

Testosterone Testing

Draft Report: Public Comment & Response

February 6, 2015

Health Technology Assessment Program (HTA) Washington State Health Care Authority PO Box 42712 Olympia, WA 98504-2712 (360) 725-5126 hca.wa.gov/hta shtap@hca.wa.gov



Testosterone Testing

Response to Public Comments on Draft Report

February 6, 2015

Prepared by:

Hayes, Inc. 157 S. Broad Street Suite 200 Lansdale, PA 19446 P: 215.855.0615 F: 215.855.5218

Response to Public Comments, Draft Report

Testosterone Testing

Hayes, Inc. is an independent vendor contracted to produce evidence assessment reports for the WA HTA program. For transparency, all comments received during the comments process are included in this response document.

Comments related to program decisions, processes, or other matters not pertaining to the evidence report are acknowledged through inclusion only. When comments cite evidence, the information is forwarded to the vendor for consideration in the evidence report.

This document responds to comments from the following parties:

- G. Steven Hammond, PhD, MD; Chief Medical Officer, Washington State Department of Corrections; comments presented on behalf of the Washington Agency Medical Directors
- Alvin M. Matsumoto, MD; Acting Head, Division of Gerontology and Geriatric Medicine, and Professor, Department of Medicine, University of Washington School of Medicine; Associate Director, Geriatric Research, Education and Clinical Center, and Director, Clinical Research Unit, VA Puget Sound Health Care System

Table 1 provides a summary of the comments with corresponding responses.

Table 1. Public Comments on Draft Report, Testosterone Testing

Comment and Source	Response
January 16, 2015 Letter – Dr. Hammond, WA Agency Medical Directors	
 The commenter highlights some of the issues surrounding testosterone testing and testosterone supplementation: "whether serum testosterone levels below the laboratory 'reference range' in adult men indicate a clinicopathological state that requires treatment." "whether an adult man with non-specific symptoms that could be associated with hypogonadism, who also has a serum testosterone level measured below a laboratory reference range, but without an otherwise well-defined hypogonadal condition, is likely to have improved health outcomes as a result of testosterone testing and consequent testosterone supplementation." Application of reference ranges derived from populations of young men to middle-aged and older men, who are "known to have continual declines in testosterone levels with increasing age." Mass media publicizing of the "low T" phenomenon. 	Thank you for your comments. No changes needed in the report.
The commenter expressed concern about equating the term androgen deficiency, which implies a pathological state, with <i>low serum</i> testosterone and suggested that this statement be included in the report: Low serum testosterone may indicate androgen deficiency. While low serum testosterone may suggest putative androgen deficiency, it must be correlated with additional clinicopathological signs to be diagnostic of hypogonadism.	Thank you for this comment. Any text suggesting that low testosterone level is equivalent to androgen deficiency as a pathological condition has been revised.
The commenter further advised that the report not equate <i>low serum testosterone</i> with <i>hypogonadism</i> in the absence of "clear signs of primary or secondary hypogonadism of known etiology."	

Comment and Source	Response
"The Hayes, Inc. technology assessment on testosterone testing includes a valuable review of the literature on current conceptions and medical practice around testosterone testing and testosterone supplementation in the setting of 'low' serum testosterone levels, but there is insufficient evidence to equate 'low serum testosterone,' even in the presence of non-specific symptoms characteristic of advancing age, with a clinicopathological condition of 'androgen deficiency.'"	Thank you for your comment. No changes needed in the report.
January 19, 2015 Email - Dr. Alvin Matsumoto, University of Washingto	n Medical School and VA Puget Sound Health Care System
"In the "Testosterone Testing – Draft Evidence Report," I am most concerned about the conclusion that testosterone therapy may be useful in improving blood sugar control (glucose and hemoglobin A1c) in men with type 2 diabetes mellitus and hypogonadism. The main evidence cited for this is the meta-analysis by Cai, et al. (2014). A more recent meta-analysis did not find improvement in glycemic control with testosterone treatment (Grossmann M, et al., Clin Endocrinol 2014 ePub, attached). The difficulty with interpreting studies that seek to determine the effect of testosterone therapy on glycemic control is the lack [of] control of changes in diabetes therapy independent of testosterone (oral hypoglycemic agents, insulin and insulin analogs, diet, exercise, weight loss). I am afraid that a conclusion that testosterone therapy improves glycemic control will lead to over- testing (screening without regard to clinical manifestations of androgen deficiency), misinterpretation of testing and over-diagnosis of hypogonadism (obese diabetics often have low total testosterone but normal free testosterone), and over-treatment with testosterone. In the absence of better data, I would eliminate this conclusion or at the very least temper it."	Thank you for calling the new systematic review by Grossman and colleagues to our attention. This review was published after the draft report was released. The findings of the meta-analysis by Grossman et al. (2014) have been added to the report and conclusions have been modified accordingly. The issue of bias due to changes in antidiabetic medications has also been addressed.
"I think that the variability reported testosterone measurements in various testosterone assays needs to be mentioned. As shown in Table	Thank you for these comments and for the reference. The results from the study by Wang et al. have been added to the
1 of the report by Wang C, et al., JCEM 89:534-543, 2004 (attached), the same quality control sample measured in different assays gave	section on <i>Analytic Validity</i> under CLINICAL BACKGROUND in the TECHNICAL REPORT and corresponding edits have been made in the

Comment and Source	Response
median testosterone levels that ranged from 215 ng/dL to 348 ng/dL. This situation has occurred because the emphasis in quality control programs has been on reproducibility within the same assay rather than accuracy of the measurement. The CDC program was initiated to provide an accuracy-based quality control for harmonization of assays (i.e., so that assay readings are more comparable); a similar program was needed to deal with the initially marked variability in cholesterol and hemoglobin A1c measurements."	Analytic Validity section of the EVIDENCE SUMMARY. Further clarification of the quality control programs offered by the CDC and the Clinical Association of Pathologists (CAP), including their voluntary nature, has been added. Additionally, a paragraph on threats to analytic validity has been added to the OVERALL SUMMARY AND DISCUSSION.
"I also think that more emphasis is needed regarding substantial variability in testosterone levels within an individual from day-to-day, up to 35%. Within individual variability was found in a study by Swerdloff RS, et al., JCEM 85:4500-4510, 2000 (page 4509, paragraph 4, attached); 30-35% of men who were found on screening to have a low testosterone < 300 ng/dL had normal average testosterone levels over a 24-hr pharmacokinetic blood sampling. Subsequently, Brambilla DJ, et al. (Clin Endocrinol 67:853-862, 2007, attached) quantified intra- individual variation in testosterone levels more formally. The bottom line is that one sample is not sufficient to assess testosterone status."	Thank you for these comments and for these references. Statistics regarding the intraindividual variability of test results within the day and between days were included in the <i>Analytic Validity</i> section of the TECHNICAL REPORT . These data have been added to the corresponding section in the EVIDENCE REVIEW . The references cited in the comments will be brought to the Health Technology Clinical Committee (HTCC) meeting.

January 16, 2015

Comments of the WA Agency Medical Directors Testosterone Testing Draft Technology Assessment Report

Presented by G. Steven Hammond, PhD, MD, Washington State Department of Corrections Chief Medical Officer

The Hayes, Inc. technology assessment highlights some of the quandaries surrounding appropriate clinical use of serum testosterone testing in adult men, as well as questions about the risks and benefits of treating men having "low" serum testosterone levels with testosterone supplementation.

The primary question raised is whether serum testosterone levels below the laboratory "reference range" in adult men indicate a clinicopathological state that requires treatment. As is noted in the report, there are a number of well-defined clinical conditions that cause hypogonadism, either primary or secondary, around which there is little controversy concerning testosterone testing or treatment with testosterone supplementation. The major question is whether an adult man with non-specific symptoms that could be associated with hypogonadism, who also has a serum testosterone level measured below a laboratory reference range, but without an otherwise well-defined hypogonadal condition, is likely to have improved health outcomes as a result of testosterone testing and consequent testosterone supplementation.

Defining a clinical condition principally on the basis of a laboratory test result, with no further etiologic diagnosis, is of questionable validity. Such practice is even more dubious when reference ranges for the lab test are statistically defined (mean + two standard deviations) in a population of young men, and are applied to middle-aged and older populations, who are known to have continual declines in testosterone levels with increasing age.

As noted in the report, there has been much publicizing of so-called "low T" in mass media, with suggestions for men to consult with their physicians about this. Such "public health" messaging often suggests, more or less overtly, that expected changes related to aging, such as decreased vigor and virility, may be related to a medical condition, i.e., "low T", with the implication that medical treatment (with testosterone) may be appropriate or even necessary.

As the report indicates, the health benefits, and safety and more so the necessity, of treating "low T" remain very much in doubt. "Low T" is not an accepted clinical diagnosis. There is not a clear case definition of "hypogonadism" associated with "below normal" serum testosterone levels and some array of symptomatology, in the absence of other findings supporting an etiologic diagnosis.

It is not warranted to equate a "low serum testosterone" with "androgen deficiency", as is done in the opening sentences of the technology assessment:

"Low serum testosterone is a form of androgen deficiency. In the present report, the term androgen deficiency can be interpreted to be equivalent to low serum testosterone."

Comments of the WA Agency Medical Directors January 16, 2015

As noted subsequently in the report, a definition of the term "low serum testosterone" is problematic, given age-related declines seen in male populations, and the many factors that affect serum testosterone and sex hormone binging globulin levels. The term androgen deficiency connotes a pathological state, whereas there is no clearly defined pathological state associated with "low testosterone" levels in aging men. It is mistaken and potentially misleading to state "the term androgen deficiency can be interpreted to be equivalent to low serum testosterone."

Under the circumstances it would be more accurate to say:

"Low serum testosterone may indicate androgen deficiency. While low serum testosterone may suggest putative androgen deficiency, it must be correlated with additional clinicopathological signs to be diagnostic of hypogonadism."

The report should eschew any equation of "low serum testosterone" with "androgen deficiency" or "hypogonadism" in clinical settings which do not include clear signs of primary or secondary hypogonadism of known etiology. Without such signs, any putative clinicopathological hypogonadal condition is hypothetical.

The Hayes, Inc. technology assessment on testosterone testing includes a valuable review of the literature on current conceptions and medical practice around testosterone testing and testosterone supplementation in the setting of "low" serum testosterone levels, but there is insufficient evidence to equate "low serum testosterone", even in the presence of non-specific symptoms characteristic of advancing age, with a clinicopathological condition of "androgen deficiency."

From: Matsumoto, Alvin M [mailto:Alvin.Matsumoto@va.gov]
Sent: Monday, January 19, 2015 6:44 PM
To: Teresa Rogstad
Cc: josh.morse@hca.wa.gov; Karen Crotty
Subject: RE: [EXTERNAL] Draft Report, WA HTA Program

Teresa, et al:

My specific comments:

- 1. I had few minor wording changes in the "Final Key Questions and Background Testosterone Testing" sheet (attached). I think that it is important to emphasize that most men male factor infertility have normal serum testosterone levels.
- 2. In the "Testosterone Testing Draft Evidence Report", I am most concerned about the conclusion that testosterone therapy may be useful in improving blood sugar control (glucose and hemoglobin A1c) in men with type 2 diabetes mellitus and hypogonadism. The main evidence cited for this is the meta-analysis by Cai, et al (2014). A more recent meta-analysis did not find improvement in glycemic control with testosterone treatment (Grossmann M, et al, Clin Endocrinol 2014 ePub, attached). The difficulty with interpreting studies that seek to determine the effect of testosterone therapy on glycemic control is the lack control of changes in diabetes therapy independent of testosterone (oral hypoglycemic agents, insulin and insulin analogs, diet, exercise, weight loss). I am afraid that a conclusion that testosterone therapy improves glycemic control will lead to over-testing (screening without regard to clinical manifestations of androgen deficiency), misinterpretation of testing and over-diagnosis of hypogonadism (obese diabetics often have low total testosterone but normal free testosterone), and over-treatment with testosterone. In the absence of better data, I would eliminate this conclusion or at the very least temper it.
- 3. I think that the variability reported testosterone measurements in various testosterone assays needs to be mentioned. As shown in Table 1 of the report by Wang C et al, JCEM 89:534-543, 2004 (attached), the same quality control sample measured in different assays gave median testosterone levels that ranged from 215 ng/dL to 348 ng/dL. This situation has occurred because the emphasis in quality control programs has been on reproducibility within the same assay rather than accuracy of the measurement. The CDC program was initiated to provide an accuracy-based quality control for harmonization of assays (i.e. so that assay readings are more comparable); a similar program was needed to deal with the initially marked variability in cholesterol and hemoglobin A1c measurements.
- 4. I also think that more emphasis is needed regarding substantial variability in testosterone levels within an individual from day-to-day, up to 35%. Within individual variability was found in a study by Swerdloff RS, et al, JCEM 85:4500-4510, 2000 (page 4509, paragraph 4, attached); 30-35% of men who were found on screening to have a low testosterone < 300 ng/dL had normal average testosterone levels over a 24-hr pharmacokinetic blood sampling. Subsequently, Brambilla DJ, et al (Clin Endocrinol 67:853-862, 2007, attached) quantified intra-individual variation in testosterone levels more formally. The bottom line is that one sample is not sufficient to assess testosterone status.</p>

My only general comment is that the "Testosterone Testing – Draft Evidence Report" was somewhat redundant and long, but was quite detailed and provided a good summary of most of the evidence-base on clinical testosterone testing.

I found the Washington State Agency Utilization and Costs interesting and found myself asking whether guidelines (such as Endocrine Society guidelines) for testosterone testing and treatment were followed in Washington, i.e. the appropriateness of utilization and costs.

I do not have the confirmed date, location and time of the public meeting in March (initially said to be on March 20th at the SeaTac Conference Center, unclear what time) or information regarding what is expected of me at the meeting. My calendar is pretty full already in March and is dynamically changing with time. So, I would appreciate more specific details as soon as possible.

Thanks,

Al

Alvin M. Matsumoto, M.D. Acting Head, Division of Gerontology and Geriatric Medicine Professor, Department of Medicine University of Washington School of Medicine

Associate Director, Geriatric Research, Education and Clinical Center Director, Clinical Research Unit V.A. Puget Sound Health Care System 1660 S. Columbian Way (S-182-GRECC) Seattle, WA 98108-1597

Phone:	206-764-2308
FAX:	206-764-2569

Intraindividual variation in levels of serum testosterone and other reproductive and adrenal hormones in men

Donald J. Brambilla*, Amy B. O'Donnell*, Alvin M. Matsumoto† and John B. McKinlay*

*New England Research Institutes, Watertown, Massachusetts, †Department of Medicine, University of Washington School of Medicine, and Geriatric Research, Education and Clinical Center, VA Puget Sound Health Care System, Seattle, USA

Summary

Background Estimates of intraindividual variation in hormone levels provide the basis for interpreting hormone measurements clinically and for developing eligibility criteria for trials of hormone replacement therapy. However, reliable systematic estimates of such variation are lacking.

Objective To estimate intraindividual variation of serum total, free and bioavailable testosterone (T), dihydrotestosterone (DHT), SHBG, LH, dehydroepiandrosterone (DHEA), dehydroepiandrosterone sulphate (DHEAS), oestrone, oestradiol and cortisol, and the contributions of biological and assay variation to the total.

Design Paired blood samples were obtained 1-3 days apart at entry and again 3 months and 6 months later (maximum six samples per subject). Each sample consisted of a pool of equal aliquots of two blood draws 20 min apart.

Study participants Men aged 30–79 years were randomly selected from the respondents to the Boston Area Community Health Survey, a study of the health of the general population of Boston, MA, USA. Analysis was based on 132 men, including 121 who completed all six visits, 8 who completed the first two visits and 3 who completed the first four visits.

Measurements Day-to-day and 3-month (long-term) intraindividual standard deviations, after transforming measurements to logarithms to eliminate the contribution of hormone level to intraindividual variation.

Results Biological variation generally accounted for more of total intraindividual variation than did assay variation. Day-to-day biological variation accounted for more of the total than did long-term biological variation. Short-term variability was greater in hormones with pulsatile secretion (e.g. LH) than those that exhibit less ultradian variation. Depending on the hormone, the intraindividual standard deviations imply that a clinician can expect to see a difference exceeding 18–28% about half the time when two measurements are made on a subject. The difference will exceed 27–54% about a quarter of the time.

Conclusions Given the level of intraindividual variability in hormone levels found in this study, one sample is generally not sufficient to characterize an individual's hormone levels but collecting more than three is probably not warranted. This is true for clinical measurements and for hormone measurements used to determine eligibility for a clinical trial of hormone replacement therapy.

(Received 22 December 2006; returned for revision 11 February 2007; finally revised 7 June 2007; accepted 7 June 2007)

Introduction

Estimates of intraindividual variation in hormone levels provide the foundation for interpreting hormone measurements, such as the reliability of one or two values as estimates of an individual's average hormone concentration, for both the clinician and the researcher. For present purposes, intraindividual variation is defined as variation around an individual's steady-state mean hormone level rather than changes in the mean itself. The steady-state mean is that individual's current average state. Systematic variation, such as the well-known changes in testosterone and other hormones with age or the relatively rapid changes that are associated with onset of certain diseases or initiation of certain medications, constitutes changes in the steady state mean.

The number of blood samples required to adequately characterize an individual's steady state hormone level increases as intraindividual variation increases. In the absence of information on this variation, it is also difficult to determine whether a difference between hormone levels on two occasions constitutes simply a fluctuation around the steady state mean or a change in the mean. The researcher has difficulty performing sample size calculations for trials in which change in hormone level is the outcome because intraindividual variation is usually the denominator of the test statistic used to compare average changes in hormone levels between treatment groups. Moreover, the number of samples required to determine eligibility, when eligibility depends on hormone level, is unknown.

Interindividual variation in the levels of testosterone and other hormones in men has received considerable attention.^{1–8} Intraindividual variation has also been examined,^{9–17} but sample sizes in previous studies were generally small and none of the studies provided estimates of intraindividual variation that would form the

Correspondence: Donald J. Brambilla, New England Research Institutes, 9 Galen Street, Watertown, Massachusetts, MA 02472, USA. Tel: +1 617-923-7747; Fax: +1 617-926-8246; E-mail: dbrambilla@neriscience.com

quantitative basis for interpreting hormone measurements. Therefore, a prospective study of variation in the levels of total, free and bioavailable testosterone (T), dihydrotestosterone (DHT), SHBG, LH, dehydroepiandrosterone (DHEA), dehydroepiandrosterone sulphate (DHEAS), oestrone, oestradiol and cortisol in men was initiated in Boston, Massachusetts, USA in May 2004. In this paper, we present estimates of day-to-day and 3-month variation in these hormones.

Methods

Subjects

Subjects for the study were selected from among the 2301 male respondents to the Boston Area Community Health (BACH) Survey.¹⁸ Subjects for the BACH survey were randomly selected from among the residents of Boston, MA, USA who were aged 30-79 years, using a weighted sampling scheme to recruit approximately equal numbers of Hispanic Americans, non-Hispanic African Americans and non-Hispanic Caucasians, and approximately equal numbers by decade of age.

For the present study, BACH Survey respondents were stratified into three categories of race/ethnicity and decade of age (30-39, 40-49, 50-59, 60-69, 70+ years) to produce 15 strata. Subjects were randomly selected from the male respondents in each stratum with the goal of obtaining approximately the same number of subjects in every stratum. A potential subject who refused or was found to be ineligible was replaced with another randomly selected from the same stratum.

