2015

Relationship between serum testosterone and fracture risk in men: a comparison of RIA and LC-MS/MS

Thach S. Tran
Jacqueline R. Center
Markus J. Seibel
John A. Eisman

University of Notre Dame Australia, john.eisman@nd.edu.au

Mark M. Kushnir

See next page for additional authors

Follow this and additional works at: http://researchonline.nd.edu.au/med_article

Part of the Medicine and Health Sciences Commons

This article was originally published as:

Original article available here:
http://www.clinchem.org/content/61/9/1182
Background: Serum testosterone can be measured by LC-MS/MS and RIA. We investigated whether the testosterone–fracture relationship was affected by the method of measurement.

Methods: We measured total testosterone (TT) by LC-MS/MS (TTLC-MS/MS) and RIA (TTRIA) in serum samples collected from 602 men whose incident fractures had been continuously ascertained by x-ray reports from 1989 to 2010. We measured bone mineral density (BMD) by dual-energy x-ray absorptiometry. The association between TT and fracture risk was assessed by the Cox proportional hazards model, taking into account the effect of age and BMD.

Results: Mean TTLC-MS/MS was higher than TTRIA by 27 ng/dL (95% CI 13–41). The concordance correlation coefficient between TT LC-MS/MS and TTRIA was 0.72 (95% CI 0.68–0.76). The Deming regression equation linking the 2 measurements was ln(TTLC-MS/MS/10) = 0.87 + 0.87 ln(TTRIA/10). The hazard ratio of fracture per SD decrease in TT was 1.32 (95% CI 1.12–1.54) for TTLC-MS/MS and 1.23 (1.06–1.43) for TTRIA. The correlation between predicted probabilities of fracture by TTLC-MS/MS and TTRIA was r = 0.96, with the mean difference being 0.01% (95% CI −6.1% to 6.2%). Slightly more patients were classified as having hypogonadism if TTRIA was used (29% vs 26%).

Conclusions: The concordance between LC-MS/MS and RIA in the measurement of serum TT was moderate. Moreover, the magnitude of association between testosterone and fracture risk in older men was largely unaffected by the method of measurement.

© 2015 American Association for Clinical Chemistry
correlation could translate into an accurate classification of hypogonadism for an individual.

From an epidemiologic point of view, measurements of TT by RIA methods are preferable given their relatively lower cost and faster throughput. Nevertheless, no data directly compare the ability of TT concentration to predict future fracture risk as measured by RIA vs LC-MS/MS. This study was designed to determine whether the fracture–testosterone relationship was affected by the method of measurement at the population and individual levels.

Materials and Methods

SETTING AND PARTICIPANTS

The study was part of the ongoing Dubbo Osteoporosis Epidemiology Study, for which study design and protocols have been described in detail previously (3). Briefly, through the electoral roll and via media campaign, all men and women ≥60 years old as of June 30, 1989, living in Dubbo (a regional city of 32000 predominantly white people in New South Wales, Australia) were invited to participate in the study. The age and sex distribution of the Dubbo population closely resembled that of the general Australian population. The study was approved by St Vincent’s Hospital Ethics Review Committee, Sydney. Written informed consent was obtained from all participants.

DATA COLLECTION

A nurse coordinator interviewed participants by administering a structured questionnaire to obtain anthropometric variables and other data including age, history of fracture after 50 years of age, history of falls in the preceding 12 months, lifestyle factors, and calcium intake. Bone mineral density (BMD) was measured at the lumbar spine and femoral neck by dual-energy x-ray absorptiometry (Lunar DPX-L; GE-Lunar). The same densitometer was used throughout the study, and the CV for the BMD measurements was 1.3% and 5.3% at the lumbar spine and femoral neck, respectively.

