant MO1 RR00425.

References

1. **Furuyama S, Mayes D, Nugent C** 1970 A radioimmunoassay for plasma testosterone. Steroids 16:415–428
2. **Chen J, Zoru E, Hallberg M, Wieland R** 1971 Antibodies to testosterone-3-bovine serum albumin applied to assay of serum 17-β-ol androgens. Clin Chem 17:581–584
3. **Dufan M, Catt K, Tsuruhara T, Ryan D** 1972 Radioimmunoassay of plasma testosterone. Clin Chem Acta 37:109–116
4. **Wang C, Youatt G, O'Connor S, Dulmanis A, Hudson B** 1974 A simple radioimmunoassay for plasma testosterone plus 5 α dihydrotestosterone. J Steroid Biochem 5:551–555
5. **Larsen PR, Kronenberg HM, Melmed S, Polonsky K** 2002 William's textbook of endocrinology, reference values. 10th ed. Philadelphia: Saunders
6. **Dorgan JF, Fears TR, McMahon RP, Friedman LA, Patterson BH, Greenhut SF** 2002 Measurement of steroid sex hormones in serum: a comparison of radioimmunoassay and mass spectrometry. Steroids 67:151–158

7. **Gelpi E** 1995 Biochemical and biochemical applications of liquid chromatography-mass spectrometry. J Chromatogr A 703:59–80
8. **Sheffield-Moore M, Urban RJ, Wolf SE, Jiang J, Catlin DH, Herndon DN, Wolfe RR, Ferrando AA** 1999 Short-term oxandrolone administration stimulates net muscle protein synthesis in young men. J Clin Endocrinol Metab 84:2705–2711
9. **Starcevic B, DiStefano E, Wang C, Catlin DH** 2003 An LC-MS-MS assay for human serum testosterone and deuterated testosterone. J Chromatogr B Analyt Technol Biomed Life Sci 792:197–204
10. **Lang JR, Bolton S** 1991 A comprehensive method of validation strategy for bioanalytical applications in the pharmaceutical industry-2. Statistical analyses. J Pharm Biomed Anal 9:435–442
11. **Shah VP, Midha KK, Findlay JWA, Hill HM, Hulse JD, McGilveray IJ, McKay G, Miller JJ, Patnaik RN, Powell ML, Tonelli A, Viswanathan CT, Yacobi A** 2000 Bioanalytic method validation—a revisit with a decade of progress. Pharm Res 17:1551–1557
12. **Wang C, Berman N, Longstreth JA, Chuapoco B, Hull L, Steiner S, Faulkner S, Dudley RE, Swerdloff RS** 2000 Pharmacokinetics of transdermal testosterone gel in hypogonadal men: application of gel at one site versus four sites: a general clinical research center study. J Clin Endocrinol Metab 85:964–969
13. **Swerdloff RS, Wang C, Cunningham G, Dobs A, Iranmanesh A, Matsumoto A, Snyder P, Weber T, Berman N, and T gel Study Group** 2000 Comparative pharmacokinetics of two doses of transdermal testosterone gel *versus* testosterone patch after daily application for 180 days in hypogonadal men. J Clin Endocrinol Metab 85:4500–4510
14. **Magari RT** 2000 Evaluating agreement between two analytical methods in clinical chemistry. Clin Chem Lab Med 38:1021–1025
15. **Linnet K** 1998 Performance of Deming regression analysis in case of misspecified analytic error ratio in method comparison studies. Clin Chem 44: 1024–1031
16. **Perisic I, Rosner B** 1999 Comparison of measures of interclass correlation: the general case of unequal group size. Stat Med 18:1451–1466
17. **Bland JM, Altman DG** 1986 Statistical method for assessment of agreement between two methods of clinical measurement. Lancet i:307–310
18. **Bland JM, Altman DG** 1999 Measuring agreement in method comparison studies. Stat Methods Med Res 8:135–160
19. **Pollock MA, Jefferson SG, Kane JW, Lomax K, Mackinnon G, Winnard CB** 1992 Method comparison—a different approach. Ann Clin Biochem 29: 556–560
20. **Dewitte K, Fierens C, Stockl D, Thienport LM** 2002 Application of the Bland-Altman plot for interpretation of method-comparison studies: a critical investigation of its practice. Clin Chem 48:799–801
21. **Linnet K** 1999 Necessary sample size for method comparison studies based on regression analysis. Clin Chem 45:882–894
22. **Taieb J, Mathian B, Millot F, Patricot M-C, Mathieu E, Queyrel N, Lacroix I, Somma-Delpero C, Boudou P** 2003 Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women, and children. Clin Chem 49:1381–1395
23. **Rodbard D, Lenox RH, Wray RL, Ramseth D** 1976 Statistical characterization of random errors in the radioimmunoassay dose-response variable. Clin Chem 22:350–358
24. **Dudley RA, Edwards P, Ekins RP, Finney DJ, McKenzie IG, Raab GM, Rodbard D, Rodgers RP** 1985 Guidelines for immunoassay processing. Clin Chem 31:1264–1271
25. **Guardabasso V, Rodbard D, Munson PJ** 1987 A dose-response analysis. Am J Physiol 252:E357–E364
26. **Draisci R, Palleschi L, Ferretti E, Lucentini L, Cammarata P** 2000 Quantitation of anabolic hormones and their metabolites in bovine serum and urine by liquid chromatography-tandem mass spectrometry. J Chromatogr A 870: 511–522
27. **Vermeulen A, Verdouck L, Kaufman JM** 1999 A critical evaluation of simple methods for the estimation of free testosterone in serum. J Clin Endocrinol Metab 84:3666–3672

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