Men were excluded if they (1) had a history of hypogonadism of known cause, such as treatment for prostate cancer, Klinefelter syndrome, Kallmann syndrome and orchidectomy; (2) were using any medications that alter hormone levels, either as the intended effect or as a side effect; (3) had cirrhosis, liver cancer, other severe liver disease, or kidney disease requiring dialysis; or (4) had a problem with blood draws, such as haemophilia, or a compromised immune system caused by HIV/AIDS, chemotherapy, radiation or other conditions. Excluded medications included anabolic steroids, androstenedione, bicalutamide (Casodex®, AstraZeneca, Wilmington, DE), cimetidine, DHEA, diethylstilbestrol, other oestrogens, dutasteride (Avodart[®], GlaxoSmithKline, Research Triangle Park, NC), finasteride (Proscar®, Merck, Whitehouse Station, NJ), glucocorticoids (prednisone, cortisone, hydrocortisone and decadron), ketoconazole, megestrol acetate (Megace®, Bristol-Myers Squibb, Princeton, NJ), opiates (morphine, codeine, oxycodone, hydrocodone, etc.), spironolactone, testosterone or any androgen, flutamide (Eulexin®, Schering-Plough, Kenilworth, NJ) and other medications for prostate cancer. Eligibility was determined from responses to the BACH survey and responses to questions at screening for this study.

Subjects were enrolled after written informed consent was obtained. The consent form, protocol, telephone scripts and contact documents were approved by the Institutional Review Board of the New England Research Institutes.

Sampling methods

Paired blood samples were obtained 1-3 days apart (median 2 days) at study entry and again 3 months and 6 months later, producing a maximum of six blood samples per subject. With this nested design, the study provided estimates of both day-to-day and 3-month variation in hormone levels. Most study visits took place in the subject's home but, if the subject so requested, they took place elsewhere, usually at his place of employment or at study headquarters at the New England Research Institutes.

At each visit, two blood samples were drawn 20 min apart to reduce the effects of pulsatile secretion on hormone levels. Blood was drawn within 4 h of the subject's awakening to control for diurnal variation.¹⁹⁻²¹ Sampling was postponed to another day if the time of awakening on a given day departed substantially from a subject's normal pattern. The two samples were placed in an ice-filled cooler for transport back to study headquarters, where they were centrifuged, equal aliquots were pooled using a calibrated automatic pipette, and the pooled samples were stored in scintillation vials at -70 °C. Samples were transferred to the endocrine laboratory of the Department of Physiology at the University of Massachusetts (UMASS) Medical School, Worcester, MA, where the assays were performed.

At the first, third and fifth study visits, a brief questionnaire was administered to identify changes in health or behaviour that might affect variation in hormone levels. Subjects who had started taking medications or had developed conditions that would have made them ineligible at the start were excluded from further participation.

Hormone assays

The methods that were used for the various assays are listed in Table 1. Free T and bioavailable T were calculated from the Sodergard equation using both a constant albumin concentration of 4.3 g/dl and measured albumin concentration.^{22,23} All samples obtained from a subject were assayed in the same run for each hormone. Differences among measurements within subjects are thus free of interassay variation. Interassay variation was estimated as described below.

Statistical methods

Descriptive statistics for interindividual differences in hormone levels were based on hormone levels at the first study visit.

Intraindividual variation was characterized using intraindividual standard deviations. Prior to analysis, hormone measurements were transformed to base 10 logarithms to eliminate a positive correlation between the standard deviation and mean hormone level that rendered application of estimates of intraindividual variation to clinical data extremely difficult. The standard deviations of the transformed data are applicable to a broad range of hormone levels. Interpretation of the estimates of intraindividual variation for untransformed hormone measurements is provided.

Intrasubject standard deviations were calculated under four sampling schemes:

- 1 Samples collected 1–3 days apart and assayed in the same run:
- $\sigma_1 = \sqrt{\sigma_D^2 + \sigma_A^2}$ 2 Samples collected 1–3 days apart, with each one assayed in a separate run: $\sigma_2 = \sqrt{\sigma_D^2 + \sigma_E^2 + \sigma_A^2}$
- **3** Samples collected 3 months apart and assayed in the same run: $\sigma_3 = \sqrt{\sigma_D^2 + \sigma_L^2 + \sigma_A^2}$

Hormone	Method	Kit or reference	Intra-assay CV	Interassay CV
Т	RIA	Coat-A-Count,	4.3	9.8
		Diagnostic Product		
		Corporation,		
D		Los Angeles, CA		
Free I	Calculated	17.18		
Bioavailable T	Calculated			
DHT	RIA	In-house assay	2.1	3.1
SHBG	CLIA	Immulite, Diagnostic	3.1	4.1
		Product Corporation,		
		Los Angeles, CA		
LH	CLIA	Immulite, Diagnostic	4.2	5.5
		Product Corporation,		
		Los Angeles, CA		
DHEA	RIA	Diagnostic Systems	2.6	5.6
		Laboratories, Webster, TX		
DHEAS	RIA	Coat-A-Count, Diagnostic	2.5	5.2
		Product Corporation,		
		Los Angeles, CA		
Oestrone (F1)	RIA	Diagnostic Systems	1.0	4.2
Oestione (E1)	NITY	Laboratories Webster TX	10	12
Opertradial (E2)	DIA	Diagnostic Systems	2.2	7.4
Oestradior (E2)	KIA	Laboratorios Webster TV	2.3	7.4
Continul	DIA	Laboratories, webster, 1X	2.4	6.4
Cortisol	KIA	Coat-A-Count, Diagnostic	5.4	6.4
		Product Corporation,		
		Los Angeles, CA		
Albumin	Colourimetric determination with Bromcresol Purple	Beckman Coulter, Brea, CA	N/A	

Table 1. Methods used to measure hormone levels

T, testosterone; DHT, dihydrotestosterone; DHEA, dehydroepiandrosterone; DHEAS, dehydroepiandrosterone sulphate; RIA, radioimmunoassay; CLIA, chemiluminescent immunoassay; N/A, not available.

4 Samples collected 3 months apart, with each one assayed in a separate run: $\sigma_4 = \sqrt{\sigma_D^2 + \sigma_L^2 + \sigma_E^2 + \sigma_A^2}$

In these equations, σ_D^2 is the variance component attributable to day-to-day variation, σ_D^2 is the component attributable to 3-month variation, σ_A^2 is the intra-assay, or within-run, component of variance and σ_E^2 is the interassay, or between-run, component of variance. The so-called batch testing in Schemes 1 and 3 is often used in clinical trials because interassay variation is excluded, which improves statistical power to detect changes in hormone levels. Schemes 2 and 4 are more relevant to clinical testing.

The day-to-day and 3-month intraindividual components of variance were estimated using a nested linear model for each hormone with subject and month within subject (baseline, 3 months, 6 months) as the predictors.²⁴ Month within subject was treated as a random effect. The error mean square provided an estimate of the sum of the day-to-day and intra-assay components, $\sigma_D^2 + \sigma_A^2$. The 3-month component, σ_L^2 , was estimated using the equation for the expected mean square for month within subject.

Intra- and interassay variation, σ_A^2 and σ_E^2 , were estimated from results from the assay controls in the nine runs per hormone in which samples were assayed in this study, using a linear model with run treated as a random effect and control (e.g. high, medium and low) treated as a fixed effect.²⁵ Each assay of cortisol, DHEAS, DHT, free T, LH and T included three controls (high, medium, low) in duplicate, while each assay of DHEA, oestrone, oestradiol and SHBG included two controls (high and low) in triplicate. The intra-assay component was estimated from the error mean square, while the interassay component was estimated from the expected mean square for the predictor run. Components of assay variation for calculated free T and calculated bioavailable T, using an albumin concentration of 4.3 g/dl, were estimated by Monte Carlo simulation using the components of assay variation of T and SHBG. Similar calculations using measured albumin concentration were not performed because data from the albumin controls were not obtained.

Estimates of the intraindividual standard deviations of calculated free and bioavailable T under Schemes 2 and 4, with albumin fixed at 4·3 g/dl, were obtained by Monte Carlo simulation. Random variates, sampled from two normal distributions, with means of zero and variances equal to the interassay components of variance for T and SHBG, respectively, were added to the observed T and SHBG values. Free T and bioavailable T were then calculated from the Sodergard^{22,23} equations as usual and the variances of log-transformed values were obtained from the nested linear model described above. The intraindividual standard deviations were estimated from the medians of the frequency distributions of the variances under Schemes 2 and 4. Estimates of the intraindividual standard deviations

of free T and bioavailable T under Schemes 2 and 4, using measured albumin concentration, were not obtained in the absence of data from the albumin controls.

Approximate 95% confidence limits for the standard deviation of each hormone under each sampling scheme were derived from the frequency distribution of 10 000 bootstrap²⁶ estimates of the standard deviation. For Schemes 2 and 4, which included interassay variation, a separate random variate, sampled from a normal distribution with mean zero and variance equal to the estimated interassay component of variance for that hormone, was added to each observed hormone measurement before bootstrapping took place.

To provide an interpretation of the standard deviations with some clinical relevance, the percentage difference between hormone levels in two samples from the same subject that would be exceeded 50% of the time was calculated under sampling Schemes 2 and 4. If hormone levels are normally distributed around an individual's steady state mean, then the percentage difference between two hormone measurements from a subject, calculated as $100 \times (x_2 - x_1)/x_1$, where x_2 is the larger and x_1 the smaller measurement, should be > $100 \times (10^{x_{055}\sqrt{2}} - 1)$ 50% of the time, where *s* is the intraindividual standard deviation and $Z_{0.75}$ is the standard normal deviate for a cumulative probability of 0.75. The difference that would be exceeded 25% of the time was calculated similarly using $Z_{0.875}$.

The extent to which repeated sampling improves the precision of estimated mean hormone level for an individual is another important aspect of intraindividual variation. Therefore, 95% confidence limits for the steady state mean were calculated for one measurement, the average of two and the average of three measurements. The standard deviations under sampling Scheme 4 were used for the calculations. If \log_{10} -transformed measurements are normally distributed, then the 95% confidence limits around the mean, M, of *n* measurements, after transforming the confidence limits back to the raw scale, are $M/10^{196s/\sqrt{n}}$ and $M \times 10^{196s/\sqrt{n}}$. Each confidence limit in the average measurement. Barring extreme fluctuations in hormone level, the bias induced by averaging *n* measurements, rather than averaging the log-transformed values and transforming the mean of the logs back to the raw scale, is small enough to be ignored for present purposes.

Systematic differences in the intraindividual standard deviation among categories of age or race/ethnicity would compromise the generalizability of the results of this study. Statistical tests for such differences were performed by calculating intraindividual standard deviations under sampling Schemes 1 and 3 for each hormone in each subject who completed all six visits and treating the standard deviations as outcome variables in an analysis of variance with age category and race/ethnicity as predictors.

In order to determine if intraindividual variation measured over 6 months might be contaminated with systematic changes, such as those associated with ageing, log-transformed hormone level was modelled as a linear function of time on study using a mixed linear model with subject treated as a random effect.²⁵

Results

Letters inviting participation were sent to 230 men: 22 refused to participate; 43 were ineligible or too ill for screening; 31 could not

Table 2. Baseline characteristics of the entire cohort

	Mean ± SD
Height	$174.7 \pm 7.4 \text{ cm}$
Weight	85·5 ± 17·8 kg
BMI	$28.0 \pm 5.4 \text{ mg/m}^2$
Waist to hip ratio	0.94 ± 0.07
Smokers	30.0%
Recreational drug use	6.0%
Depressive symptoms	11.9%
Alcohol use (average drinks per day):	
None	36.4%
> 0, < 1	35.3%
$\geq 1, < 3$	21.6%
≥3	6.7%

BMI, body mass index.

be contacted; and 134 agreed to participate. Eight men completed only the first two visits and three men completed only the first four visits. Of the 11 men with incomplete data, 6 men were dropped from the study after they began taking medications or developed medical conditions that altered hormone levels, 4 men withdrew voluntarily or were lost to follow-up and 1 man died. Two subjects, who were determined to be ineligible for the study after completing all six visits, were excluded from the analysis. All available hormone measurements from the remaining 132 subjects were used to obtain the standard deviations reported here.

The cohort included 43 Caucasians, 46 African Americans and 45 Hispanic Americans. There were 20 respondents in the first age category (30-39 years) and 28 or 29 respondents in each of the other four age categories. Other baseline characteristics of the participants are provided in Table 2. The hormone levels in the participants in the study span broad ranges of values (Table 3). Some of the extreme values in the table, such as the minimum for total T and the maxima for LH and SHBG, may indicate various disease states. As health status and medication use were obtained by self-report, the possibility that some subjects were taking unreported medications or had unreported conditions affecting hormone levels cannot be ruled out. However, the majority of values are within normal ranges, so underreporting of diseases or medications was uncommon.

Median time from awakening to first blood draw was 1·67 h (5th and 95th percentiles: 0·42 h and 3·12 h). Most (92·5%) of the 760 blood draws took place between 06·00 h and noon but 37 samples were drawn between 16·30 h and 18·00 h, and the remaining 20 samples were drawn between 12·00 h and 16·25 h. Draw times were fairly tightly clustered within subjects, although they varied considerably among subjects because of work schedules, among other factors.

Tests for systematic changes in hormone levels during the study indicated that mean 6-month change was < 4% for all hormones and < 2% for half of the hormones. Total T had the smallest mean change at 0.8%. All of these changes were quite small compared to the intraindividual standard deviations described below. For more than half of the hormones, the direction of the change indicated by the

Tabl	le 3.	Descriptive	statistics	for	hormone	leve	ls at t	he i	first	stud	y v	isit
------	-------	-------------	------------	-----	---------	------	---------	------	-------	------	-----	------

		Percentile					
Hormone	Minutes	5th	25th	50th	75th	95th	Maximum
Total T (nmol/l)	3.5	6.7	11.4	15.2	18.9	26.9	38.8
Free T (pmol/l)*	48.5	152.6	228.8	305.1	377.9	523.5	755.8
Free T (pmol/l)†	48.5	159.5	228.8	301.6	391.8	530.5	828.6
Bioavailable T (nmol/l)*	1.1	3.6	5.4	7.1	8.8	12.2	17.8
Bioavailable T (nmol/l)†	1.1	3.7	5.4	7	9.2	12.4	19.4
Albumin (g/dl)	3	3.4	3.9	4.1	4.4	4.8	5.2
Cortisol (nmol/l)	165.5	202	292.4	358.7	458	579.4	847
DHEA (nmol/l)	4.3	7	11.6	16.7	23.2	44	64.5
DHEAS (umol/l)	0.35	0.89	1.76	3.08	5.16	11.23	15.61
DHT (nmol/l)	0.34	0.55	0.76	0.96	1.24	1.65	2.03
Oestrone (pmol/l)	37	77.7	103.6	129.5	155.3	210.8	310.7
Oestradiol (pmol/l)	37	62.9	96.2	122	151.6	196	273.7
LH (mIU/ml) (76 subjects, 30–59 years)	1.26	2.52	3.77	4.96	6.53	11.1	17.0
LH (mIU/ml) (56 subjects, 60-79 years)	1.04	1.81	4.26	6.13	8.23	11.8	24.9
SHBG (nmol/l)	7.59	16.8	25.8	36.9	51	79.6	118

*From the Sodergard equations assuming albumin at 4.3 g/dl.

†From the Sodergard equations using measured albumin.

T, testosterone; DHT, dihydrotestosterone; DHEA, dehydroepiandrosterone; DHEAS, dehydroepiandrosterone sulphate.

fitted parameter disagreed with the age-related changes reported previously.^{2,6–8} Thus, the differences over time were more likely noise than systematic variation.

Using a critical P-value of 0.05, only 7 of 56 tests for variation of the short-term and long-term intraindividual standard deviations with age category and race/ethnicity produced statistically significant results. All but two of the other 49 tests had P > 0.20. Linear models that included both predictors accounted for no more than 10% of the variation $(r^2 \le 0.10)$ in the standard deviation for all hormones considered. The short-term and long-term standard deviations for oestradiol varied with race/ethnicity (short-term: P = 0.0178; longterm: P = 0.0139). Non-Hispanic African Americans had the highest average standard deviations, while Hispanic Americans had the lowest. However, the differences between mean standard deviations were small enough that a single estimate was reasonably representative of all three groups. The short-term intraindividual standard deviations of cortisol (P = 0.0366), DHEA (P = 0.0380) and LH (P = 0.0097) varied with age, as did the long-term standard deviations of DHT (P = 0.0381) and LH (P = 0.0320). For cortisol, DHEA and DHT, plots of intraindividual standard deviations against age failed to demonstrate any coherent pattern that could be interpreted as systematic changes in the level of variation with age. In all three cases, the highest mean standard deviation was found in the middle age category (50-59 years) and in each case, the elevated mean was attributable to a small number of subjects with elevated variation, rather than an overall difference between age groups. A single estimate of the short-term and long-term standard deviation for each hormone was deemed reasonably representative of the group as a whole.

On the other hand, the mean short-term and long-term standard deviations of LH were higher in the first three age groups (30-59 years) than in the last two age groups (60-79 years).

There was no clear evidence of an age-related trend within either the first three or the last two age groups. Therefore, standard deviations for LH were calculated separately for men \leq 59 years and those \geq 60 years.

Estimates of σ_A^2 , σ_E^2 , $\sigma_D^2 + \sigma_A^2$ and σ_L^2 are provided in Table 4 and intraindividual standard deviations based on these estimates are provided in Table 5 with bootstrap 95% confidence limits. Units are not specified in these tables because the variance and standard deviation of log-transformed values do not depend on the units in which a concentration is expressed. The 3-month standard deviations were only 12–60% (median 26%) larger than the day-to-day standard deviations, indicating that intraindividual variation measured over a relatively short interval of, for instance, a few days, would capture more than half the total variation that would be seen over a few months. This follows from the relatively small size of the 3-month component, σ_t^2 , in Table 4.

Setting aside albumin and SHBG, cortisol, DHEA and LH had the largest intraindividual standard deviations in the study, whereas oestradiol, DHT, total T and oestrone had the smallest. The differences in intraindividual variation between these two groups of hormones were caused mainly by differences in biological variation, not assay variation, as is clear from the components of variance in Table 4. The intraindividual standard deviations for free T and bioavailable T by the Sodergard equation were expected to be larger than the standard deviations for total T because the variation for the two fractions includes variation attributable to both total T and SHBG. The differences, however, are rather small, reflecting the small standard deviations for SHBG.

The percentage difference between two hormone measurements on the same subject that would be exceeded 25% or 50% of the time, when the measurements were made a few days or a few months apart,

858 *D. J. Brambilla* et al.

Hormone	σ_A^2	σ_E^2	$\sigma_D^2 + \sigma_A^2$	σ_L^2
Total T	1.767×10^{-3}	2.507×10^{-3}	4.313×10^{-3}	1.976×10^{-3}
Free T*	3.730×10^{-3}	2.630×10^{-3}	5.433×10^{-3}	1.774×10^{-3}
Free T†	_	_	5.600×10^{-3}	1.894×10^{-3}
Bioavailable T*	3.730×10^{-3}	2.630×10^{-3}	5.433×10^{-3}	1.774×10^{-3}
Bioavailable T†	_	_	5.600×10^{-3}	1.894×10^{-3}
Albumin	_	_	5.777×10^{-4}	1.361×10^{-4}
Cortisol	1.122×10^{-3}	1.069×10^{-3}	8.578×10^{-3}	3.451×10^{-3}
DHEA	3.438×10^{-4}	3.249×10^{-4}	8.766×10^{-3}	3.568×10^{-3}
DHEAS	7.226×10^{-4}	2.996×10^{-4}	6.253×10^{-3}	1.564×10^{-3}
DHT	1.238×10^{-3}	2.842×10^{-4}	3.711×10^{-3}	1.917×10^{-3}
Estrone	3.723×10^{-3}	8.911×10^{-4}	3.605×10^{-3}	1.543×10^{-3}
Estradiol	1.431×10^{-3}	3.342×10^{-4}	4.715×10^{-3}	1.588×10^{-3}
LH (76 subjects, 30–59 years)	8.833×10^{-4}	7.488×10^{-5}	1.212×10^{-2}	5.447×10^{-4}
LH (56 subjects, 60–79 years)	8.833×10^{-4}	7.488×10^{-5}	$6 \cdot 105 \times 10^{-3}$	2.095×10^{-3}
SHBG	$5 \cdot 117 \times 10^{-4}$	$3{\cdot}101\times10^{-4}$	1.306×10^{-3}	2.049×10^{-3}

Table 4. Estimates of the variance components for log-transformed hormone measurements

*From the Sodergard equations assuming albumin at 4.3 g/dl.