Incident fractures were continuously ascertained from 1989 to 2010 by review of x-ray reports from all 3 radiological services for the entire Dubbo area. A study coordinator determined the circumstances surrounding each fracture by phone call after each fracture. The analysis included only low-trauma and nonpathologic fractures with a definite report of fracture. Fractures were excluded from the analysis if they had clearly resulted from major trauma (motor vehicle accidents) or underlying diseases (cancer or Paget disease) or were fractures of digits, skull, or cervical spine.

LABORATORY MEASUREMENTS

We measured TT concentrations by LC-MS/MS \( (TT_{LC-MS/MS}) \) and RIA \( (TT\text{RIA}) \) in 602 (69.4%) of the 868 men who had been participating by July 2004 and were followed up at biennial intervals. Men with and without a serum sample \((n=259)\) were comparable with regard to baseline characteristics, such as age, weight, height, and BMD.

Most participants had nonfasting blood samples collected in the morning in plain tubes, which were centrifuged at 14000g. Serum samples were removed and stored in 2-mL Eppendorf tubes at \(−80 \text{ °C} \) until analysis. Serum TT concentrations were measured by LC-MS/MS \((9)\) at ARUP Laboratories. The same samples were also measured by a commercial RIA (Delfia; PerkinElmer, Wallac Oy) at the Bone Biology Laboratory, ANZAC Research Institute, Sydney. The limit of quantification and method imprecision for TT were 3.0 ng/dL and <10% (LC-MS/MS) and 8.6 ng/dL and 6.8% (RIA), respectively. Intraassay CVs were 5.6% and 4.5% for LC-MS/MS and RIA, respectively.

Serum concentrations of sex hormone–binding globulin (SHBG) were determined by RIA (Delfia) at the Bone Biology Laboratory, ANZAC Research Institute, Sydney, with CVs of 10.2%, 5.3%, and 8.3% at high (14.6 \( \mu \text{g/mL} \)), mid-range (6.4 \( \mu \text{g/mL} \)), and low (2.2 \( \mu \text{g/mL} \)) concentrations, respectively. Estradiol concentrations were measured by LC-MS/MS \((10)\) at ARUP Laboratories. The limit of quantification for estradiol was 1.5 \( \text{pg/mL} \) (5.5 pmol/L), and method imprecision was <10%.

STATISTICAL METHODS

We determined the extent of between-method agreement in TT by comparing the measured values of TT obtained by RIA with those obtained by LC-MS/MS as evaluated by Deming regression \((11)\) and Bland–Altman plot \((12)\). We used the \( \kappa \) correlation coefficient \((13)\) to quantify the concordance in clinical diagnosis of male hypogonadism, with a threshold of 300 ng/dL (10.4 nmol/L) \( (1)\).

We used the Cox proportional hazards model to assess the association between TT and fracture risk. In this method, the time to fracture was the outcome, and measurements of TT were predictors. Two independent models were considered: \(a\) fracture as a function of TT\text{RIA} and \(b\) fracture as a function of TT\text{LC-MS/MS}. In each model, we assessed the magnitude of association by the hazard ratio and 95% CI per SD decrease in TT concentration. The hazard ratio was further adjusted for known risk factors, such as age, body weight, femoral neck BMD, prior fracture, dietary calcium intake, SHBG concentration, and smoking status. Femoral neck BMD was considered in the model because it was less likely than lumbar spine BMD to be affected by degenerative changes. Because the distributions of testosterone con-
centrations, SHBG, estradiol, and dietary calcium intake were skewed, we applied a natural logarithmic (ln) transformation of observed values with the formula: ln(x + c), where x is the original value and c is a normalized constant. A correlation of the predicted probability of fracture obtained from the 2 predictive models was determined. All statistical analyses were performed with the R statistical environment on a Windows platform (14).