†From the Sodergard equations using measured albumin.

 σ_A^2 : Intra-assay variance

 σ_E^2 : Interassay variance

 $\sigma_D^2 + \sigma_A^2$: Day-to-day variance + intra-assay variance

 σ_L^2 : Long-term or 3-month variance

T, testosterone; DHT, dihydrotestosterone; DHEA, dehydroepiandrosterone;

DHEAS, dehydroepiandrosterone sulphate.

Table 5. Estimates of the interindividual standard deviation at Visit 1 and the intraindividual standard deviation of \log_{10} -transformed hormone concentration for four study designs, using data from all visits, with 95% bootstrap confidence limits in parentheses

		Sampling design			
	Interindividual standard deviation	1	2	3	4
Total T	0.182	0.066 (0.057, 0.076)	0.083 (0.076, 0.093)	0.079 (0.070, 0.090)	0.094 (0.086, 0.104)
Free T*	0.170	0.074 (0.065, 0.084)	0.094 (0.084, 0.101)	0.085 (0.076, 0.096)	0.104 (0.093, 0.111)
Free T†	0.171	0.075(0.066, 0.085)	-	0.087 (0.078, 0.098)	-
Bioavailable T*	0.170	0.074 (0.065, 0.084)	0.094 (0.079, 0.097)	0.085 (0.076, 0.096)	0.104 (0.092, 0.111)
Bioavailable T†	0.171	0.075 (0.066, 0.085)	-	0.087 (0.078, 0.098)	_
Albumin	0.048	0.024 (0.021, 0.027)	-	0.027 (0.024, 0.030)	_
Cortisol	0.141	0.093 (0.082, 0.103)	0.098 (0.089, 0.110)	0.110 (0.098, 0.124)	0.114 (0.101, 0.127)
DHEA	0.236	0.094 (0.083, 0.101)	0.095 (0.084, 0.101)	0.111 (0.101, 0.120)	0.113 (0.102, 0.121)
DHEAS	0.335	0.079 (0.064, 0.095)	0.081 (0.066, 0.096)	0.088 (0.077, 0.101)	0.090 (0.079, 0.102)
DHT	0.145	0.061 (0.054, 0.069)	0.063 (0.056, 0.071)	0.075 (0.065, 0.086)	0.077 (0.067, 0.090)
Oestrone	0.157	0.060 (0.055, 0.066)	0.067 (0.066, 0.096)	0.072 (0.065, 0.080)	0.078 (0.079, 0.102)
Oestradiol	0.138	0.069 (0.063, 0.075)	0.071 (0.066, 0.078)	0.079 (0.072, 0.088)	0.082 (0.074, 0.090)
LH (subjects 30-59 years)	0.200	0.110 (0.095, 0.125)	0.111 (0.096, 0.126)	0.112 (0.100, 0.126)	0.113 (0.100, 0.126)
LH (subjects 60-79 years)	0.236	0.078 (0.068, 0.089)	0.079 (0.069, 0.090)	0.090 (0.080, 0.101)	0.091 (0.081, 0.102)
SHBG	0.211	0.036 (0.028, 0.045)	0.040 (0.034, 0.047)	0.058 (0.051, 0.065)	0.061 (0.054, 0.067)

*From the Sodergard equations assuming albumin at 4.3 g/dl.

†From the Sodergard equations using measured albumin.

Design 1: a short-term study with batch testing.

Design 2: a short-term study with samples tested when collected.

Design 3: a six-month study with batch testing.

Design 4: a six-month study with samples tested when collected.

Table 6. 95% confidence limits for steady state mean hormone level, expressed as multipliers of the result of 1 measurement or the average of 2 or 3 measurements

Hormone	l measurement	Mean of 2 measurements	Mean of 3 measurements
Total T	0.65, 1.53	0.74, 1.35	0.78, 1.28
Free T*	0.63, 1.60	0.72, 1.39	0.76, 1.31
Bioavailable T*	0.63, 1.60	0.72, 1.39	0.76, 1.31
Cortisol	0.60, 1.68	0.69, 1.44	0.74, 1.35
DHEA	0.60, 1.66	0.70, 1.43	0.75, 1.34
DHEAS	0.67, 1.50	0.75, 1.33	0.79, 1.26
DHT	0.71, 1.41	0.78, 1.28	0.82, 1.22
Oestrone	0.70, 1.42	0.78, 1.28	0.82, 1.22
Oestradiol	0.69, 1.44	0.77, 1.30	0.81, 1.24
LH	0.63, 1.60	0.72, 1.39	0.76, 1.31
SHBG	0.76, 1.31	0.82, 1.21	0.85, 1.17

*From the Sodergard equations assuming albumin at 4.3 g/dl.

T, testosterone; DHT, dihydrotestosterone; DHEA, dehydroepiandrosterone; DHEAS, dehydroepiandrosterone sulphate.

were calculated to aid in interpreting the standard deviations. For DHT, oestrone and oestradiol, the difference between two hormone measurements should exceed 15–17% or 18–20% half the time, when the measurements are made a few days or a few months apart, respectively. At the larger standard deviations that characterize free T, bioavailable T, cortisol and DHEA, differences should exceed 22–24% or 25–28% half the time, for measurements made a few days or a few months apart, respectively. Percentage differences that would be exceeded 25% of the time ranged from 27% to 31%, for DHT, oestrone and oestradiol measured a few days apart, to 46–54% for free T, bioavailable T, cortisol and DHEA measured a few months apart.

To clarify these calculations, consider two measurements of free T made a few months apart and suppose that one of the two values was 175 pmol/l. The probability that the other value was < 140 pmol/l or > 220 pmol/l is 0.50 and the probability that it was < 120 pmol/l or > 238 pmol/l is 0.25. If one of the two values was 300 pmol/l, then the probability that the other measurement was < 240 pmol/l or > 380 pmol/l is 0.50 and the probability that it was < 205 pmol/l or > 440 pmol/l is 0.25.

Another way to assess the results is to consider the impact of intraindividual variation on a diagnosis of abnormally high or low hormone levels. Consider, for example, total T and suppose that values < 8.67 nmol/l (250 ng/dl) are considered possibly indicative of hypogonadism. Of 121 subjects who completed all six visits, 15 had total T < 8.67 nmol/l at the first visit but only 6 of these 15 had average values < 8.67 nmol/l over all six visits. This outcome probably reflects the regression to the mean that can occur when subjects are selected on the basis of values that are on one side of a specified threshold. Of the 15 subjects, 3 had average values > 10.40 nmol/l (300 ng/dl) which many clinicians would consider to be within the normal range for young men. Reducing the threshold to 6.93 nmol/l (200 ng/dl) does not eliminate the problem. Of 7 men with total T < 6.93 nmol/l at Visit 1, 3 had average values over six visits that

were > 6.93 nmol/l. One average was between 6.93 nmol/l and 8.67 nmol/l, one was between 8.67 nmol/l and 10.40 nmol/l and the third was > 10.40 nmol/l. These counts do not include the two men who were excluded after it was determined that they were not eligible for the study. In both cases, T on Visit 1 and average T were < 6.93 nmol/l. On the other hand, 5 of 10 subjects with average values < 8.67 nmol/l over the first two visits had average values > 8.67 nmol/l over all six visits but none had average values > 10.40 nmol/l. Thus, some improvement in diagnostic accuracy can be obtained by averaging values from two blood samples.

The 95% confidence limits for the steady state mean, based on one measurement, the average of two and the average of three, are provided in Table 6. The values in the table are the multipliers, $10^{\pm 196s/\sqrt{n}}$, that were defined earlier for the measured value or average of two or three measured values. For example, if total T is measured once, then the confidence limits are 65% and 153% of the measured value. The values in the table demonstrate the gain in precision that results from collecting more than one sample from a subject. The confidence interval based on the average of two measurements of total T is approximately 30% narrower than the width of the interval around a single measurement, while the interval around the average of three measurements is 43% narrower than the width around a single measurement.

As an example of the gain in precision with repeated testing, suppose that total T from a subject, based on one measurement or the average of two or three measurements, is at the 5th percentile in Table 3 (6.7 nmol/l). Using the multipliers in Table 6, the 95% confidence interval for the steady state mean is $4\cdot39-10\cdot23$ nmol/l, based on one measurement, $4\cdot97-9\cdot04$ nmol/l, based on the average of two measurements, and $5\cdot25-8\cdot55$ nmol/l, based on the average of three measurements.

The extent to which batch testing reduces intraindividual variation by eliminating interassay variation can be determined by comparing the standard deviations for Schemes 1 and 2 or those for Schemes 3 and 4 in Table 5. For LH, DHEA, DHEAS, cortisol and DHT, batch testing reduced the intraindividual standard deviation by < 5%, indicating that interassay variation makes only a small contribution to total variation when samples from a subject are assayed separately. For total T and the fractions, batch testing produced reductions of 16-21% in the intraindividual standard deviations. Thus, batch testing would lead to a fairly substantial increase in statistical power or reduction in sample size in studies in which the end-point is change in total T or a T fraction over time.

Discussion

This study provided estimates of day-to-day and 3-month intraindividual variation in total T, free T and bioavailable T, seven other adrenal and reproductive hormones and SHBG in a large cohort of generally healthy, community-dwelling, middle-aged to older men of diverse ethnicity. We expect the results to be broadly generalizable because the subjects were randomly sampled from the community.

The relatively narrow bootstrap confidence limits in Table 5 indicate that the differences between the standard deviations for LH, DHEA and cortisol on the one hand and total T, DHT, oestrone and oestradiol on the other are not the result of chance but reflect real

differences in intraindividual variation. As noted earlier, differences in assay variation do not explain the differences in total variation between these two sets of hormones. Differences in ultradian variation brought on by pulsatile release could account for some of the differences in total variation. Pooling equal aliquots of two samples, as was done in this study, reduces but does not eliminate the effect of pulsatile release. Thus, all other sources of variation being equal, the hormone with a greater degree of ultradian variation will display greater intraindividual variation even with the sampling strategy used in this study. For example, LH displays greater ultradian variation and had larger intraindividual standard deviations than total T.^{27,28}

The standard deviations presented here are based on a specific protocol for blood draws and a specific set of assays. It is important to consider the effects of departures from the procedures used in this study on estimates of intraindividual variation in hormone levels. Sampling times were fairly tightly clustered within men to reduce the effects of diurnal variation on intraindividual variation. The amplitude of diurnal variation for total T, free T and bioavailable T, SHBG, LH, FSH and DHEAS may actually exceed the intraindividual standard deviations obtained in this study.^{19,20,29-36} Relaxing the restrictions on intrasubject variation of time of collection would likely increase the intraindividual standard deviations considerably, whereas further restricting the time of collection would probably reduce the standard deviations. The amplitude of diurnal fluctuation, at least for total T, may be reduced in elderly men^{19,30} although possibly not in middle-aged men.³⁶ The extent to which time of day must be controlled in sample collection may therefore depend on a subject's age.

Measurements of hormone levels are often based on a single blood draw, rather than on a pool of equal aliquots from two draws. Hormone levels based on a single draw will be more variable than levels based on a pool of equal aliquots of two draws but the difference in variation is likely to be small. Pooling equal aliquots of two samples drawn 20 min apart will reduce the day-to-day component of variation, σ_D^2 , but it will not alter the long-term component, σ_L^2 or the assay components, σ_A^2 or σ_E^2 . The extent to which total intraindividual variation, σ_T^2 is reduced will depend on the relative contribution of the day-to-day component to the total. It will also depend on the relative contribution of variation over the short interval between blood draws to the day-to-day component because variation over the interval between draws is the only part of day-to-day variation that will be affected by the pooling. Given the other sources of variation that contribute to the total, the effect of pooling on the intraindividual standard deviation is likely very small, especially if one focuses on intraindividual variation over 6 months without batch testing (Scheme 4). It is difficult to be more specific than this without information on the effect of pooling two aliquots on intraindividual variation. This issue is currently under investigation.

Although differences in assay variation among methods will affect intraindividual variation, the effect will be diluted when assay variation is combined with biological variation (σ_D^2 or $\sigma_D^2 + \sigma_L^2$ in the models described earlier). Consider, for example, a study of total T in which an assay is used for which both the intra- and interassay component of variance is two-thirds the component in the assay used in this study. Assuming no change in the biological components, reductions of only 11% and 8.5%, respectively, for the day-to-day and 3-month intraindividual standard deviations, can be calculated from the variance components in Table 4.

New assays for steroid hormones are being used by commercial laboratories and developed by investigators, such as approaches based on liquid chromatography tandem mass spectrometry (LC-MS/SM) and gas chromatography mass spectrometry (GC-MS). The new and existing methods may not measure exactly the same molecular species or isoforms. For example, the antibodies in a method based on radioimmunoassay (RIA) may measure immunoreactive molecules that are not detected by LC-MS/MS or GC-MS. Furthermore, protein hormones, such as LH, FSH and SHBG circulate in various glycosylated isoforms that are differentially recognized by the antibodies used in different RIAs. If the various species or isoforms of a given hormone exhibit different levels of biological variation, then switching to a different method may alter both the assay and biological components of intraindividual variation. Therefore, it may be necessary to repeat the measurements made here using the newer assays in order to determine if the estimates of biological variation are affected by the assay used.

With these caveats in mind, the standard deviations presented here have a number of uses in clinical monitoring of hormone levels and in trials of hormone replacement therapy and in other investigations in which hormones are of interest. As shown in Table 6, confidence limits for steady state mean hormone level for an individual can be calculated from the results of one, two or several measurements. One important application of the confidence limits arises when a threshold of concern has been defined, such as a threshold for total T level below which hypogonadism might be suspected or diagnosed. The standard deviations provide a basis for determining how far a measured T level or the average of a series of measurements of T must be from the threshold to conclude that an individual's steady state mean is reliably above or below the threshold. They also provide a basis for deciding whether the difference between two measurements is large enough to represent a change in hormone level rather than a fluctuation around the steady state mean, for developing eligibility criteria for a clinical trial of hormone replacement therapy in hypogonadal men and for doing sample size calculations for a trial in which change in hormone level is the outcome of interest. The calculations for these uses of the intraindividual standard deviations are straightforward for a statistician; for the sake of brevity, we forego detailed exposition of the methods.

The calculations summarized in Table 6 show that averaging the results from repeated hormone measurements on the same individual reduces the width of the confidence interval for the steady state mean. However, the increment by which the confidence interval is narrowed and precision is improved declines with each additional sample. Reasonably large gains can be made by averaging the results of two or three measurements but, in many cases, the precision gained with further measurements is likely to be too small to be meaningful.

While repeated sampling may improve the precision with which mean hormone level is estimated, there are some circumstances under which such repetition may not be necessary. For example, repeated sampling may not be required if there is a consistent pattern to the results, such as low total T coupled with abnormally high LH and perhaps physical symptoms of hypogonadism. The clinician may find such a cluster of signs and symptoms to be sufficient for a diagnosis without obtaining follow-up blood samples.

In addition to averaging the results of two or more samples from the same subject, intraindividual variation can also be reduced by averaging the results of repeated assays of the same sample. While repeated assays of a sample will reduce assay variation, however, they will not reduce biological variation. Averaging the results from repeated samples reduces both assay and biological variation. Therefore, repeated assays are less effective than repeated samples at reducing total variation. Moreover, the assay components of variance in Table 4 are generally smaller than the biological components, further limiting the gain from repeated assays.

Many clinicians are aware of the problems created by intraindividual variation in hormone levels when interpreting clinical measurements of hormone levels. Researchers are all too aware of the difficulties encountered in designing studies involving hormone levels as eligibility criteria or end-points when information on intraindividual variation is not available. The measurements of intraindividual variation provided here should have broad application clinically and in research in endocrinology.

Acknowledgements

This study was supported by grant number AG23027 from the National Institute on Ageing of the National Institutes of Health, USA.

References

- 1 Vermeulen, A. & Deslypere, J.P. (1985) Testicular endocrine function in the ageing male. *Maturitas*, **7**, 273–279.
- 2 Gray, A., Feldman, H.A., McKinlay, J.B. & Longcope, C. (1991) Age, disease, and changing sex hormone levels in middle-aged men: results of the Massachusetts Male Aging Study. *Journal of Clinical Endocrinology and Metabolism*, **73**, 1016–1025.
- 3 Wu, A.H., Whittemore, A.S., Kolonel, L.N., John, E.M., Gallagher, R.P., West, D.W., Hankin, J., Teh, C.Z., Dreon, D.M. & Paffenbarger, R.S. (1995) Serum androgens and sex hormone-binding globulins in relation to lifestyle factors in older African-American, white, and Asian men in the United States and Canada. *Cancer Epidemiology Biomarkers and Prevention*, 4, 735–741.
- 4 Hsieh, C.C., Signorello, L.B., Lipworth, L., Lagiou, P., Mantzoros, C.S. & Trichopoulos, D. (1998) Predictors of sex hormone levels among the elderly: a study in Greece. *Journal of Clinical Epidemiology*, **51**, 837–841.
- 5 Denti, L., Pasolini, G., Snfelici, L., Benedetti, R., Cecchetti, A., Ceda, G.P., Ablondi, F. & Valenti, G. (2000) Aging-related decline of gonadal function in healthy men: correlation with body composition and lipoproteins. *Journal of the American Geriatrics Society*, 48, 51– 58.
- 6 Feldman, H.A., Longcope, C., Derby, C.A., Johannes, C.B., Araujo, A.B., Coviello, A.D., Bremner, W.J. & McKinlay, J.B. (2002) Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. *Journal of Clinical Endocrinology and Metabolism*, 87, 589–598.
- 7 Gapstur, S.M., Gann, P.H., Kopp, P., Colangelo, L., Longcope, C. & Liu, K. (2002) Serum androgen concentrations in young men: a

longitudinal analysis of associations with age, obesity, and race. The CARDIA male hormone study. *Cancer Epidemiology Biomarkers and Prevention*, **11**, 1041–1047.

- 8 Kaufman, J.M. & Vermeulen, A. (2005) The decline of androgen levels in elderly men and its clinical and therapeutic implications. *Endocrine Reviews*, 26, 833–876.
- 9 Diver, M.J. (2006) Analytical and physiological factors affecting the interpretation of serum testosterone concentration in men. *Annals* of *Clinical Biochemistry*, **43**, 3–12.
- 10 Fox, C.A., Ismail, A.A., Love, D.N., Kirkham, K.E. & Loraine, J.A. (1972) Studies on the relationship between plasma testosterone levels and human sexual activity. *Journal of Endocrinology*, **52**, 51–58.
- 11 Morley, J.E., Patrick, P. & Perry, H.M. 3rd. (2002) Evaluation of assays available to measure free testosterone. *Metabolism*, 51, 554– 559.
- 12 Nieschlag, E. & Ismail, A.A. (1970) Diurnal variations of plasma testosterone in normal and pathological conditions as measured by the technique of competitive protein binding. *Journal of Endocrinology*, 46, 3–4.
- 13 Vermeulen, A. & Verdonck, G. (1992) Representativeness of a single point plasma testosterone level for the long term hormonal milieu in men. *Journal of Clinical Endocrinology and Metabolism*, 74, 939–942.
- 14 Couwenbergs, C., Knussmann, R. & Christiansen, K. (1986) Comparisons of the intra-and inter-individual variability in sex hormone levels of men. *Annals of Human Biology*, **13**, 63–72.
- 15 Ricos, C. & Arbos, M.A. (1990) Quality goals for hormone testing. Annals of Clinical Biochemistry, 27, 353–358.
- 16 Valero-Politi, J. & Fuentes-Arderiu, X. (1993) Within- and betweensubject biological variations of follitropin, lutropin, testosterone, and sex-hormone-binding globulin in men. *Clinical Chemistry*, **39**, 1723–1725.
- 17 Ahokoski, O., Virtanen, A., Huupponen, R., Scheinin, H., Salminen, E., Kairisto, V. & Irjala, K. (1998) Biological day-to-day variation and daytime changes of testosterone, follitropin, lutropin and oestradiol-17beta in healthy men. *Clinical Chemistry and Laboratory Medicine*, 36, 485–491.
- 18 McKinlay, J.B. & Link, C.L. (2007) Measuring the urologic iceberg: design and implementation of the Boston Area Community Health (BACH) Survey. *European Urology*, **52**, 389–396.
- 19 Bremner, W.J., Vitiello, M.V. & Prinz, R.N. (1983) Loss of circadian rhythm in blood testosterone levels with aging in normal men. *Journal of Clinical Endocrinology and Metabolism*, 56, 1278–1281.
- 20 Gray, A., Berlin, J.A., McKinlay, J.B. & Longcope, C. (1991) An examination of research design effects on the association of testosterone and male aging: results of a meta-analysis. *Journal of Clinical Epidemiology*, 44, 671–684.
- 21 Brambilla, D.J., McKinlay, S.M., McKinlay, J.B., Weiss, S.R., Johannes, C.B., Crawford, S.L. & Longcope, C. (1996) Does collecting repeated blood samples from each subject improve the precision of estimated steroid hormone levels? *Journal of Clinical Epidemiology*, 49, 345–350.
- 22 Sodergard, R., Backstrom, T., Shanbhag, V. & Carstensen, H. (1982) Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins at body temperature. *Journal of Steroid Biochemistry*, **16**, 801–810.
- 23 Vermeulen, A., Verdonck, L. & Kaufman, J. (1999) A critical evaluation of simple methods for the estimation of free testosterone in serum. *Journal of Clinical Endocrinology and Metabolism*, 84, 3666–3672.
- 24 Searle, S.R., Casella, G. & McCulloch, C.E. (1992) Variance Components. John Wiley & Sons, New York.

- 25 Fitzmaurice, G.M., Laird, N.M. & Ware, J.H. (2004) *Applied Longitudinal Analysis.* John Wiley & Sons, New York.
- 26 Davison, A.C. & Hinkley, D.V. (1997) *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge, UK.
- 27 Pincus, S.M., Mulligan, T., Iranmanesh, A., Gheorghiu, S., Godschalk, M. & Veldhuis, J.D. (1996) Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males. *Proceedings of the National Academy of Sciences* of the USA, **93**, 14100–14105.
- 28 Mulligan, T., Iranmanesh, A., Gheorghiu, S., Godschalk, M. & Veldhuis, J.D. (1995) Amplified nocturnal luteinizing hormone (LH) secretory burst frequency with selective attenuation of pulsatile (but not basal) testosterone secretion in healthy aged men: possible Leydig cell desensitization to endogenous LH signaling a clinical research center study. *Journal of Clinical Endocrinology and Metabolism*, **80**, 3025–3031.
- 29 Nicolau, G.Y., Haus, E., Lakatua, D.J., Bogdan, C., Sackett-Lundeen, L., Popescu, M., Berg, H., Petrescu, E. & Robu, E. (1985) Circadian and circannual variations of FSH, LH, testosterone, dehydroepiandrosterone-sulfate (DHEA-S) and 17-hydroxy progesterone (17 OH-Prog) in elderly men and women. *Endocrinologie*, 23, 223–246.
- 30 Tenover, J.S., Matsumoto, A.M., Clifton, D.K. & Bremner, W.J. (1988) Age-related alterations in the circadian rhythms of pulsatile

luteinizing hormone and testosterone secretion in healthy men. *Journal of Gerontology*, **43**, M163–M169.

- 31 Plymate, S.R., Tenover, J.S. & Bremner, W.J. (1989) Circadian variation in testosterone, sex hormone-binding globulin, and calculated nonsex hormone-binding globulin bound testosterone in healthy young and elderly men. *Journal of Andrology*, **10**, 366–371.
- 32 Winters, S.J. (1991) Diurnal rhythm of testosterone and luteinizing hormone in hypogonadal men. *Journal of Andrology*, **12**, 185–190.
- 33 Cooke, R.R., McIntosh, J.E. & McIntosh, R.P. (1993) Circadian variation in serum free and non-SHBG-bound testosterone in normal men: measurements, and simulation using a mass action model. *Clinical Endocrinology*, **39**, 163–171.
- 34 Mermall, H., Sothern, R.B., Kanabrocki, E.L., Quadri, S.F., Bremner, F.W., Nemchausky, B.A. & Scheving, L.E. (1995) Temporal (circadian) and functional relationship between prostate-specific antigen and testosterone in healthy men. *Urology*, 46, 45–53.
- 35 Gupta, S.K., Lindemulder, E.A. & Sathyan, G. (2000) Modeling of circadian testosterone in healthy men and hypogonadal men. *Journal* of *Clinical Pharmacology*, **40**, 731–738.
- 36 Diver, M.J., Imtiaz, K.E., Ahmad, A.M., Vora, J.P. & Fraser, W.D. (2003) Diurnal rhythms of serum total, free and bioavailable testosterone and of SHBG in middle-aged men compared with those in young men. *Clinical Endocrinology*, **58**, 710–717.

Effects of testosterone treatment on glucose metabolism and symptoms in men with type 2 diabetes and the metabolic syndrome: a systematic review and meta-analysis of randomized controlled clinical trials

Mathis Grossmann*'t, Rudolf Hoermann*, Gary Wittert‡ and Bu B. Yeap§'¶

*Department of Medicine Austin Health, University of Melbourne, †Endocrine Unit, Austin Health, Heidelberg, Vic., ‡Discipline of Medicine, Royal Adelaide Hospital, University of Adelaide, Adelaide, SA, §School of Medicine and Pharmacology, University of Western Australia and ¶Department of Endocrinology and Diabetes, Fremantle and Fiona Stanley Hospitals, Perth, WA, Australia

Summary

Context The effects of testosterone treatment on glucose metabolism and other outcomes in men with type 2 diabetes (T2D) and/or the metabolic syndrome are controversial.

Objective To perform a systematic review and meta-analysis of placebo-controlled double-blind randomized controlled clinical trials (RCT) of testosterone treatment in men with T2D and/or the metabolic syndrome.

Data sources A systematic search of RCTs was conducted using Medline, Embase and the Cochrane Register of controlled trials from inception to July 2014 followed by a manual review of the literature.

Study selection Eligible studies were published placebo-controlled double-blind RCTs published in English.

Data extraction Two reviewers independently selected studies, determined study quality and extracted outcome and descriptive data.

Data synthesis Of the 112 identified studies, seven RCTs including 833 men were eligible for the meta-analysis. In studies using a simple linear equation to calculate the homeostatic model assessment of insulin resistance (HOMA1), testosterone treatment modestly improved insulin resistance, compared to placebo, pooled mean difference (MD) -1.58 [-2.25, -0.91], P < 0.001. The treatment effect was nonsignificant for RCTs using a more stringent computer-based equation (HOMA2), MD -0.19 [-0.86, 0.49], P = 0.58). Testosterone treatment did not improve glycaemic (HbA1c) control, MD -0.15 [-0.39, 0.10], P = 0.25, or constitutional symptoms, Aging Male Symptom score, MD -2.49 [-5.81, 0.83], P = 0.14).

Conclusions This meta-analysis does not support the routine use of testosterone treatment in men with T2D and/or the metabolic syndrome without classical hypogonadism. Additional studies are needed to determine whether hormonal interventions are warranted in selected men with T2D and/or the metabolic syndrome.

(Received 11 August 2014; returned for revision 17 September 2014; finally revised 9 October 2014; accepted 4 November 2014)

Introduction

A large body of epidemiological evidence shows that, in men, low serum testosterone is associated with insulin resistance and the associated conditions metabolic syndrome and type 2 diabetes (T2D). In meta-analyses of case-control studies, total testosterone levels are consistently lower in men with T2D or with the metabolic syndrome compared to controls, although the degree of the reduction is modest, ranging from -1.61 nM (95%CI -2.56 to -0.65) to -2.99 nM (95%CI -3.59 to -2.40).^{1–3} In addition, low testosterone levels predict incident T2D or the metabolic syndrome in some⁴ but not all⁵ longitudinal studies.

From a mechanistic perspective, this relationship between low testosterone and dysglycaemia is complex, and at least in part, mediated by changes in body composition, in particular visceral fat mass.⁶ It is also bidirectional: On the one hand, androgen deprivation increases fat mass and insulin resistance,⁷ and, on the other hand, weight loss increases both insulin sensitivity and testosterone levels⁸ Preclinical evidence reviewed elsewhere^{6,9} suggests that testosterone regulates stem cells and differentiated adipocytes and myocytes to promote metabolically favourable changes in body composition and glucose metabolism.

The hypothesis that testosterone treatment improves measures of glucose metabolism has been tested in a number of interventional studies, which collectively have yielded inconclusive results. In this study, therefore we sought to conduct a systematic review and meta-analysis of the effects of testosterone therapy

Correspondence: Mathis Grossmann, Department of Medicine Austin Health, The University of Melbourne, 145 Studley Road, Heidelberg, Vic., 3084, Australia. Tel.: +613 9496 5000; Fax: +613 9496 3365; E-mail: mathisg@unimelb.edu.au

on glucose metabolism in men with T2D or the metabolic syndrome. In contrast to previous meta-analyses in this area,^{2,3,10} we limited our analysis to placebo-controlled double-blind randomized controlled clinical trials (RCT) and include two recent such studies^{11,12} that have not been considered previously.

Materials and methods

In this study, we followed the reporting recommendations made in the PRISMA (Preferred reporting items for systematic reviews and meta-analyses) statement.¹³ The PRISMA statement was designed to improve the quality of systematic reviews or metaanalyses. The statement lists 27 items to include when conducting and reporting such a study. In this study, all 27 items were included.

Eligibility criteria

Eligible studies were defined in a protocol as fully published (English language) double-blind randomized controlled clinical trials that assessed the effects of testosterone therapy in men with diagnosed metabolic syndrome and/or T2D on measures of glucose metabolism.

Search strategy

We conducted a comprehensive search of the literature using the electronic databases Medline, Embase and the Cochrane register of controlled trials from inception to July 2014. The search strategy was developed in consultation with an experienced medical research librarian using a broad range of relevant search terms (full research strategies are available in the supplementary information). In addition, reference lists of potentially eligible articles and relevant reviews were searched by hand. Study selection was conducted by two independent reviewers (M.G. and B.B.Y). Studies included by both reviewers were compared and disagreement resolved by consensus and third party adjudication. Only placebo-controlled double-blind RCTs were eligible.

Data extraction

Two investigators (M.G and B.B.Y) independently extracted the relevant data using a standardized form. Data extracted from each eligible RCT included demographic information, diagnosis of T2D and/or the metabolic syndrome, number of participants, baseline and on treatment testosterone levels, type and duration of treatment. Disagreements were resolved by consensus and third party adjudication.

Quality assessment

Two reviewers (M.G and B.B.Y), working in duplicate, assessed the methodological quality of each eligible RCT using the full 25-item CONSORT checklist of information to be included when reporting a RCT.¹⁴ The CONSORT checklist was designed to improve the quality of RCT reporting. Therefore, the quality of a published RCT can be assessed by quantifying how many of the 25 recommended criteria are reported. A high quality RCT will report all 25 items, and the quality of a RCT correlates inversely with the number of reported items.

Primary outcomes

The primary outcomes of interest were the mean differences (MD) in insulin resistance (assessed by homeostatic model assessment of insulin resistance, HOMA-IR) and in glycaemic control (assessed by HbA1c) between treatment and control groups. In addition, standardized mean differences (SMD) were derived. For each eligible RCT, MD \pm Standard Deviation (SD) from baseline to end of trial in each group, treated and controls were retrieved for HOMA-IR and HbA1c. Where SD was not given, it was estimated from SEM or from the 95% confidence interval of the MD. For obtaining SMD, mean differences of individual trials were standardized prior to meta-analysing. In RCTs reporting an open-label extension phase, only the initial double-blind placebo-controlled phase was considered.

Secondary outcomes

The secondary outcomes were effects of testosterone on symptoms, cardiovascular risk markers and adverse effects. We metaanalysed constitutional symptoms reported by Aging Male Symptom score. Due to between-trial heterogeneity and inconsistent reporting, effects of testosterone on sexual symptoms, cardiovascular risk markers and adverse effects could not be meta-analysed, but were instead reported in a descriptive fashion.

Data synthesis and statistical analysis

Three investigators (M.G., B.B.Y and G.W.) independently verified and collated the extracted data to provide a descriptive synthesis of key characteristics and a quantitative synthesis of effect size estimates for each RCT.

The consistency or heterogeneity of the results among various trials in a meta-analysis is an important statistical measure. Hence, the variation in effect beyond chance was tested by using both the Cochran's Q chi-squared test and the I^2 test, which describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error, therefore providing a quantitative measure of nonrandom differences observed across studies. An $I^2 > 30\%$ indicates a moderate intertrial heterogeneity. Given the fact that considerable inconsistency existed among the trials, we used a random effects model to estimate the average true difference (MD) in HOMA-IR or HBA1c in the treatment group, compared to placebo. The choice of differing HOMA-IR modelling employed in the trials, HOMA1 or HOMA2, was assessed in the meta-analysis by adding this information as a moderator variable. SMD are based on Hedges' g with appropriate correction for a negative bias. The Hedges g' with appropriate correction for negative bias was used to calculate the SMD, as recommended by the Cochrane Library, which in contrast to the MD relates the size of the intervention

effect to the variability in each study. In a moderator analysis, we examined the impact of a covariate (moderator variable) on the effect. In addition, we used a weighted restricted maximum-likelihood estimator for fitting the models, and the model was implemented by the metafor package (version 1.9.4) in the R statistical program (R for Mac version 3.1.1).^{15,16}

Graphical data presentation included Forest plots and Funnel plots. A Funnel plot is a scatterplot of treatment effect against a measure of study size-related imprecision (such as the standard error used here) on the vertical axis and serves as a visual aid for detecting bias. In addition to this graphical method, a regression test (regtest) was employed to formally test for the presence of asymmetry in the Funnel plot.

The study is an analysis of published data that does not require specific approval by an ethics committee.

Results

Study identification and descriptive data synthesis

Figure 1 shows the flow chart summarizing the identification of eligible RCTs. Seven RCTs were eligible for analysis, randomizing between 22 and 220 men for a total of 833 men (449 treated with testosterone and 384 receiving placebo) (Table 1).^{11,12,17–21} Major inclusion criteria in all studies were T2D and/or the metabolic syndrome, and low or low-normal serum testosterone levels. Presence of symptoms suggestive of androgen deficiency was an inclusion criterion in five studies.^{12,17–20} Major exclusion criteria were contraindications to testosterone treatment (*e.g.* prostate disease, uncontrolled sleep apnoea, increased haematocrit), previous testosterone or anabolic treatment within the last 6–12 months, and uncontrolled T2D (variously defined HbA1c > 9–10%). 531 men had T2D either with or without the metabolic syndrome, and 302 men had the metabolic syndrome, but no diagnosis of T2D. Mean age of participants ranged from



Fig. 1 Flow chart summarizing the identification of studies included for the meta-analysis.

44 to 64 years, BMI from 24–35 kg/m² and baseline total testosterone levels from 6·7 to 10·1 nm. Durations of the double-blind treatment phase ranged from 3 to 12 months, and six RCTs used intramuscular and one RCT transdermal testosterone (Table 1). The studies were conducted in Italy,¹⁷ India,¹⁹ United Kingdom,^{12,18} Russia²¹ and one was a multinational RCT conducted in eight European Countries.²⁰ Median CONSORT scores ranged from 16 to 24, where 0 denotes the lowest and 25 the highest quality.

Effects of testosterone therapy on insulin resistance (HOMA-IR)

HOMA-IR estimates were derived by different methods in the trials. While five studies^{17–21} were using the original HOMA1 equation (fasting insulin in mUl/L x fasting glucose in mM)/22·5), two trials^{11,12} employed the computer-based HOMA2 model.²² Combining all seven trials resulted in a large intertrial heterogeneity (Cochrane Q-test <0·0001, $I^2 = 76\%$). The heterogeneity was markedly reduced when accounting for the choice of the HOMA-IR model in the analysis (Q_E test P = 0.21, $I^2 = 36.4\%$).

Figure 2a shows the MDs in HOMA-IR after testosterone therapy between treatment and control groups. While testosterone treatment significantly improved insulin resistance, compared to placebo, in the HOMA1-based studies (pooled MD -1.58 [-2.25, -0.91], P < 0.001), the treatment effect was non-significant for the HOMA2 trials (-0.19, [-0.86, 0.49], P = 0.58). In the moderator analysis, the HOMA-IR-model-related shift was 1.39, [0.44, 2.34], P = 0.004. When excluding the study ranking lowest in the quality score,¹⁷ the treatment effect for HOMA1 models decreased slightly to -1.22, [-1.92, -0.53], P = 0.0006. The SMD (Hedges' g) for HOMA-IR across all trials was -0.34 [-0.51, -0.16], P < 0.001, suggesting a small to moderate treatment effect (Fig. 2b). A Funnel plot including a regtest for funnel plot asymmetry (z = 0.20, P = 0.84) revealed no significant publication bias (Figure S1).

Effects of testosterone therapy on glycaemic control (HbA1c)

The intertrial heterogeneity for the testosterone effect on HbA1c was significant among the trials (Cochrane Q-test P < 0.001, $I^2 = 77.3\%$). The pooled difference was minor and did not reach statistical significance (-0.15, [-0.39, 0.10], P = 0.25), as shown in Fig. 3. The funnel plot showed no indication for a publication bias or significant asymmetry (Figure S2, regtest z = 0.82, P = 0.41). Excluding the qualitatively weakest trial¹⁷ proved again statistically inconsequential (-0.10, [-0.34, 0.42],P = 0.42). Omitting individual trials in leave-one-out scenarios proved statistically inconsequential for all individual RCTs with the exception of the trial by Gianatti et al.¹¹ When omitting this trial,¹¹ the reduction in HbA1c was significant $(-0.24 \ [-0.43,$ -0.06], P = 0.02), but failed to retain significance when applying a more robust Knapp Hartung test (P = 0.16). The pooled difference expressed in terms of SMD was nonsignificant, -0.50[-1.37, 0.36], P = 0.25.