Results

BETWEEN-METHOD AGREEMENT OF RIA AND LC-MS/MS MEASUREMENTS

We evaluated data from 602 men aged 73 (6) years [mean (SD)] whose serum TT concentrations were measured by both RIA and LC-MS/MS (Table 1). Approximately 66% of men were age ≥70. TT concentrations were weakly correlated with serum estradiol (Pearson correlation coefficient 0.45 for TTLC-MS/MS and 0.35 for TTRIA) and SHBG (0.35 for TTLC-MS/MS and 0.36 for TTRIA). Serum TTLC-MS/MS and TTRIA were 430 (206) ng/dL [14.9 (7.1) nmol/L] and 403 (196) ng/dL [14 (6.8) nmol/L], respectively. The median (interquartile range) of serum TTLC-MS/MS was 398 (296 to 555) ng/dL [13.8 (10.3–19.2) nmol/L], and of serum TTRIA, 386 (280 to 525) ng/dL [13.4 (9.7–18.2) nmol/L]. Mean TTRIA was lower than that of TTLC-MS/MS by 27 ng/dL (0.9 nmol/L) [95% CI 13–41, P (paired t-test) = 0.0002]. The concordance correlation coefficient between the 2 methods of measurement was 0.72 (95% CI 0.68–0.76). There was no evidence that the between-method difference was systematically related to the means (Fig. 1). The Deming regression equation describing the relationship between TT concentration measured by LC-MS/MS and RIA was ln(TTLC-MS/MS) = 0.87 ln(TTRIA) + 0.87. This equation suggested that the RIA method overestimated TT concentrations by 13% compared with the LC-MS/MS method (R² = 0.54). Free testosterone concentrations were estimated from TT and SHBG concentrations with the empirical algorithm proposed by Sartorius et al. (15).

DIAGNOSTIC CONCORDANCE AT THE INDIVIDUAL LEVEL

With the criterion of TT ≤300 ng/dL (10.4 nmol/L), 156 men (26%) were classified as testosterone deficient by LC-MS/MS (Table 2). With the same criterion, the prevalence of testosterone deficiency was 29% (n = 176) by RIA. However, the concordance in the diagnosis between 2 methods of measurement was modest, with a φ coefficient of 0.55 and κ statistic of 0.57 (95% CI 0.47–0.62). Among 156 men with testosterone deficiency by TTLC-MS/MS, 45 (28.8%) were classified as eugonadal by TTRIA. On the other hand, of the 446 men classified as eugonadal by TTLC-MS/MS, 65 (14.6%) were considered testosterone deficient by TTRIA. Slightly more patients were classified as low testosterone if TTRIA was used, regardless of the serum testosterone thresholds (Fig. 2).

CONCORDANCE IN THE MAGNITUDE OF THE ASSOCIATION BETWEEN TT CONCENTRATION AND FRACTURE

During the median 7.8 years of follow-up, 112 men sustained a fragility fracture. The incidence of fracture was

Table 1. Baseline characteristics of 602 participants in the study.a

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>72.59 (5.68)</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>78.35 (12.93)</td>
</tr>
<tr>
<td>Height, cm</td>
<td>172.37 (6.32)</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>26.34 (3.79)</td>
</tr>
<tr>
<td>Dietary calcium, mg/day</td>
<td>635.62 (342.24)</td>
</tr>
<tr>
<td>Past or present smoker</td>
<td>356 (59.14)</td>
</tr>
<tr>
<td>Lumbar spine BMD, g/cm²</td>
<td>1.26 (0.22)</td>
</tr>
<tr>
<td>Lumbar spine T-score</td>
<td>0.48 (1.87)</td>
</tr>
<tr>
<td>Femoral neck BMD, g/cm²</td>
<td>0.91 (0.15)</td>
</tr>
<tr>
<td>Femoral neck T-score</td>
<td>−1.08 (1.28)</td>
</tr>
<tr>
<td>Serum TT measured by LC-MS/MS ng/dL</td>
<td>430.17 (206.18)</td>
</tr>
<tr>
<td>nmol/L</td>
<td>14.93 (7.15)</td>
</tr>
<tr>
<td>Serum free testosterone calculated from TTLC-MS/MS ng/dLb</td>
<td>5.91 (2.53)</td>
</tr>
<tr>
<td>ng/dL</td>
<td>0.20 (0.09)</td>
</tr>
<tr>
<td>Serum TT measured by RIA, ng/dL</td>
<td>403.01 (195.53)</td>
</tr>
<tr>
<td>nmol/L</td>
<td>13.98 (6.78)</td>
</tr>
<tr>
<td>Serum free testosterone calculated from TTRIA, ng/dLb</td>
<td>5.55 (2.45)</td>
</tr>
<tr>
<td>ng/dL</td>
<td>0.19 (0.08)</td>
</tr>
<tr>
<td>Serum estradiol, pg/mL</td>
<td>20.27 (9.32)</td>
</tr>
<tr>
<td>pmol/L</td>
<td>74.41 (34.21)</td>
</tr>
<tr>
<td>Serum SHBG, µg/mL</td>
<td>5.35 (2.30)</td>
</tr>
<tr>
<td>nmol/L</td>
<td>47.59 (20.46)</td>
</tr>
<tr>
<td>Previous fracture</td>
<td>112 (18.60)</td>
</tr>
</tbody>
</table>