Reference	Major inclusion criteria	N (treated)	N (placebo)	Mean age (years)	Mean BMI (m/kg ²)	HbAlc (%)	TT Baseline (nM)	Treatment	Duration (RCT phase)	TT Achieved (nM)
Aversa <i>et al.</i> ¹⁷	$MetSTT \leq 11 \text{ nM}$	40	10	57	31	5.7-6.6	8.7	TU im 1 g	12 months	14.2
Gopal <i>et al.</i> ¹⁹	01 r1 ≤ 230 pM × 2 T2DFT <225 pM	22 (crosso) randomize placebo-co	⁄er, d blinded ntrolled)	44	24	0.7	10.1	T im every T im every 15 days × 6	3 months	N.R.
Jones et al. ²⁰	T2D or MetSTT <11 nm or FT <255 pm × 2	108	112	60	32	7.2% (T2D)	9.4	T gel 60 mg/day	6 months	19.5
Kalinchenko et al. ²¹	MetSTT <12 nm or FT <225 pm	113	71	52	35	N.R.	6.7	TU im 1 g at 0, 6 and 18 weeks	30 weeks	13.1
Kapoor <i>et al.</i> ¹⁸	T2DTT <12 nM \times 2	24 (crossov randomize blind place controlled)	/er, d double- !bo-	64	33	7.3	8.6	T im 200 mg every 2 weeks	3 months (1 months washout)	12.8
Hackett <i>et al.</i> ^{12,24}	T2D TT 8·1–12 nM or ≤8 nM (FT 0·18–0·25 or <0.18 nM) average/2	97	102	62	33	7.6	9.1	TU im 1 g at 0, 6 and 18 weeks	30 weeks (RCT phase)	11.2
Gianatti <i>et al.</i> ^{11,25}	T2D TT ≤ 12 nm average/2	45	43	62	31	7.1	8.6	TU im 1 g at 0, 6, 18 and 30 weeks	40 weeks	13.6
Reference	HOMA-IR			Hb/	A1c (%)			Study design	00	Quality score out of 25)
Aversa <i>et al.</i> ¹⁷	$T - 2.10 \pm 3.5 PBO$	(T –	$0.2 \pm 0.4 PBO$			Mixed Model	1	4
Gopal <i>et al.</i> ¹⁹	T $0.3 \pm 2.61PBO$ T $0.38 \pm 2.61PBO$ $1.72 \pm 3.49MD -$	-1.67 ±		T – T 4.5	0.06 ± 1.87PBO 55MD 1.75 ± 5.3	$-2.04 \pm$ 35(CI - 12.46,	8-95)	.L.LI	1	6
Jones et al. ²⁰	$^{4.29}(CI - 0.91, 10$ T $-0.68 \pm 2.81PB_{-0.07} \pm 2.50$	(c7.(ц Т	$0.06 \pm 1.87 \text{PBO}$	-2.04 ± 4.55		TTI	2	1
Kalinchenko et al. ²¹	$T - 1.49 \pm 2.56 (S)$ $0.20 + 0.84 (SFM)$	EM)PBO		N/A	_			Mixed Model	1	6
Kapoor <i>et al.</i> ¹⁸ Hackett <i>et al.</i> ^{12,24}	MD T -1.73 ± 0.00 T 4.06 ± 2.02 to 4 2.58PBO 3.67 ± 2 2.52MAD 0.15 ± 2	$ \frac{1}{67} (\text{SEM}) P = 67 (\text{SEM}) P = -100 \text{SEM} $ $ \frac{1}{64} \pm 2.64 \text{ to } 3.94 \pm 2.19(\text{CI} - 0.92 \text{ sec}) $	0.02 . 1.22)	MD T 7- 1-2	(-0.37 ± 0.17) ((74 ± 1.31 to 7.6 26PBO 7.47 ± 1. 0.11 ± 0.84(CI -	SEM) $P = 0.03$ S8 \pm 24 to 7.54 \pm -0.34, 0.13)	; I-24MAD	Per Protocol Mixed Model	1	1
Gianatti et al. ^{11,25}	MAD - 0.08(CI - 0.08)).31, 0.47), P =	= 0.23	MA	D 0.36(CI 0.0, 0	(7), P = 0.05		Mixed Model	2	4

Table 1. Characteristics of the randomized controlled trials included

 $\ensuremath{\mathbb{C}}$ 2014 John Wiley & Sons Ltd Clinical Endocrinology (2014), $0, \ 1{-}8$

Effects of testosterone treatment on other cardiovascular risk markers

Three^{11,12,18} of the seven studies reported modest (-0.21 to 0.40 mmol) reductions in total cholesterol with testosterone treatment relative to placebo, with the largest decrease in the smallest study.¹⁸ LDL cholesterol was reduced (by - 0.26 mM) in one study,¹¹ and modest decreases in HDL cholesterol were



Standardized mean difference

Fig. 2 Mean difference (MD) (with 95%CI) in HOMA-IR across randomized controlled trials between testosterone- and placebo-treated groups. (a) MDs were stratified in a moderator analysis by the HOMA-IR model used, as indicated by the grey diamonds. (b) Random effects model of the standardized mean difference (SMD) in HOMA-IR across trials with different HOMA models. SMD is based on Hedges' g (see Materials and methods). The studies by Gianatti¹¹ and Hackett¹² used HOMA2, all others HOMA1 to estimate insulin resistance.

Aversa et al. 2010 Kapoor et al. 2006 Gopal et al. 2010 Jones et al. 2011 Gianatti et al. 2014 Hackett et al. 2013		-1.10 [-2.10, -0.10] -0.37 [-0.46, -0.28] -0.198 [-0.08, 4.04] -0.12 [-0.33, 0.09] 0.20 [-0.09, 0.49] -0.13 [-0.41, 0.15]
RE Model	•	_0.15 [_0.39 , 0.10]
ļ		
-4.00	0.00 2.00	4.00 6.00
	Mean differend	ce

Fig. 3 Mean difference (with 95%CI) in HbA1c across randomized controlled trials between testosterone- and placebo-treated groups.

found in two studies.^{11,20} None of the studies showed effects on triglyceride or blood pressure levels. One study reported a significant decrease in Lipoprotein a.²⁰ The three $RCTs^{11,17,21}$ that assessed C-reactive protein (CRP) levels all showed that testosterone treatment reduced CRP levels, although this was just short of statistical significance (P = 0.05) in one study.¹¹

Effects of testosterone treatment on symptoms

Effects of testosterone treatment on symptoms consistent with androgen deficiency were reported as secondary outcome measures either within the same publication^{18–20} or in a separate manuscript.^{23–25} Constitutional symptoms using the Aging Male Symptom score were reported by 4 RCTs,^{20,23–25} including a total of 691 men.

The trials showed considerable heterogeneity with respect to their mean differences in Aging Male Symptom scores between the testosterone and control groups (Q-test P < 0.001, $I^2 = 92\%$). Overall, the pooled treatment effect of testosterone administration on the symptom score was nonsignificant, MD -2.49 [-5.81, 0.83], P = 0.14) (Fig. 4). The funnel plot (Figure S3) and regtest (z = -0.23, P = 0.82) were sufficiently balanced with a wide variation. The SMD missed significance, -0.35 [-0.76, 0.06], P = 0.09.

Effects on sexual function could not be meta-analysed due to the use of different instruments to assess sexual function across RCTs. Of the four trials that reported effects on sexual function, one study reported a modest improvement erectile function²⁴ whereas three studies did not find significant effects. Two studies reported an increase in sexual desire,^{20,24} whereas overall sexual function improved in only one of the four studies.²³

Adverse events

Adverse event reporting was inconsistent among trials. The reported incidences of serious adverse events were none (in three studies)^{18,19,21} 2.0%,¹⁷ 2.0%¹² 7.7%²⁰ and 9.1%¹¹ with no differences between testosterone and placebo groups. A total of 24 cardiovascular related events were reported in three studies (testosterone group *vs* placebo group: 0 *vs* 1,¹⁷ 3 *vs* 3¹¹ and 5 *vs* 12²⁰), whereas the other four studies did not report any cardiovascular related events. The overall incidence was 2.9%, with



Fig. 4 Mean difference (with 95%CI) in Aging Male symptom score across randomized controlled trials between testosterone- and placebo-treated groups.

eight events occurring in testosterone treated, and 16 in placebo-treated patients, including one fatal cardiovascular event occurring in a man receiving placebo. Significant increases in haematocrit were reported in four,^{11,17,20,21} and significant increases in PSA in one¹¹ of the seven RCTs.

Discussion

In this systematic review and meta-analysis, using rigorous inclusion criteria, we have identified seven double-blind placebo-controlled RCTs that assess the effects of testosterone treatment on glucose metabolism in men with T2D and/or the metabolic syndrome and low to low-normal testosterone levels. Although the RCTs were of moderate to high quality, there was a large between trial heterogeneity. Differences in the choice of HOMA-IR-based methodology to estimate insulin resistance was identified as a major contributor as accounting for HOMA-IR methodology eliminated significant between trial heterogeneity. Collectively, the results indicate that testosterone treatment modestly improved insulin resistance in RCTs that estimated insulin resistance by using the HOMA1 equation. In contrast, in RCTs that used a computer-based model to determine insulin resistance (HOMA2),²² insulin resistance was not improved. While HOMA1 gives a good approximation of beta cell function relative to gold standard insulin clamp studies, the HOMA2 model attempts to account for variations in hepatic and peripheral glucose resistance, for increases in the insulin secretion curve for plasma glucose concentrations above 10 mm, and for the contribution of circulating proinsulin [https://www.dtu.ox. ac.uk/ homacalculator/]. HOMA2 may thus provide a more precise physiological basis for estimation of insulin resistance.²² Therefore, testosterone therapy results in improved glucose-insulin profiles as assessed in the fasting state using HOMA1, which would be consistent with a beneficial effect to ameliorate insulin resistance. However, this may be modulated indirectly via factors which are captured by HOMA2, possibly explaining the disparity in trial results when the newer model is used.

Testosterone treatment had no significant effect on glycaemic control, assessed by HbA1c. Although we did not identify evidence of publication bias, we are aware of one unpublished RCT in 180 men with T2D with baseline HbA1c of 7.0-9.5% and total testosterone < 10.4 nm.²⁶ In this RCT, testosterone treatment had no significant effect on HbA1c or on HOMA-IR, compared to placebo. Insufficient information precluded inclusion of this RCT into the meta-analysis, although it is expected that inclusion of this relatively large negative study would have further reduced the difference in glycaemic outcomes between testosterone and placebo groups. Similarly to the trial by Gianatti,11 this unpublished RCT recruited exclusively patients with established T2D. Reductions in insulin resistance with testosterone treatment were predominantly reported in RCTs that included men with the metabolic syndrome but without established T2D.^{20,21} This raises the possibility that testosterone treatment may be more effective in improving glycaemic outcomes in men with the metabolic syndrome compared to men with established T2D.

Our results differ from previous meta-analyses in this area, which have generally shown more favourable effects of testosterone treatment on glucose metabolism.^{2,3,10} These previous metaanalyses were smaller, including between 228 and 483 participants, showed larger between trial heterogeneity (I^2 38–82%), and were not restricted to placebo-controlled double-blind RCTs, but instead included nonblinded, open-label trials. In addition, the two more recent studies using more stringent HOMA2^{11,12} modelling were not included.

Given that men in all the studies included had relatively well controlled T2D at baseline, the effects of testosterone treatment in men with poorly controlled T2D are unknown. Of note, RCTs in such populations would be more difficult to conduct, given the efficacy of standard antidiabetic medications. In addition, baseline testosterone levels in RCT participants were only modestly reduced, and whether testosterone treatment improves glucose metabolism in men with more marked reductions in testosterone levels remains unknown. Experimental studies of induced hypogonadism have not identified a serum testosterone threshold below which insulin resistance increases.²⁷ Consistent with this, the inverse relationship of insulin resistance with testosterone levels in men with T2D does not have a clear breakpoint and remains present even in men with testosterone levels extending into the normal range.²⁸ Marked reductions in testosterone are relatively uncommon in men with T2D and require careful assessment for underlying classical hypogonadism.8 Such men may well require testosterone treatment irrespective of glycaemic considerations.

While there is evidence that testosterone treatment can improve glucose metabolism in preclinical studies by a variety of cellular mechanisms,^{8,9} the degree of which they are operative in men is unknown. Testosterone treatment modestly increases muscle mass and decreases fat mass,²⁹ changes expected to be metabolically favourable. Of the meta-analysed RCTs, only one study¹¹ reported changes in body composition assessed by rigorous methodology. Despite the expected decrease in fat mass and increase in muscle mass, in that study insulin resistance was not improved. Similarly, in the unpublished RCT, insulin resistance was not improved despite a significant increase in lean body mass.²⁶ In a recent study of 57 obese men, testosterone treatment did not improve insulin sensitivity assessed by euglycaemic clamps, despite significant decreases in fat mass and increases in total fat mass,³⁰ and in a chemical castration study of healthy young men, no changes in insulin resistance were observed despite significant increases in fat mass.³¹ One explanation for this apparent paradox is the observation that in men with T2D testosterone treatment preferentially reduced the amount of subcutaneous, but not of the metabolically more active visceral fat.¹¹ Similarly, testosterone treatment had no significant effect on visceral adipose tissue in most but not all RCTs conducted in overweight or obese men not specifically selected for the presence of T2D and/or the metabolic syndrome.8 Additional studies are needed to clarify whether testosterone acts selectively on subcutaneous compared to visceral fat, or whether specific circumstances exist which lead to one or other reservoir being affected.

Given that the metabolic syndrome and T2D are slowly progressive conditions, it remains possible that longer duration of testosterone treatment, beyond the maximum duration of 12 months in current RCTs (Table 1) may have more marked effect on glucose metabolism. While uncontrolled registry studies have demonstrated progressive improvements in glucose metabolism in men treated with testosterone up to 6 years,³² these observations have yet to be confirmed in controlled trials.

With respect to the effects of testosterone treatment on other cardiovascular risk factors, effects in the different RCTs were relatively modest and consistent with findings from RCTs in of testosterone therapy in men from the general population.³³ Effects on lipids were modest and may be neutral from a cardiovascular perspective as both decreases in pro-atherogenic lipid fractions (total cholesterol, LDL cholesterol, lipoprotein a) and in the theoretically cardioprotective HDL cholesterol were reported. Triglyceride levels did not change, consistent with the absence of documented changes in visceral fat. While none of the RCTs found an effect on blood pressure, modest decreases in CRP were seen, although whether this is independent of changes in body composition is not known.

Testosterone treatment did not have a significant effect on constitutional symptoms suggestive of androgen deficiency, as assessed by the relatively nonspecific Aging Male Symptoms score. Heterogeneity in the instruments used to assess sexual function precluded a meta-analysis of sexual function. Some^{20,23,24} but not all²⁵ RCTs reported improvements in sexual parameters although these were variable between the studies, including erectile function,²⁴ sexual desire^{20,23,24} and overall sexual function. In one RCT,²⁵ sexual and constitutional symptoms were worse in men with depression and microvascular complications, but did not correlate with testosterone levels, suggesting that these nonspecific symptoms are confounded by comorbidities. This provides a potential explanation for the relatively modest and inconsistent symptomatic benefit of testosterone treatment in this population, which is distinct from the marked symptomatic benefits in men with classical, pathologically based hypogonadism who have unequivocally low testosterone levels and objective evidence of androgen deficiency.³⁴

In these relatively short term RCTs as expected, serious adverse events were few. Consistent with men in the general population,³³ an increase in haematocrit was the most commonly reported adverse event, occurring in four of the seven RCTs. Cardiovascular events were numerically lower in men receiving testosterone a difference largely driven by the TIMES2 study.²⁰ Although the number of events was very low, this may provide some assurance given the uncertainties surrounding the cardiovascular safety of testosterone treatment.³⁵

Strengths of the present meta-analysis include, in contrast to previous meta-analyses,^{2,3,10} the strict selection of double-blind placebo-controlled RCTs including two recent RCTs^{11,12} using more rigorous HOMA2 methodology to estimate insulin resistance, which allowed identification of sources of heterogeneity and the quantification of summary estimates across all published RCTs. We used the PRISMA statement to conduct our research,¹³ and carefully evaluated study quality using the full 25-item CONSORT checklist.¹⁴ Limitations include the limited number and relative modest size of included studies, lack of

access to individual patient level data and the fact that included studies not always fully complied with CONSORT reporting criteria, which may weaken the precision of some of our estimates. Moreover, while none of the RCTs used 'gold standard' measurements of insulin resistance, HOMA-IR measurements have shown to correlate well with clamp-based technologies.²²

In conclusion, the results of this meta-analysis do not support the routine use of testosterone treatment to improve glucose metabolism or constitutional symptoms in men with relatively well controlled T2D and/or the metabolic syndrome and modest reductions in testosterone levels. This reinforces recommendations that, at the current state of evidence, lifestyle measures and use of standard therapy to optimize glycaemic control and comorbidities should remain the first line approach.⁸

Acknowledgements

The authors thank Cheryl Hamill and the staff of the Medical Library, Fremantle Hospital, for assistance with the literature search.

M Grossmann was supported by a National Health and Medical Research Council of Australia Career Development Fellowship (# 1024139).

Conflict of interest

Mathis Grossmann received research funding from Bayer Schering and speaker's honoraria from Besins Healthcare. Rudolf Hoermann has nothing to disclose. Gary Wittert received research funding from Bayer Schering, Eli Lilly and Lawley Pharmaceuticals, speaker's honoraria from Bayer Schering and is a member of an Eli Lilly advisory board. Bu B Yeap received speaker's honoraria and conference support from Bayer Schering and Eli Lilly and is a member of an Eli Lilly Advisory Board.

References

- 1 Ding, E.L., Song, Y., Malik, V.S. *et al.* (2006) Sex differences of endogenous sex hormones and risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA*, **295**, 1288–1299.
- 2 Corona, G., Monami, M., Rastrelli, G. *et al.* (2010) Type 2 diabetes mellitus and testosterone: a meta-analysis study. *International Journal of Andrology*, **34**, 528–540.
- 3 Corona, G., Monami, M., Rastrelli, G. *et al.* (2011) Testosterone and metabolic syndrome: a meta-analysis study. *The Journal of Sexual Medicine*, **8**, 272–283.
- 4 Laaksonen, D.E., Niskanen, L., Punnonen, K. *et al.* (2004) Testosterone and sex hormone-binding globulin predict the metabolic syndrome and diabetes in middle-aged men. *Diabetes Care*, **27**, 1036–1041.
- 5 Bhasin, S., Jasjua, G.K., Pencina, M. *et al.* (2011) Sex hormonebinding globulin, but not testosterone, is associated prospectively and independently with incident metabolic syndrome in men: the framingham heart study. *Diabetes Care*, **34**, 2464–2470.
- 6 Grossmann, M. (2014) Testosterone and glucose metabolism in men: current concepts and controversies. *Journal of Endocrinology*, 220, R37–R55.

- 7 Hamilton, E.J., Gianatti, E., Strauss, B.J. *et al.* (2011) Increase in visceral and subcutaneous abdominal fat in men with prostate cancer treated with androgen deprivation therapy. *Clinical Endocrinology (Oxford)*, **74**, 377–383.
- 8 Grossmann, M. (2011) Low testosterone in men with type 2 diabetes: significance and treatment. *Journal of Clinical Endocrinology and Metabolism*, **96**, 2341–2353.
- 9 Rao, P.M., Kelly, D.M. & Jones, T.H. (2013) Testosterone and insulin resistance in the metabolic syndrome and T2DM in men. *Nature Reviews. Endocrinology*, **9**, 479–493.
- 10 Cai, X., Tian, Y., Wu, T. *et al.* (2014) Metabolic effects of testosterone replacement therapy on hypogonadal men with type 2 diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. *Asian Journal of Andrology*, 16, 146–152.
- 11 Gianatti, E.J., Dupuis, P., Hoermann, R. *et al.* (2014) Effect of testosterone treatment on glucose metabolism in men with type 2 diabetes: a randomized controlled trial. *Diabetes Care*, 37, 2098–2107.
- 12 Hackett, G., Cole, N., Bhartia, M. *et al.* (2013) Testosterone Replacement therapy improves metabolic parameters in hypogonadal men with type 2 diabetes but not in men with coexisting depression: the BLAST Study. *The Journal of Sexual Medicine*, **11**, 840–856.
- 13 Moher, D., Liberati, A., Tetzlaff, J. *et al.* (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, **151**, 264–269, W264.
- 14 Moher, D., Hopewell, S., Schulz, K.F. *et al.* (2010) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, c869.
- 15 R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://wwwR-projectorg/ (accessed 7 June 2014).
- 16 Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, **36**, 1–48.
- 17 Aversa, A., Bruzziches, R., Francomano, D. *et al.* (2010) Effects of testosterone undecanoate on cardiovascular risk factors and atherosclerosis in middle-aged men with late-onset hypogonadism and metabolic syndrome: results from a 24-month, randomized, double-blind, placebo-controlled study. *The Journal of Sexual Medicine*, 7, 3495–3503.
- 18 Kapoor, D., Goodwin, E., Channer, K.S. *et al.* (2006) Testosterone replacement therapy improves insulin resistance, glycaemic control, visceral adiposity and hypercholesterolaemia in hypogonadal men with type 2 diabetes. *European Journal of Endocrinology*, **154**, 899–906.
- 19 Gopal, R.A., Bothra, N., Acharya, S.V. *et al.* (2010) Treatment of hypogonadism with testosterone in patients with type 2 diabetes mellitus. *Endocrine Practice*, **16**, 570–576.
- 20 Jones, T.H., Arver, S., Behre, H.M. *et al.* (2011) Testosterone replacement in hypogonadal men with type 2 diabetes and/or metabolic syndrome (the TIMES2 Study). *Diabetes Care*, **34**, 828–837.
- 21 Kalinchenko, S.Y., Tishova, Y.A., Mskhalaya, G.J. *et al.* (2010) Effects of testosterone supplementation on markers of the metabolic syndrome and inflammation in hypogonadal men with the metabolic syndrome: the double-blinded placebo-controlled Moscow study. *Clinical Endocrinology (Oxford)*, **73**, 602–612.
- 22 Wallace, T.M., Levy, J.C. & Matthews, D.R. (2004) Use and abuse of HOMA modeling. *Diabetes Care*, **27**, 1487–1495.

- 23 Giltay, E.J., Tishova, Y.A., Mskhalaya, G.J. *et al.* (2010) Effects of testosterone supplementation on depressive symptoms and sexual dysfunction in hypogonadal men with the metabolic syndrome. *The Journal of Sexual Medicine*, **7**, 2572–2582.
- 24 Hackett, G., Cole, N., Bhartia, M. *et al.* (2013) Testosterone replacement therapy with long-acting testosterone undecanoate improves sexual function and quality-of-life parameters vs. placebo in a population of men with type 2 diabetes. *The Journal of Sexual Medicine*, **10**, 1612–1627.
- 25 Gianatti, E.J., Dupuis, P., Hoermann, R. *et al.* (2014) Effect of testosterone treatment on constitutional and sexual symptoms in men with type 2 diabetes in a randomized, placebo-controlled clinical trial. *Journal of Clinical Endocrinology and Metabolism* **99**, 3821–3828.
- 26 Unpublished accessed at Available from: http://www.solvaypharmaceuticals.com/static/wma/pdf/1/3/4/4/2/S176.2.101.pdf (accessed 28 May 2014).
- 27 Singh, A.B., Hsia, S., Alaupovic, P. *et al.* (2002) The effects of varying doses of T on insulin sensitivity, plasma lipids, apolipoproteins, and C-reactive protein in healthy young men. *Journal of Clinical Endocrinology and Metabolism*, **87**, 136–143.
- 28 Grossmann, M., Thomas, M.C., Panagiotopoulos, S. *et al.* (2008) Low testosterone levels are common and associated with insulin resistance in men with diabetes. *Journal of Clinical Endocrinology and Metabolism*, **93**, 1834–1840.
- 29 Isidori, A.M., Giannetta, E., Greco, E.A. *et al.* (2005) Effects of testosterone on body composition, bone metabolism and serum lipid profile in middle-aged men: a meta-analysis. *Clinical Endocrinology (Oxford)*, **63**, 280–293.
- 30 Juang, P.S., Peng, S., Allehmazedeh, K. *et al.* (2013) Testosterone with dutasteride, but not anastrazole, improves insulin sensitivity in young obese men: a randomized controlled trial. *The Journal of Sexual Medicine*, **11**, 563–573.
- 31 Woodhouse, L.J., Gupta, N., Bhasin, M. *et al.* (2004) Dosedependent effects of testosterone on regional adipose tissue distribution in healthy young men. *Journal of Clinical Endocrinology and Metabolism*, **89**, 718–726.
- 32 Haider, A., Yassin, A., Doros, G. *et al.* (2014) Effects of long-term testosterone therapy on patients with "diabesity": results of observational studies of pooled analyses in obese hypogonadal men with type 2 diabetes. *International Journal of Endocrinology*, **2014**, 683515.
- 33 Fernandez-Balsells, M.M., Murad, M.H., Lane, M. *et al.* (2010) Clinical review 1: adverse effects of testosterone therapy in adult men: a systematic review and meta-analysis. *Journal of Clinical Endocrinology and Metabolism*, **95**, 2560–2575.
- 34 Snyder, P.J., Peachey, H., Berlin, J.A. *et al.* (2000) Effects of testosterone replacement in hypogonadal men. *Journal of Clinical Endocrinology and Metabolism* **85**, 2670–2677.
- 35 Endocrine Society Statement (2014) The Risk of Cardiovascular Events in Men Receiving Testosterone Therapy. http://www. endocrine.org/~/media/endosociety/Files/ Advocacy and Outreach/Position Statements/Other Statements/The Risk of Cardiovascular Events in Men Receiving Testosterone Therapy.pdf (accessed 28 May 2014)

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.

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

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

(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 $\pm 20\%$ 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)

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 a 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 to	otal testosterone (ng/	dl) external qu	ality control program	(College of American	Pathologists, sample Y-04)
----------	----------------------	------------------------	-----------------	-----------------------	----------------------	----------------------------

Instrument/assay	No. of	Mean		CV	N 1.	Range	
	labs	(ng/dl)	SD		Median	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 $102 \pm 1\%$ (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 steroidstripped serum samples spiked with T (purity ascertained by celite

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

TABLE 2. Characteristics of the methods

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 antigenantibody 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 μ l 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, *lower 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 sp (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 \pm 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 $+48.1 \pm 7.5$, $+33.8 \pm 11.1$, 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 $+99 \pm 11$ 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 \pm 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 \pm 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 \pm 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 Or-tho 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 I	Deming regres	sion 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)	$-39.2 (-73.7 \text{ to } -4.2^{a})$	0.948(0.910 - 0.967)
Roche Elecsys	1.167 (1.112-1.222)	$-75.5 (-102 \text{ to } -49.1^{a})$	0.965(0.939 - 0.978)
Vitros ECi	1.233(1.136 - 1.330)	$-118.4 (-160.5 \text{ to } -76.4^{a})$	0.954(0.921 - 0.971)
DCP Immulite	0.881 (0.838-0.924)	$-28.6 (-49.8 \text{ to } -7.4^{a})$	$0.925 (0.628 - 0.969^{b})$
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)

	DDG DIA				DDG I I''	P. C. I
	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 T value <100 ng/dl	20/101 (19.8%) 7/11 (63.6)	40/101 (39.6%) 10/11 (90.9%)	26/101 (25.7%) 6/11 (55.5%)	40/101 (39.7%) 10/11 (90.9%)	49/101 (48.5%) 7/11 (63.6%)	51/101 (50.4%) 6/11 (55.5%)

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

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 fourparameter 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 \pm 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 \pm 20% of the reference method in over 60% of the samples. At low T concentrations ($\leq 100 \text{ ng/dl}$), 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
- Chen J, Zoru E, Hallberg M, Wieland R 1971 Antibodies to testosterone-3bovine serum albumin applied to assay of serum 17-β-ol androgens. Clin Chem 17:581–584
- 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
- Larsen PR, Kronenberg HM, Melmed S, Polonsky K 2002 William's textbook of endocrinology, reference values. 10th ed. Philadelphia: Saunders
 Dorgan JF, Fears TR, McMahon RP, Friedman LA, Patterson BH, Greenhut
- 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

- Gelpi E 1995 Biochemical and biochemical applications of liquid chromatography-mass spectrometry. J Chromatogr A 703:59–80
- 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
- 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
- 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
- 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
- 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
 Swerdloff RS, Wang C, Cunningham G, Dobs A, Iranmanesh A, Matsumoto
- 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
- 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
- Bland JM, Altman DG 1999 Measuring agreement in method comparison studies. Stat Methods Med Res 8:135–160
- Pollock MA, Jefferson SG, Kane JW, Lomax K, Mackinnon G, Winnard CB 1992 Method comparison—a different approach. Ann Clin Biochem 29: 556–560
- 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
- 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
- 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
- Guardabasso V, Rodbard D, Munson PJ 1987 A dose-response analysis. Am J Physiol 252:E357–E364
- 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.

Long-Term Pharmacokinetics of Transdermal Testosterone Gel in Hypogonadal Men*

RONALD S. SWERDLOFF, CHRISTINA WANG, GLENN CUNNINGHAM, ADRIAN DOBS, ALI IRANMANESH, ALVIN M. MATSUMOTO, PETER J. SNYDER, THOMAS WEBER, JAMES LONGSTRETH, NANCY BERMAN, AND THE TESTOSTERONE GEL STUDY GROUP[†]

Divisions of Endocrinology, Departments of Medicine / Pediatrics, Harbor-University of California-Los Angeles Medical Center and Research and Education Institute (R.S.S., C.W., N.B.), Torrance, California 90509; Veterans Affairs Medical Center, Baylor College of Medicine (G.C.), Houston, Texas 77030; The Johns Hopkins University (A.D.), Baltimore, Maryland 21287; Veterans Affairs Medical Center (A.I.), Salem, Virginia 24153; Veterans Affairs Puget Sound Health Care System, University of Washington (A.M.M.), Seattle, Washington 98108; University of Pennsylvania Medical Center (P.J.S.), Philadelphia, Pennsylvania 19104; Duke University Medical Center (T.W.), Durham, North Carolina 27705; Unimed Pharmaceuticals, Inc. (J.L.), Deerfield, Illinois 60015

ABSTRACT

Transdermal delivery of testosterone (T) represents an effective alternative to injectable androgens. Transdermal T patches normalize serum T levels and reverse the symptoms of androgen deficiency in hypogonadal men. However, the acceptance of the closed system T patches has been limited by skin irritation and/or lack of adherence. T gels have been proposed as delivery modes that minimize these problems. In this study we examined the pharmacokinetic profiles after 1, 30, 90, and 180 days of daily application of 2 doses of T gel (50 and 100 mg T in 5 and 10 g gel, delivering 5 and 10 mg T/day, respectively) and a permeation-enhanced T patch (2 patches delivering 5 mg T/day) in 227 hypogonadal men. This new 1% hydroalcoholic T gel formulation when applied to the upper arms, shoulders, and abdomen dried within a few minutes, and about 9-14% of the T applied was bioavailable. After 90 days of T gel treatment, the dose was titrated up (50 mg to 75 mg) or down (100 mg to 75 mg) if the preapplication serum T levels were outside the normal adult male range. Serum T rose rapidly into the normal adult male range on day 1 with the first T gel or patch application. Our previous study showed that steady state T levels were achieved 48-72 h after first application of the gel. The pharmacokinetic parameters for serum total and free T were very similar on days 30, 90, and 180 in all treatment groups. After repeated daily application of the T formulations for 180 days, the average serum T level over the 24-h sampling period (C_{avg}) was highest in the 100 mg T gel group (1.4- and 1.9-fold higher than the C_{avg} in the 50 mg T gel and T patch groups, respectively). Mean serum steady state T levels remained stable over the 180 days of T gel application. Upward dose adjustment from T gel 50 to 75 mg/day did not significantly increase the Cave, whereas downward dose adjustment from 100 to 75 mg/day reduced serum T levels to the normal range for most patients. Serum free T levels paralleled those of serum total T, and the percent free T was not changed with transdermal T preparations. The serum dihydrotestosterone C_{avg} rose 1.3-fold above baseline after T patch application, but was more significantly increased by 3.6- and 4.6-fold with T gel 50 and 100 mg/day, respectively, resulting in a small, but significant, increase in the serum dihydrotestosterone/T ratios in the two T gel groups. Serum estradiol rose, and serum LH and FSH levels were suppressed proportionately with serum T in all study groups; serum sex hormone-binding globulin showed small decreases that were significant only in the 100 mg T gel group. We conclude that transdermal T gel application can efficiently and rapidly increase serum T and free T levels in hypogonadal men to within the normal range. Transdermal T gel provided flexibility in dosing with little skin irritation and a low discontinuation rate. (J Clin Endocrinol Metab 85: 4500-4510, 2000)

THE SKIN IS an attractive route for systemic delivery of steroids. Transdermal preparations of testosterone (T) provide a useful delivery system for normalizing serum T levels in hypogonadal men and preventing the clinical symptoms and long-term effects of androgen deficiency (1–5). Currently available transdermal patches are applied to the

scrotal skin (Testosderm) or to other parts of the body (Androderm and Testoderm TTS). The former requires preparation of the scrotal skin with hair clipping or shaving to optimize adherence of the patches. The permeationenhanced T patch (Androderm) is associated with skin irritation in about a third of the patients, and 10–15% of subjects have been reported to discontinue the treatment because of

Received December 28, 1999. Revision received March 25, 2000. Rerevision received June 30, 2000. Accepted August 30, 2000.

Address all correspondence and requests for reprints to: Christina Wang, M.D., General Clinical Research Center, Harbor-University of California-Los Angeles Medical Center, 1000 West Carson Street, Torrance, California 90509-2910. E-mail: wang@gcrc.humc.edu.

^{*} This work was supported by grants from Unimed Pharmaceuticals, Inc. The work at Harbor-University of California-Los Angeles Medical Center was supported by NIH Grant M01-RR-00425 to the General Clinical Research Center. The work at Duke University Medical Center was performed at the General Clinical Research Center supported by NIH Grant M01-RR-0030.

[†] The Testosterone Gel Study Group includes: S. Berger, The Chicago Center for Clinical Research (Chicago, IL); E. Dula, West Coast Clinical Research (Van Nuys, CA); J. Kaufman, Urology Research Options (Aurora, CO); G. P. Redmond, Center for Health Studies (Cleveland, OH); S. Scheinman and H. W. Hutman, South Florida Bioavailability Clinic (Miami, FL); S. L. Schwartz, Diabetes and Glandular Disease Clinic, P.A. (San Antonio, TX); C. Steidle, Northeast Indiana Research (Fort Wayne, IN); J. Susset, MultiMed Research (Providence, RI); G. Wells, Alabama Research Center, L.L.C. (Birmingham, AL); and R. E. Dudley, S. Faulkner, N. Rehousky, G. Ringham, W. Singleton, and K. Zunich, Unimed Pharmaceuticals, Inc. (Deerfield, IL).

chronic skin irritation (6, 7). Preapplication of corticosteroid cream at the site of application of the Androderm patch has been reported to decrease the incidence and severity of the skin irritation (8). The most recently approved nonscrotal T patch (Testoderm TTS) causes less skin irritation (itching in about 12% and erythema in 3% of the subjects), but adherence of the patch to the skin poses a problem in some subjects (9, 10). Despite these limitations of local irritation and adherence to skin, the various T patches provide a steady state delivery of T to the circulation that mimics the normal diurnal rhythm of serum T at the low to mid normal adult male range (11–17). The long-term use of these transdermal androgen delivery patches has been shown to be efficacious in maintaining sexual function, secondary sexual characteristics, and bone and muscle mass in hypogonadal young and elderly men (5, 18-21).

T and other steroids can also be applied to the skin in open systems. When T is applied to the skin surface as a hydroalcoholic gel, the gel dries rapidly, and the steroid is absorbed into the stratum corneum, which serves as a reservoir. The reservoir in the skin releases T into the circulation slowly over several hours, resulting in steady state serum levels of the hormones (22). Our previous short-term (7-14 days) pharmacokinetic studies of both T and 5α-dihydrotestosterone (DHT) transdermal hydroalcoholic gels showed that the androgens were absorbed, and peak levels of the applied androgens occurred 18-24 h after initial application. With continued application of the gel for 7-14 days, steady serum levels of androgens were maintained (23, 24). About 9–14% of the T in the gel applied to the skin is bioavailable (24). We also demonstrated that application of the T gel (100 mg/day)at a single site or four separate sites resulted in serum T levels at the upper limit of the normal range, with about 23% higher serum levels when the gel was applied at four sites. In the 7to 14-day studies, neither T nor DHT gel produced skin irritation in the small number of subjects studied (23, 24). In the present study we investigated the detailed pharmacokinetics and tolerability of T gel (AndroGel) at two dosages (50 and 100 mg/day) and T patch after repeated daily dosing for 180 days in a large number of hypogonadal men (n = 227) recruited from 16 centers across the United States.

TABLE 1. Baseline characteristics of the hypogonadal men

Subjects and Methods

Subjects

Two hundred and twenty-seven hypogonadal men were recruited, randomized, and studied in 16 centers in the United States. About one third of the subjects were randomized into each treatment group (Table 1). The patients were between 19-68 yr of age and had single morning serum T levels at screening of 10.4 nmol/L (300 ng/dL) or less. The screening serum T concentrations were measured at each center's clinical laboratory. Previously treated hypogonadal men were withdrawn from T ester injection for at least 6 weeks and from oral or transdermal androgens for 4 weeks before the screening visit. Aside from the hypogonadism, the subjects were in good health, as evidenced by medical history, physical examination, complete blood count, urinalysis, and serum biochemistry. If the subjects were taking lipid- lowering agents or tranquilizers, the doses were stabilized for at least 3 months before enrollment. The subjects had no history of chronic medical illness or alcohol or drug abuse. The subjects had a normal rectal examination, a prostate-specific antigen level of less than 4 ng/mL, and a urine flow rate of more than 12 mL/s before enrollment to the study. They were excluded if they had a generalized skin disease that might affect T absorption or a prior history of skin irritability with the nonscrotal T patch (Androderm). Subjects with body weight of less than 80 or more than 140% of ideal body weight and subjects taking medications known to alter the cytochrome P450 enzyme systems were also excluded from this study.

T gel and patch

T gel (AndroGel) was manufactured by Besins Iscovesco (Paris, France) and supplied by Unimed Pharmaceuticals, Inc. (Deerfield, IL). The formulation is a hydroalcoholic gel containing 1% T (10 mg/g). We have previously shown that about 9-14% of the steroid in the gel applied is available to the body. Thus, 10 g gel applied to the skin contain 100 mg T and delivers approximately 10 mg T to the body (23, 24). Approximately 250 g gel were packaged in multidose glass bottles that delivered 2.27 g gel for each actuation of the pump. Patients assigned to the 50 mg T in 5 g gel group were given one bottle of T gel and one bottle of placebo gel (vehicle only); those assigned to the 100 mg T in 10 g gel were dispensed two bottles of the active T gel. All patients applied T gel or placebo gel at four separate sites each day (right and left upper arms/shoulders and right and left abdomen). On day 1 of the study, the patients were instructed to depress the pump of one of the bottles once, and the gel was applied to the right upper arm/shoulder. Then, using the same bottle, a second dose of gel was delivered and applied to the left upper arm/shoulder. The second bottle was then used with the actuation of the pump for gel to be applied to the right abdomen and the second actuation to the left abdomen. On the following day, the application sites were reversed. Alternate application sites continued throughout the study. After application of the gel to the skin, the gel

Treatment group	T patch (5 mg/day)	T gel (50 mg/day)	T gel (100 mg/day)
No. of subjects enrolled	76	73	78
Age (yr)	51.1	51.3	51.0
Range (yr)	28 - 67	23 - 67	19 - 68
Ht (cm)	179.3 ± 0.9	175.8 ± 0.8	178.6 ± 0.8
Wt (kg)	92.7 ± 1.6	90.5 ± 1.8	91.6 ± 1.5
Serum T (nmol/L) at screening ^{a}	6.40 ± 0.41	6.44 ± 0.39	6.49 ± 0.37
Causes of hypogonadism			
Primary hypogonadism	34	26	34
Secondary hypogonadism	15	17	12
Aging	6	13	6
Normogonadotropic hypogonadism	21	17	26
Yr diagnosed	5.8 ± 1.1	4.4 ± 0.9	5.7 ± 1.24
No. previously treated with T (%)	50 (65.8)	38 (52.1)	46 (59.0)
Duration of treatment (vr)	5.8 ± 1.0	5.4 ± 0.8	4.6 ± 0.7

 a Screening serum T concentrations were measured before enrollment in each study center's clinical laboratory and not at the central laboratory.

dried within a few minutes. The patients washed their hands with soap and water thoroughly after gel application. After 90 days the subjects titrated to the 75 mg/day T gel dose were supplied with three bottles, one containing placebo and two containing T gel. The subjects were instructed to apply one actuation from the placebo bottle and three actuations from the T gel bottle to four different sites of the body as described above.