a Data are mean (SD) or n (%).
b Free testosterone (pmol/L) = 24.00314 × TT/log10 SHBG − 0.04599 × TT² [Sartorius et al. (15)].
Fig. 1. TT measurement comparison between RIA and LC-MS/MS.
(A), Concordance correlation coefficient between \(T_{\text{TRIA}}\) concentration (x axis) and \(T_{\text{LC-MS/MS}}\) concentration (y axis). (B), Bland–Altman plot of the mean ln(TT concentration) of the 2 methods (x axis) and percentage difference in ln(TT concentration) (y axis).
3.4 per 100 person-years (95% CI 3.3–3.5). Most fractures occurred at the hip (26%), vertebrae (44%), and nonvertebrae (81%). Men with fracture had a lower baseline TT concentration than those without a fracture, measured by either LC-MS/MS [mean difference 35 ng/dL (1.2 nmol/L); 95% CI 7 to 77] or RIA [mean difference 45 ng/dL (1.6 nmol/L); 95% CI 5–85]. The hazard ratio of fracture per SD lower TT concentration was 1.23 (95% CI 1.06–1.43) by RIA and 1.41 (1.19–1.64) by LC-MS/MS. The association between TT and fracture risk remained statistically significant after adjusting for SHBG, age, BMD, weight, and lifestyle factors (Table 3). Similarly, every SD lower in calculated free testosterone concentration, computed from TTLC-MS/MS and TTria, was associated with a 25% and 22% increase in fracture risk, respectively (see Supplemental Table 1, which accompanies the online version of this article at http://www.clinchem.org/content/vol61/issue9).

On the basis of the association between TT and fracture, we estimated the probabilities of any fracture, hip fracture, vertebral fracture, and nonvertebral fracture with TTria (denoted by \( P_{\text{RIA}} \)) and TTLC-MS/MS (\( P_{\text{LC-MS/MS}} \)). Fig. 3 shows the correlation between \( P_{\text{RIA}} \) and \( P_{\text{LC-MS/MS}} \). For each fracture site, the coefficient of correlation between \( P_{\text{RIA}} \) and \( P_{\text{LC-MS/MS}} \) was consistently >0.96, and the mean difference in the predicted probability of fracture was 0.01% (95% CI –6.1% to 6.2%) for any fracture.

### Discussion

Accurate measurement of serum testosterone concentrations is essential for the diagnosis and management of male hypogonadism (1). In this study, we used TT as a
Fracture Risk and Testosterone Measurement Methods

Table 3. Association between TT and fracture risk: analysis for TT_{LC-MS/MS} and TT_{RIA}.