T patches (Androderm) were provided, each delivering 2.5 mg/day T, which is the recommended replacement dose for androgen replacement therapy. The patients were instructed to apply two T patches to a clean dry area of skin on the back, abdomen, upper arms, or thighs once per day. Application sites were rotated, with an approximately 7-day interval between applications to the same site. T gel or patches were applied at approximately 0800 h each morning for 180 days.

In the T gel group, treatment compliance was estimated as the percentage of T gel actually used compared with the theoretical amount of T gel that could have been used. The actual amount of T gel used was measured as the difference in weight of the dispensed and returned T gel bottles. The theoretical weight of T gel that could have been used was calculated as 2.27 g/actuation × days in study × 2, 3, or 4 actuations depending on whether the dose of T gel was 50, 75, or 100 mg, respectively. In the T patch group, the actual number of patches used was compared with the theoretical number that could have been used calculated as days in study × 2 patches/day.

Study design

The study is a randomized, multicenter (16 centers), parallel study including 2 doses of T gel and a single dose of T patches. A placebo group was not included because 6-month placebo treatment of hypogonadal men was not believed to be justifiable, as untreated hypogonadism will result in impaired libido, decreased strength, bone mineral loss, and other clinical defects. The study was double blinded until day 90 with respect to the T gel groups and open label for the T patch group. For the first 3 months of the study (days 1-90), the subjects were randomized to receive 50 mg/day T gel (in 5 g gel delivering about 5 mg T/day), 100 mg/day T gel (in 10 g gel delivering about 10 mg T/day), or 2 patches delivering 5 mg T/day (T patch). In the following 3 months (days 91-180), the subjects were administered 1 of the following treatments: 50 mg/day T gel, 100 mg/day T gel, 75 mg/day T gel, or 5.0 mg/day T patch. Patients who were applying T gel had a single, preapplication serum T measurement made on day 60; if the levels were within the normal range (10.4-34.7 nmol/L; 300-1000 ng/dL), they remained on their original dose. Men with T levels at 60 days of treatment less than 10.4 nmol/L and who were applying 50 mg T gel and those with T levels more than 34.7 nmol/L who had received 100 mg T gel were then assigned to the 75 mg/day T gel group for days 91–180. No changes in dose were made to subjects randomized to T patch.

On days 0, 1, 30, 90, and 180 subjects had multiple blood samples for T and free T measurements at 30, 15, and 0 min before and 2, 4, 8, 12, 16, and 24 h after T gel or patch application. Brief history and physical examinations were performed, and any complaints or adverse events were documented in the subject's records. In addition, subjects returned to each study center on days 60, 120, and 150 for a single blood sampling before application of the gel or patch. Serum DHT, estradiol (E_2), FSH, LH, and sex hormone-binding globulin (SHBG) were measured in samples collected before gel or patch application on days 0, 30, 60, 90, 120, 150, and 180. Sera for hormones were stored frozen at -20 C until assay. All samples for a patient for each hormone were measured in the same assay whenever possible. In addition, the subjects were examined for any adverse effects and skin irritation.

Hormone assays

Except for the screening serum T concentration, which was measured at each center's clinical laboratory, all hormone assays were performed at the Endocrine Research Laboratory of the Harbor-University of California-Los Angeles Medical Center. Serum T levels were measured after extraction with ethyl acetate and hexane by a specific RIA using reagents from ICN Biomedicals, Inc. (Costa Mesa, CA). The cross-reactivities of the antiserum used in the T RIA were 2.0% for DHT, 2.3% for androstenedione, 0.8% for 3β -androstanediol, 0.6% for etiocholanolone, and less than 0.01% for all other steroids tested. The lower limit of quantitation of serum T measured by this assay was 0.87 nmol/L (25 ng/dL). The mean accuracy (recovery) of the T assay, determined by spiking steroid free serum with varying amounts of T (0.9–52 nmol/L), was 104% (range, 92–117%). The intra- and interassay coefficients of the T assay were 7.3% and 11.1% at the normal adult male range, which in our laboratory was 10.33–36.17 nmol/L (298–1043 ng/dL). Serum free T was measured by RIA of the dialysate after an overnight equilibrium dialysis, using the same RIA reagents as in the T assay. The lower limit of quantitation of serum free T using this equilibrium dialysis method was estimated to be 22 pmol/L. When steroid-free serum was spiked with increasing doses of T in the adult male range, increasing amounts of free T were recovered, with a coefficient of variation that ranged from 11–18.5%. The intra- and interassay precisions of free T were 15% and 16.8%, respectively, for adult normal male values (121–620 pmol/L, 3.48–17.9 ng/dL).

Serum DHT was measured by RIA after potassium permanganate treatment of the sample followed by extraction. The methods and reagents of the DHT assay were provided by Diagnostic Systems Laboratories, Inc. (Webster, TX). The cross-reactivities of the antiserum used in the RIA for DHT were 6.5% for $3\beta\text{-androstanediol},~1.2\%$ for $3\alpha\text{-}$ androstanediol, 0.4% for 3a-androstanediol glucuronide, 0.4% for T (after potassium permanganate treatment and extraction), and less than 0.01 for other steroids tested. This low cross-reactivity against T was further confirmed by spiking steroid free serum with T (35 nmol/L, 1000 ng/dL) and taking the samples through the DHT assay. The results even on spiking with over 35 nmol/L T were less than 0.1 nmol/L DHT. The lower limit of quantitation of serum DHT in this assay was 0.43 nmol/L. All values below this value were reported as less than 0.43 nmol/L. The mean accuracy (recovery) of the DHT assay, determined by spiking steroid free serum with varying amounts of DHT from 0.43-9 nmol/L, was 101% (range, 83-114%). The intra- and interassay coefficients of variation for the DHT assay were 7.8% and 16.6%, respectively, for the adult male range, which in our laboratory was 1.06-6.66 nmol/L (30.7-193.2 ng/dL).

Serum E_2 levels were measured by a direct assay without extraction with reagents from ICN Biomedicals, Inc. The intra- and interassay coefficients of variation of E_2 were 6.5% and 7.1%, respectively, for normal adult male range (E_2 , 63–169 pmol/L, 17.1–46.1 pg/mL). The lower limit of quantitation of the E_2 was 18 pmol/L. All values below this value were reported as 18 pmol/L. The cross-reactivities of the E_2 antibody were 6.9% for estrone, 0.4% for equilenin, and less than 0.01% for all other steroids tested. The accuracy of the E_2 assay was assessed by spiking steroid free serum with an increasing amount of E_2 (18–275 pmol/L). The mean recovery of E_2 compared with the amount added was 99.1% (range, 95–101%).

Serum SHBG levels were measured by assay kits obtained from Delfia (Wallac, Inc., Gaithersburg, MD). The intra- and interassay precisions were 5% and 12%, respectively, for the adult normal male range (10.8–46.6 nmol/L). Serum FSH and LH were measured by highly sensitive and specific fluoroimmunometric assays with reagents provided by Delfia (Wallac, Inc., Gaithersburg, MD). The intraassay coefficient of variations for LH and FSH fluoroimmunometric assays were 4.3% and 5.2%, respectively, and the interassay variations for LH and FSH were 11.0% and 12.0%, respectively (adult normal male range: LH, 1.0–8.1 U/L; FSH, 1.0–6.9 U/L). For both LH and FSH assays, the lower limit of quantitation was 0.2 IU/L. All samples obtained from the same subject were measured in the same assay.

Statistical analyses

Descriptive statistics for each of the hormone levels were calculated. Before analysis, each variable was examined for its distributional characteristics and, if necessary, transformed to meet the requirements of a normal distribution. There were no significant differences between the study sites on any of the parameters; therefore, the data presented were pooled for all of the centers. The pharmacokinetic parameters for each full sampling day were determined by noncompartmental methods. The pharmacokinetics of T gel were assessed using the area under the curve from $0-24 h (AUC_{0-24})$ generated by the 24 h of multiple blood sampling for T on days 1, 30, 90, and 180. The AUC was computed using the linear trapezoid method. The average T concentration over the 24 h after gel application (C_{avg}) was calculated as the AUC₀₋₂₄ divided by 24 h.

All data in the figures and tables show the treatment mean $(\pm SEM)$

by time and/or day for each of the three groups of subjects based on the treatment from days 0–90 and for each of the five groups from days 91–180. However, because the final treatment groups (five groups) for the subjects receiving T gel were no longer randomized, statistical comparisons between groups were only performed until day 90 using the original treatment assignments (50 or 100 mg T gel or patch) as the independent groups. Comparisons between groups were performed using one-way ANOVA or the Kruskal-Wallace test (accumulation ratio, fluctuation index) followed by posttest contrasts. Analysis of the effects was performed using repeated measures ANOVA. The χ^2 test was used to compare rates. Analyses of change from day 0 to day 180 within treatment groups were performed within each of the five groups based on pattern using paired *t* tests. Comparisons resulting in $P \leq 0.05$ were considered statistically significant. SAS version 6.12 was used for all analyses (SAS Institute, Inc., Chicago, IL).

Results

Subjects

A total of 227 patients were enrolled: 73, 78, and 76 were randomized to the 50 mg/day T gel (T gel 50), 100 mg/day T gel (T gel 100), and T patch groups, respectively (Table 1). There were no significant differences in the patients' characteristics at baseline (height, weight, and previous T treatment). Thirty-five to 45% of the patients in each treatment group had primary hypogonadism (Klinefelter's syndrome, anorchia, testicular failure); 15-25% had well defined secondary hypogonadism (Kallman's syndrome, hypothalamic pituitary disease, pituitary tumor). The other patients had low serum T and normal or low normal LH levels. These were ascribed to aging (based on age >60 yr), or normogonadotropic hypogonadism. These patients did not have brain imaging to exclude hypothalamic-pituitary disease. Their primary physician did not deem that brain scans were indicated. After completion of day 90, 55 of the subjects in the T patch, 67 in the T gel 50, and 73 in the T gel 100 groups agreed to continue for another 3 months (days 91-180). The discontinuation rate (21 of 76, 27.6%) in the T patch group was higher (P = 0.0002) than those in the T gel groups (50 mg: 6 of 73, 8.2%; 100 mg: 5 of 78, 6.4%). Most of the discontinuation in the T patch group was due to adverse skin reaction based on the subjects' complaints and records. After 90 days of treatment, patients randomized initially to the T gel groups had dose adjustment if their preapplication serum T level was below 10.4 or above 34.7 nmol/L on day 60. Twenty subjects who had received 50 mg/day T gel had their dose increased to 75 mg/day; 20 who had received 100 mg/day T gel decreased their dose to 75 mg/day. The exceptions were 1 100 mg T gel patient who was adjusted to 50 mg/day and 1 50 mg T gel patient who decreased the dose to 25 mg/day. Before approval of the long-term follow-up study, 3 patients who were receiving T patch until day 90 were switched to T gel 50 from days 91-180 because of skin irritation from the patches. The data for these 3 patients as well as for the single subject who was changed from 100 to 50 mg/day were analyzed as the T gel 50 group from days 91-180. The number of subjects enrolled in the study from days 91-180 was 195, with 51 receiving T gel 50, 40 receiving T gel 75, 52 receiving T gel 100, and 52 continuing on the patch.

Treatment compliance

From days 1–90, the mean treatment compliance rates were 89.8%, 93.1%, and 96.0% for the T patch, T gel 50, and T gel 100 groups, respectively. During days 1–180 (the 6-month study period), the mean compliance rate was 86.3% for the T patch and 93.3%, 111.4%, and 96.5% for the 50, 75, and 100 mg/day T gel groups, respectively.

Pharmacokinetics of serum T concentrations (Table 2 and Fig. 1)

At baseline (day 0) average serum T concentrations over 24 h (C_{avg}) were similar in the three groups and were below the normal adult range (Fig. 1). In all three groups, during the 24-h baseline period the mean maximum T levels (C_{max}) occurred between 0800–1000 h (0–2 h in Fig. 1), and the minimum (C_{min}) T levels occurred 8–12 h later, demonstrating the expected diurnal variation of serum T.

About 35% of the patients in each group (24 of 73 subjects for the T gel 50, 26 of 78 subjects for the T gel 100, and 25 of 76 subjects for T patch) had C_{avg} within the lower normal adult male range on day 0. (The C_{avg} of serum T levels at baseline in the subjects with normal serum T on day 0 were 13.3 ± 0.4 , 13.3 ± 0.5 , and 13.0 ± 0.5 nmol/L in the T patch, T gel 50, and T gel 100 groups, respectively.) However, over 55% of these subjects had one or more serum T measurements below 10.4 nmol/L during the course of day 0. All except three of the subjects met the enrollment criterion of serum T less than 10.4 nmol/L at screening (measured at each center's laboratory). These three subjects were enrolled during a brief period when the admission serum T level was raised to 12.1 nmol/L (350 ng/dL) or less by the sponsor. The C_{avg} of serum T in the three treatment groups on day 90 after transdermal T application was different between those with low (T patch, 11.8 ± 0.8 ; T gel 50, 17.2 ± 1.2 ; T gel 100, $25.9 \pm$ 1.4 nmol/L) or normal (T patch, 14.5 \pm 0.7; T gel 50, 25.1 \pm 2.4; T gel 100, 29.5 \pm 1.9 nmol/L) baseline serum T levels. This was anticipated; however, statistical analyses with twoway ANOVA showed that the status (Cave) of serum T at baseline of more than or less than 10.4 nmol/L had no significant interaction with treatment. Thus, the differential response to transdermal T treatment was not confounded by the pretreatment serum T concentrations. Inclusion of these subjects did not influence the pharmacokinetic results of the treatment groups. Thus, in all subsequent pharmacokinetic analyses, all subjects in a treatment group were analyzed together regardless of whether their Cavy of serum T on day 0 was more than or less than 10.4 nmol/L.

On day 1 after the first application of transdermal T, serum T rose most rapidly in the T patch group, reaching a C_{max} between 8–12 h (T_{max}), plateaued for another 8 h, then declined to the baseline. Serum T rose steadily to the normal range after T gel application, with C_{max} achieved by 22 and 16 h in the T gel 50 and T gel 100 groups, respectively.

On days 30 and 90, serum T followed a similar pattern as on day 1 in the T patch group. In the T gel groups, serum T levels were at steady state, showing small and variable increases after treatment. After gel application on both days 30 and 90, the C_{avg} in the T gel 100 group was 1.4-fold higher than that in the T gel 50 group and was 1.9-fold higher than

TABLE 2. Serum T pharmacokinetic parameters after transdermal application of T gel or patch

Parameters	Parameters		T gel (50 mg/day)		T gel (100 mg/day)
Day 0					
Č _{ave} (nmol/	/L)	8.22 ± 0.55	8.22 ± 0.53		8.60 ± 0.55
C _{max} (nmol	/L)	10.89 ± 0.71	11.37 ± 0.72		11.55 ± 0.76
C _{min} (nmol	/L)	6.07 ± 0.42	6.07 ± 0.42		6.52 ± 0.44
Day 1					
C _{avg} (nmol/	/L)	16.71 ± 0.82	13.80 ± 0.63		17.82 ± 0.90
C _{max} (nmol	/L)	22.36 ± 1.13	19.42 ± 1.09		25.86 ± 1.39
C _{min} (nmol	/L)	8.04 ± 0.53	7.90 ± 0.50		8.67 ± 0.57
T _{max} (h)		11.8	22.1		16.0
Day 30					
C _{avg} (nmol/	L)	14.62 ± 0.17	19.62 ± 1.12		27.46 ± 1.18
C _{max} (nmol	/L)	19.96 ± 0.92	30.37 ± 1.99		41.60 ± 1.94
C _{min} (nmol	/L)	8.15 ± 0.50	12.52 ± 6.36		17.51 ± 0.94
T _{max} (h)		11.3	7.9		7.8
Day 90					
C _{avg} (nmol/	L)	14.46 ± 0.68	19.17 ± 1.06		27.46 ± 1.12
C _{max} (nmol	/L)	20.70 ± 1.05	29.33 ± 1.91		41.74 ± 2.31
C _{min} (nmol	/L)	7.38 ± 0.46	12.27 ± 0.63		17.37 ± 0.78
$T_{max}\left(h\right)$		8.1	4.0		7.9
Dox 180	Tnatah	T gel	T gel	T gel	T gel
Day 100	1 paten	(50 mg/day)	(50 to 75 mg/day)	(100 to 75 mg/day)	(100 mg/day)
C _{ave} (nmol/L)	14.14 ± 0.88	19.24 ± 1.18	15.60 ± 3.68	25.79 ± 2.55	24.72 ± 1.05
C_{max}^{nrg} (nmol/L)	20.04 ± 1.31	28.78 ± 1.81	23.58 ± 3.72	38.48 ± 3.72	37.55 ± 2.17
C _{min} (nmol/L)	7.69 ± 0.62	12.86 ± 0.86	10.47 ± 2.47	17.51 ± 1.85	16.82 ± 0.78
$T_{max}(h)$	10.6	5.8	2.0	7.8	7.7

 C_{avg} (nmol/L), Time-averaged concentration over 24-h dosing interval determined by AUC₀₋₂₄/24; C_{max} (nmol/L), maximum concentration during 24-h dosing interval; C_{min} (nmol/L), minimum concentration during 24-h dosing interval; T_{max} , time at which C_{max} occurred.

that in the T patch group (P = 0.0001). The variation in serum concentration over the day [fluctuation index = ($C_{max} - C_{min}$)/ C_{avg}] was similar in the three groups. On days 30 and 90, the accumulation ratio, which is defined as the increase in daily exposure to T with continued transdermal application (calculated as AUC_{day 30 or 90}/AUC_{day 1}) was 0.94 ± 0.04 for the T patch group showing no accumulation, whereas the accumulation ratios at 1.53 ± 0.09 and 1.9 ± 0.18 were significantly higher (P = 0.0001) in the T gel 50 and 100 groups, respectively. This indicates that the T gel preparations had a longer effective half-life than the T patch (Table 2 and Fig. 2).

On day 180, the serum T concentrations achieved and the pharmacokinetic parameters were similar to those on days 30 and 90 in those patients who continued in their initial randomized treatment groups (Fig. 1 and Table 2). For the patients who switched from T gel 50 or 100 to T gel 75, their C_{avg} on day 180 was 20.84 \pm 1.76 nmol/L, midway between the C_{avg} in the T gel 50 (19.24 \pm 1.18 nmol/L) and T gel 100 $(24.72 \pm 6.08 \text{ nmol/L})$ groups. Examination of Table 2 and Fig. 1 shows that the patients titrated to this T gel 75 group were not homogeneous. On day 180, the C_{avg} in the patients in the T gel 100 group who converted to 75 mg/day on day 90 was 1.7-fold higher than the C_{avg} in the patients titrated to T gel 75 from 50 mg/day. Despite adjusting the dose up by 25 mg/day in the T gel 50 to 75 group, the C_{avg} remained lower than for those remaining in the 50 mg group. In the T gel 100 to 75 group, the Cavy became similar to those achieved by patients remaining in the T gel 100 group without dose titration.

The increase in AUC_{0-24 h} on days 30, 90, and 180 from the pretreatment baseline (net AUC_{0-24 h}) showed dose proportionality. The mean for the net AUC_{0-24 h} from day 0 to day

30 or 90 was about 1.7-fold higher for T gel 100 than for T gel 50 patients (T gel 50: day 30, 268 \pm 28; day 90, 263 \pm 29 nmol/L·h; T gel 100: day 30, 446 \pm 30; day 90, 461 \pm 27 nmol/L·h). A 4.3 nmol/L (125 ng/dL) mean increase in the serum T C_{avg} level was produced by each 25 mg/day of T gel. The increases in AUC_{0-24 h} from the pretreatment baseline achieved by the T gel 100 and T gel 50 groups were approximately 2.9- and 1.7-fold higher than those resulting from application of the T patch (day 30, 154 \pm 18; day 90, 157 \pm 20 nmol/L·h).

The preapplication serum T levels in the T patch group remained at the lower limit of the normal range throughout the entire treatment period. Serum T levels after T gel application reached steady state at about 1–2 days after the initial application (24). Thereafter, the mean serum T levels remained at about 17–20 nmol/L in the T gel 50 group and about 22–30 nmol/L in the T gel 100 group (Fig. 2, *upper panel*).

Pharmacokinetics of serum free T concentration

At baseline (day 0), serum free T C_{avg} was similar in all three groups (T patch, 167 \pm 14; T gel 50, 154 \pm 14; T gel 100, 150 \pm 13 pmol/L) and was at the lower limit of the adult male range (121–620 pmol/L). The detailed pharmacokinetic parameters of serum free T on days 1, 30, 90, and 180 mirrored those of serum total T as described above (data not shown). Similar to the total T results, the free T C_{avg} achieved by the T gel 100 group was 1.4- and 1.7-fold higher than those in the T gel 50 and T patch groups, respectively (P = 0.001).

The preapplication mean free T levels throughout the treatment period in all three groups were within the normal



FIG. 1. Serum T concentrations (mean \pm SE) before (day 0) and after transdermal T applications on days 1, 30, 90, and 180. Time 0 h was 0800 h, when blood sampling usually began. On day 90, the dose in the subjects applying T gel 50 or 100 was up- or down-titrated if their preapplication serum T levels were below or above the normal adult male range, respectively. In this and subsequent figures the *dotted lines* denote the adult male normal range, and the *dashed lines* and *open symbols* represent subjects whose T gel dose were adjusted.

range, with the T gel 100 group maintaining higher free T levels than both the T gel 50 and T patch groups (Fig. 2, *middle panel*). The calculated percent free T (free T/T \times 100) remained between 1.6–2.2% before and throughout the transdermal T treatment period. Exogenous T replacement did not significantly alter the percent free T in any of the treatment groups (Fig. 2, *lower panel*).

Serum DHT concentrations

The pretreatment mean serum DHT concentrations were between 1.24-1.45 nmol/L, which were near the lower limit of the normal range (1.06-6.66 nmol/L) and were not different among the three groups (Fig. 3, upper panel). After T patch application mean serum DHT levels rose to about 1.3-fold above the baseline, whereas serum DHT increased to 3.6-fold (within the normal range) and 4.8-fold (at the upper limit of the normal range) above the baseline after application of T gel 50 and 100 (P = 0.0001), respectively, throughout the 180 days. Examination of the DHT to T ratio (Fig. 3, middle panel) showed that this ratio was not significantly altered in the T patch group (P = 0.078), whereas in the T gel 50 and 100 groups, the DHT to T ratio increased significantly from a baseline of 0.2 to between 0.23-0.29 and 0.29-0.33, respectively, during the treatment period (P = 0.0001 for both groups). The mean serum total androgen levels (calculated as the sum of serum T + DHT levels for each time point) achieved by T gel 100 throughout the treatment period were

1.4- and 2.5-fold higher than those in the T gel 50 (\sim 20 nmol/L) and T patch (\sim 10 nmol/L) groups, respectively (*P* = 0.0001; Fig. 3, *lower panel*). Adjustment of the T gel dose on day 90 did not significantly affect the serum DHT levels, DHT/T ratios, or total androgen levels.

Serum E_2 concentrations

The baseline mean serum E_2 levels were at the lower normal range and were not different in the three treatment groups. After transdermal T application, mean serum estradiol increased to stable levels by an average of 9.2% in the T patch during the treatment period, 30.9% in the T gel 50 group, and 45.5% in the T gel 100 group (P = 0.001; Fig. 4).

Serum SHBG concentrations

The serum SHBG levels were similar and within the adult male range in the three treatment groups at baseline. After T replacement, serum SHBG levels showed a small decrease in all three groups (P = 0.0046; data not shown), which was most marked in the T gel 100 group (baseline, 26.6 ± 2.0; day 90, 23.6 ± 2.7; day 180, 24.0 ± 1.7 nmol/L; P = 0.0095).

Suppression of serum gonadotropin levels

Because of the wide variability in the baseline serum LH and FSH levels, these were expressed as the percent change from baseline in response to T replacement (Fig. 5). The mean

FIG. 2. Preapplication serum T (*upper panel*), free T (*middle panel*), and percent free T (*lower panel*) concentrations during daily treatment with T gel or patch from days 1-90 (*left panel*) and days 90-180 (*right panel*). On day 90, the dose in the T gel groups was changed in some subjects, as described in Fig. 1. \blacktriangle , T patch, \blacksquare , T gel $50; \bigcirc$, T gel $100; \Box$, T gel 50 to $75; \bigcirc$, T gel 100 to 75.



percent suppression of serum LH levels was least in the T patch group (between ~30–40%), intermediate in the T gel 50 group (between ~55–60%), and most marked in the T gel 100 group (between ~80–85%; P < 0.01). The suppression of serum FSH paralleled that of serum LH levels. In the subjects with primary hypogonadism, mean serum LH and FSH levels were suppressed to within the normal range after both doses of T gel administration, but remained above the normal range after T patch application. The suppression of serum gonadotropins occurred in all hypogonadism.

Discussion

We have shown in this study that transdermal application of this new hydroalcoholic T gel formulation (AndroGel) to a large area of skin (arms, shoulders, and abdomen) at 50 and 100 mg/day (in 5 and 10 g gel, delivering approximately 5 and 10 mg T/day, respectively) resulted in dose proportional increases in serum T in a large number of hypogonadal men. After the first application of T gel, serum T levels gradually climbed to reach a maximum level after 48–72 h, as shown in our previous report (24). On repeated application, as illustrated by the pharmacokinetics, parameters on days 30, 90, and 180 remained remarkably similar and steady serum T levels were maintained, with small and variable peaks of serum T after each application. The T levels achieved with the T patch showed little evidence of accumulation (accumulation ratio, \sim 1) with repeated application. The accumulation ratios were higher in both T gel groups (1.5–1.9) on day 30, consistent with the longer lasting elevations of serum T. With continued application of T gel, the accumulation rates showed no further increases, suggesting no further accumulation on days 90 and 180.

Dose titration of T gel to 75 mg was initiated after day 90 in the hypogonadal men who had serum T levels above or below the normal range. Because of study design there was



FIG. 3. Preapplication serum DHT concentration (*upper panel*), DHT/T ratio (*middle panel*), and DHT and T concentrations (*lower panel*) during daily transdermal T treatment from days 1-90 (*left panel*) and days 90-180 (*right panel*). \blacktriangle , T patch; \blacksquare , T gel 50; \heartsuit , T gel $100; \Box$, T gel 50 to $75; \bigcirc$, T gel 100 to 75.

no dose adjustment within the T patch group. Increasing the number of T patches to three or four a day could have resulted in increases in the mean serum T concentrations (16), but might have led to an even higher dropout rate because of skin irritation in some subjects. The patients who were converted from the T gel 50 to 75 mg/day, despite increasing the dose by 50%, had average serum T levels lower than those remaining in the T gel 50 group. It is uncertain whether these lower responders to T gel might be less compliant or are biologically different. The former may be possible in some individuals, as about one third of the subjects had a lower mean compliance rate of 80%, and the average serum T levels

attained were related to the mean compliance rate. Alternatively, some patients might have low absorption and high clearance of T either in the basal state or after induction by exogenous T. Downward titration of the T gel dose from 100 to 75 mg/day was effective in decreasing the mean serum T level in the group by 15% and lowering the serum T concentration to the normal range in 16 of 19 of these hypogonadal men.

The present study examined a new transdermal open system, T gel, together with the available standard closed T patch system. A placebo group was not included because of ethical problems associated with withdrawing or delaying T 200

Serum Estradiol

60

90

Day

90

120

150

180

Serum Estradiol (pmol/L) 150 100 50 30 0 LH Change from baseline(%) LH Change 20 0 -20 -40 -60 -80 -100 FIG. 5. Percent change in serum LH 0 30 (upper panel) and FSH (lower panel) FSH Change from baseline(%) **FSH** Change 20 0 -20 -40





from baseline values after transdermal T replacement therapy from days 1-90 (*left panel*) and days 90-180 (*right panel*). \blacktriangle , T patch; \blacksquare , T gel 50; \bigcirc , T gel 100; □, T gel 50 to 75; ○, T gel 100 to 75.

replacement in hypogonadal men for a prolonged 6-month study period. Despite a relatively higher dropout rate, pharmacokinetic data obtained from this large group of hypogonadal men treated with this T patch were similar to those previously reported (14, 15).

Serum free T levels rose after transdermal T gel or T patch

application, paralleling those of serum T. The percent free T did not change significantly with T treatment. The results were corroborated by the small decreases, probably not clinically significant, in serum SHBG observed after transdermal T replacement in all three groups. The results indicated that when T is administered by the transdermal route, the lack of the first pass effect of the liver resulted in minor, if any, decreases in SHBG.

T gel application resulted in mean serum DHT that tripled after application of 50 mg T gel and rose nearly 5-fold with 100 mg T gel treatment. As 5α -reductase is present in nongenital skin (25), the increase in DHT/T ratios in the 100 and 50 mg gel groups could be explained by the higher conversion in the skin of T to DHT as a result of the large area of skin surface exposed to T in the gel groups compared with the very small area of skin exposed to the T patch. Increased DHT/T ratios have been observed with the transdermal scrotal patch, where even greater DHT/T ratios were noted (11-13). DHT is a potent androgen that is not back-convertible to T or aromatizable to E₂. Serum levels of T and DHT are not equivalent in all aspects of biological action, but certainly both have major actions on multiple androgen-dependent target organs. The biological impact of the moderately greater increase in DHT after T gel application is unclear other than its additive effect on total androgen action. Serum E₂ levels showed small and proportionate increases after transdermal T application that may be important for the known beneficial effects of estrogens on serum lipid levels, vascular endothelium reactivity, and bone resorption.

The biological activity of the T replacement in the hypogonadal men was evidenced by the consistent suppression of serum gonadotropin levels in the patients after transdermal T applications. The suppression of gonadotropins was proportional to the serum T levels achieved by the T patch or T gel. The marked and consistent suppression of gonadotropins observed after T gel 100 treatment suggested that such a modality of T delivery could be used in a male contraceptive regimen.

All patients were diagnosed to have male hypogonadism by their primary physician. In each of the three treatment groups, the same proportion (\sim 30–35%) of subjects had subnormal serum T levels at screening (assayed at each center's clinical laboratory), but their average serum T levels over 24 h were within the normal range when studied at baseline (on another day and assayed at the central laboratory). Serum T in a population of men is to a great extent a continuum. The selection of men that had serum T levels below 10.4 nmol/L at screening would inevitably allow some subjects to have serum T above this arbitrarily defined threshold (approximately <2 SD below the mean for young adult men) on subsequent measurements. The admission criterion requiring a serum T concentration of 10.4 nmol/L or less is arbitrary and necessary for the design of a clinical study; however, there is no definite evidence that there is a threshold level of T at which biological response changes. The well known intrasubject variability from day to day and the differences between T assays using different reagents and methods might account for this discrepancy between screening and baseline levels. It is also not uncommon in clinical practice that on repeat serum T measurements, some hypogonadal

patients would have serum T levels that fluctuate in and out of the statistical normal range. In practice, if symptomatic, many if not most of these men received androgen replacement therapy. The situation for assessment of pharmacokinetic parameters after administration of naturally occurring substances (e.g. T) poses different problems from those after administration of non-naturally occurring substances in the body. Ultimate serum levels attained in dynamic closed loop endocrine systems are complex and include integration of T levels (with endogenous serum T decreasing while serum T rises from exogenous administration), the characteristics of the formulation, the generic and individualized metabolic factors, and the duration of treatment. Although serum T levels attained in the groups with low or normal baseline levels were different, statistical analyses showed that the relative response to T transdermal treatment was not affected by the initial value. Thus, inclusion of these subjects did not influence the treatment comparison.

We conclude that transdermal T gel application can efficiently elevate serum T and free T levels in hypogonadal men into the mid to upper normal range within the first day of application, achieve steady state within a few days, and maintain serum T levels with once daily repeated applications. Although serum DHT/T ratios were raised after T gel applications, these ratios remained within the normal range. Serum E_2 levels were increased, and gonadotropin levels were suppressed in proportion to serum T levels. The pharmacokinetic profile and the dose proportionality observed after T gel application indicate that this transdermal delivery system may provide dose flexibility and serum T levels from the low to the high normal adult male range.

Acknowledgments

We thank Barbara Steiner, R.N., B.S.N.; Carmelita Silvino, R.N.; the nurses at the General Clinical Research Center (Harbor-University of California-Los Angeles Medical Center, Torrance, CA); Emilia Cordero, R.N. (V.A. Medical Center, Houston, TX); Tam Nguyen (The Johns Hopkins University, Baltimore, MD); Nancy Valler (V.A. Medical Cen-ter, Salem, VA); Janet Gilchriest (V.A. Puget Sound Health Care System, Seattle, WA); Helen Peachey, R.N., M.S.S. (University of Pennsylvania Medical Center, Philadelphia, PA); Mike Shin and Cheryl Franklin-Cook (Duke University Medical Center, Durham, NC); K. Todd Keylock (The Chicago Center for Clinical Research, Chicago, IL); Brenda Fulham (West Coast Clinical Research, Van Nuys, CA); Shari L. DeGrofft (Urology Research Options, Aurora, CO); Mary Dettmer (Center for Health Studies, Cleveland, OH); Jessica Bean and Maria Rodriguez (South Florida Bioavailability Clinic, Miami, FL); George Gwaltney, R.N. (Diabetes and Glandular Disease Clinic, P.A., San Antonio, TX); Peggy Tinkey (Northeast Indiana Research, Fort Wayne IN); Bill Webb (MultiMed Research, Providence, RI); and Linda Mott (Alabama Research Center, L.L.C., Birmingham, AL) for study coordination, and other support staff of each study center for their dedicated effort in conducting these studies. F. Ziel, M. D. (Kaiser Permanente Southern California) referred many patients to Harbor-University of California-Los Angeles Medical Center for this study. We thank A. Leung, H.T.C.; S. Baravarian, Ph.D.; Vince Atienza, B.Sc.; Magdalene Que, B.Sc.; Joy Whetstone, B.Sc.; Stephanie Griffiths, M.Sc.; Maria La Joie, B.Sc.; and Ellen Aquino, B.Sc., for their skillful technical assistance with many hormonal assays; Laura Hull, B.A., for data management and graphical presentations; and Sally Avancena, M.A., for preparation of the manuscript.

References

1. Bhasin S, Gabelnick HL, Spieler JM, Swerdloff RS, Wang C. 1996 Pharmacology, biology and clinical applications of androgen. New York: Wiley Liss.

- 2. Wang C, Swerdloff RS. 1997 Androgen replacement therapy. Ann Med 29:365–70.
- Nieschlag E, Behre HM. 1998 Testosterone: action, deficiency, substitution, 2nd Ed. Berlin: Springer.
 Wang C, Swerdloff RS. 1999 Androgen replacement therapy, risks and ben-
- wang C, Swerdioff KS. 1999 Androgen replacement therapy, hisks and benefits. In: Wang C, ed. Male reproductive function. Boston: Kluwer; 157–172.
 McClellan KJ, Goa KL. 1998 Transdermal testosterone. Drugs. 55:253–258.
- Jordan WP. 1997 Allergy and topical irritation associated with transdermal testosterone administration: a comparison of scrotal and non-scrotal transdermal systems. Am J Contact Dermat. 8:108–113.
- Jordan WP Jr, Atkinson LE, Lai C. 1998 Comparison of the skin irritation potential of two testosterone transdermal systems: an investigational system and a marketed product. Clin Ther. 20:80–87.
- Wilson DE, Kaidbey K, Boike SC, Jorkasky DK. 1998 Use of topical corticosterol pretreatment to reduce the incidence and severity of skin reactions associated with testosterone transdermal delivery. Clin Ther. 20:229–306.
- Yu Z, Gupta SK, Hwang SS, Kipnes MS, Mooradian AD, Snyder PJ, Atkinson LE. 1997b Testosterone pharmacokinetics after application of an investigational transdermal system in hypogonadal men. J Clin Pharmacol. 37:1139–1145.
- Yu Z, Gupta SK, Hwang SS, Cook DM, Duckett MJ, Atkinson LE. 1997a Transdermal testosterone administration in hypogonadal men: comparison of pharmacokinetics at different sites of application and at the first and fifth days of application. J Clin Pharmacol. 37:1129–3118.
- Findlay JC, Place V, Snyder PJ. 1989 Treatment of primary hypogonadism in men by the transdermal administration of testosterone. J Clin Endocrinol Metab. 68:369–373.
- Cunningham GR, Cordero E, Thornby JI. 1989 Testosterone replacement with transdermal therapeutic systems. Physiological serum testosterone and elevated dihydrotestosterone levels. JAMA. 261:2525–2530.
- Nieschlag E, Bals-Pratsch M. 1989 Transdermal testosterone. Lancet. 1:1146–1147.
- Meikle AW, Mazer NA, Moellmer JF, Stringham JD, Tolman KG, Sanders SW, Odell WD. 1992 Enhanced transdermal delivery of testosterone across nonscrotal skin produces physiological concentrations of testosterone and its metabolites in hypogonadal men. J Clin Endocrinol Metab. 74:623–628.

- Meikle AW, Arver S, Dobs AS, Sanders SW, Rajaram L, Mazer NA. 1996 Pharmacokinetics and metabolism of a permeation-enhanced testosterone transdermal system in hypogonadal men: influence of application site–a clinical research center study. J Clin Endocrinol Metab. 81:1832–1840.
- Brocks DR, Meikle AW, Boike SC, Mazer NA, Zariffa N, Audet PR, Jorkasky DK. 1996 Pharmacokinetics of testosterone in hypogonadal men after transdermal delivery: influence of dose. J Clin Pharmacol. 36:732–739.
- Wilson DE, Meikle AW, Boike SC, Failes AJ, Etheredge RC, Jorkasky DK. 1998 Bioequivalence assessment of a single 5 mg/day testosterone transdermal system versus two 2.5 mg/day systems in hypogonadal men. J Clin Pharmacol. 38:54–59.
- Arver S, Dobs AS, Meikle AW, Caramelli KE, Rajaram L, Sanders SW, Mazu NA. 1997 Long-term efficacy and safety of a permeation-enhanced testosterone transdermal system in hypogonadal men. Clin Endocrinol (Oxf). 47:727–737.
- Behre HM, von Eckardstein S, Kliesch S, Nieschlag E. 1999 Long term substitution of hypogonadal men with transcrotal testosterone over 7–10 years. Clin Endocrinol (Oxf). 50:629–635.
- Snyder PJ, Peachey H, Hannoush P, et al. 1999 Effect of testosterone treatment on bone mineral density in men over 65 years of age. J Clin Endocrinol Metab. 84:1966–1972.
- Snyder PJ, Peachey H, Hannoush P, et al. 1999 Effect of testosterone treatment on body composition and muscle strength in men over 65 years of age. J Clin Endocrinol Metab. 84:2647–2653.
- 22. Sitruk-Ware R. 1989 Transdermal delivery of steroids. Contraception. 39:1–20.
- Wang C, Iranmanesh A, Berman N, et al. 1998 Comparative pharmacokinetics of three doses of percutaneous dihydrotestosterone gel in healthy elderly men–a clinical research center study. J Clin Endocrinol Metab. 83:2749–2757.
- Wang C, Berman N, Longstreth JA, et al. 2000 Pharmacokinetics of transdermal testosterone gel in hypogonadal men: application of gel at one site versus four sites. J Clin Endocrinol Metab. 85:964–969.
- Russell DW. 1993 Tissue distribution and ontogeny of steroid 5α reductase isozyme expression. J Clin Invest. 92:903–910.
- Wang C, Swerdloff RS. 1999 Male contraception in the 21st century. In: Wang C, ed. Male reproductive function. Norwell: Kluwer; 303–319.