<table>
<thead>
<tr>
<th></th>
<th>TT_{LC-MS/MS}</th>
<th></th>
<th>TT_{RIA}</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hazard ratio (95% CI)</td>
<td>P</td>
<td>Hazard ratio (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td>Unadjusted</td>
<td>1.41 (1.19–1.64)</td>
<td>&lt;0.001</td>
<td>1.23 (1.06–1.43)</td>
<td>0.006</td>
</tr>
<tr>
<td>Adjusted for SHBG</td>
<td>1.43 (1.23–1.67)</td>
<td>&lt;0.001</td>
<td>1.37 (1.20–1.54)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Adjusted for SHBG and age</td>
<td>1.28 (1.10–1.49)</td>
<td>0.002</td>
<td>1.12 (0.98–1.30)</td>
<td>0.1</td>
</tr>
<tr>
<td>Adjusted for SHBG, age, and weight</td>
<td>1.30 (1.11–1.52)</td>
<td>0.001</td>
<td>1.18 (1.02–1.37)</td>
<td>0.02</td>
</tr>
<tr>
<td>Adjusted for age, BMD, SHBG, weight, fracture history, calcium intake, and smoking status</td>
<td>1.32 (1.12–1.54)</td>
<td>0.001</td>
<td>1.23 (1.06–1.43)</td>
<td>0.005</td>
</tr>
</tbody>
</table>

*Estimated for every SD decrease of ln(TT concentration). The equivalent SD was 206 ng/dL for back-transformation.

**Note:** TT = total testosterone; SHBG = sex-hormone-binding globulin; BMD = bone mineral density.

**Table 3.** Association between TT and fracture risk: analysis for TT_{LC-MS/MS} and TT_{RIA}.

- **Hazard ratio (95% CI)**
- **P**

**Table 3** shows the association between total testosterone (TT) and fracture risk in men, using two different methods: TT_{LC-MS/MS} and TT_{RIA}.

**Predictor of Fracture Risk:** We chose this predictor because TT measurement is clinically recommended as both the initial and the confirmed test for diagnosis of androgen deficiency. Measurement of free or bioavailable testosterone is suggested for a subgroup of patients whose TT concentrations are close to the lower limit of the reference range, although its direct measurement is costly, laborious, and not always possible in local laboratories. Furthermore, free testosterone calculated from TT and SHBG concentrations was found to have excellent predictive capability and very good performance.

**Calculation of Free Testosterone:** Calculated free testosterone has been used in most studies. At present, RIA and LC-MS/MS are commonly used for measuring TT, but it is not clear whether the discordance between the two methods could affect the classification of hypogonadism. In this study, we showed that the correlation between TT concentrations determined by RIA and LC-MS/MS was reasonably high, consistent with previous observations.

**Consistency in Classification:** The higher proportion of low TT concentrations of our participants had TT concentrations below the lower limit of quantification or statistical outliers were excluded from their analysis, leaving 101 more centralized samples than ours.

**Excellent Agreement:** The discordance between RIA and mass spectrometry has been reported to range from 0.92 for the automated multipurpose immunoassays from Bayer (Centaur) to 0.98 for a commercially available RIA kit (DPC-RIA, Core Endocrine Laboratory, Penn State University- Hershey Medical Center). All methods except the DPC-RIA, however, had an intercept of the Deming regression significantly different from that of LC-MS/MS, which is commonly considered a reference method. The inconsistency between that study and ours could be a result of differences in the population studied, RIA methods used, or method of data analysis. The participants in the study of Wang et al. (8) were much younger and less likely to have a low TT concentration than our participants. The other reason for the better agreement identified in the Wang et al. study is method of analysis. All values that were below the lower limit of quantification or statistical outliers were excluded from their analysis, leaving 101 more centralized samples than ours.
We found that, at the individual level, there was a substantial discordance in the diagnosis of hypogonadism between RIA and LC-MS/MS. Indeed, the \( \kappa \) statistic was only 0.57, a little bit higher than chance agreement (0.5).

We consider that the poor rate of agreement in the diagnosis could be due to the loss of information when a continuous variable is dichotomized (22), especially when a cutpoint is selected arbitrarily (23). Statistically, a population measure, i.e., a correlation coefficient, is not necessarily applicable to an individual, since the former focuses more on a mean than absolute values. On the other hand, an absolute value of TT is used to diagnose whether an individual is hypogonadal. Second, to fulfill clinical classification of male hypogonadism, TT concentration is dichotomized by an arbitrary value. A threshold of low TT concentration has long been controversial (1, 4). The Endocrine Society Position Statement considered a TT concentration of \( \leq 200 \) ng/dL as hypogonadal, whereas a TT concentration of 200–320 ng/dL was considered to be equivocal (4). In contrast, a TT concentration of \( \geq 346 \) ng/dL is reported as healthy in

---

**Fig. 3.** Correlation between predicted probabilities of fracture by \( \text{TT}_{\text{LC-MS/MS}} \) and \( \text{TT}_{\text{RIA}} \) concentration.
Fracture Risk and Testosterone Measurement Methods

the European Association of Urology recommendation (24). Without any strong biological plausibility background, a cutpoint of 300 ng/dL, which is the lower limit of the reference range for TT concentrations in healthy young men in some but not all laboratories (7), has been recently used as a threshold of low testosterone. In our study, a third of TT concentrations that fell within the hypogonadal range turned out to be within reference intervals on repeat measurement (25), whereas a TT concentration below the reference range in a 24-h period was reported in ≥15% of healthy young men (26). The imprecision of TT measurement, regardless of the laboratory method, emphasizes the need to confirm a low TT concentration by repeat measurement before establishing a diagnosis of hypogonadism for men with symptoms of androgen deficiency (7).

An important contribution of this study is that the discordance between TTLC-MS/MS and TTRIA had little effect on the predicted probability of fractures. We found that each SD decrease of TT concentration was associated with a 32% and 23% increase in fracture risk by LC-MS/MS and RIA, respectively. However, the correlation between predicted probability of fracture on the basis of TTTRIA and TTLC-MS/MS was excellent (R ≥ 0.96). More importantly, an absolute difference between TTTRIA and TTLC-MS/MS (27 ng/dL or 0.9 nmol/L) was only 13% of an SD of TT concentrations, making the between-method discordance very unlikely to alter the testosterone–fracture relationship.

Emerging evidence suggests that apart from testosterone, estrogen also plays important roles in skeletal maturation and mineralization in men (27–29). This has some biologic basis, since the majority of estrogens in elderly men are derived from androgens by peripheral conversion (30). In our study, TT concentrations significantly correlated with serum estradiol (r = 0.45 for TTLC-MS/MS and 0.35 for TTTRIA) and SHBG (0.35 for TTLC-MS/MS and 0.36 for TTTRIA).

Our findings should be interpreted within the context of potential strengths and weaknesses. With 602 serum samples analyzed for TT determination by both RIA and LC-MS/MS, this is the largest study ever conducted to examine the accuracy of the immunoassays with mass spectrometry as a reference method, using patient samples with available clinical information. In addition to the robust design, a sufficiently large sample is known to improve certainty of the findings. The study participants had been followed for a reasonably long period, with sufficient fracture incidence for a meaningful analysis. All fractures that occurred within the studied period were ascertained by x-ray reports. However, because serum samples were not collected consistently in the morning, a potential random measurement error could have occurred. This error, if any, would not significantly alter the findings in our elderly cohort, since a circadian rhythm of TT concentrations was reported to be absent in elderly men (31).

In summary, these data show that there is a moderate concordance in total testosterone measurements between the RIA and LC-MS/MS methods (r = 0.72). Although this moderate concordance could substantially misclassify the status of hypogonadism, it has little effect on the prediction of fracture.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: A.L. Rockwood, University of Utah.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honaria: None declared.
Research Funding: J. Center, institution funding from National Health and Medical Research Council and BUPA Health Foundation.
Expert Testimony: None declared.
Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References


Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213–20.


