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

This article was originally published as:
Tran, T. S., Center, J. R., Seibel, M. J., Eisman, J. A., Kushnir, M. M., Rockwood, A. L., & Nguyen, T. V. (2015). Relationship between serum testosterone and fracture risk in men: a comparison of RIA and LC-MS/MS. *Clinical Chemistry, 61* (9), 1182-1190.

Original article available here:
http://www.clinchem.org/content/61/9/1182

Clinical Chemistry 61:9
1182–1190 (2015)

Relationship between Serum Testosterone and Fracture Risk in Men: A Comparison of RIA and LC-MS/MS

Thach S. Tran,[1] Jacqueline R. Center,[1,2] Markus J. Seibel,[3] John A. Eisman,[1,2,4,5] Mark M. Kushnir,[6] Alan L. Rockwood,[6] and Tuan V. Nguyen[1,7,8*]

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 ($TT_{LC\text{-}MS/MS}$) and RIA (TT_{RIA}) 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 $TT_{LC\text{-}MS/MS}$ was higher than TT_{RIA} by 27 ng/dL (95% CI 13–41). The concordance correlation coefficient between $TT_{LC\text{-}MS/MS}$ and TT_{RIA} was 0.72 (95% CI 0.68–0.76). The Deming regression equation linking the 2 measurements was $\ln(TT_{LC\text{-}MS/MS} + 10) = 0.87 + 0.87 \times \ln(TT_{RIA} + 10)$. The hazard ratio of fracture per SD decrease in TT was 1.32 (95% CI 1.12–1.54) for $TT_{LC\text{-}MS/MS}$ and 1.23 (1.06–1.43) for TT_{RIA}. The correlation between predicted probabilities of fracture by $TT_{LC\text{-}MS/MS}$ and TT_{RIA} 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 TT_{RIA} 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.

Measurement of testosterone is commonly used in the diagnosis of androgen deficiency and epidemiologic studies of association. It has been recommended that a serum total testosterone (TT)[9] concentration <300 ng/dL (equivalent to 10.4 nmol/L) is considered hypogonadal *(1)*. Recent epidemiologic studies have suggested that hypogonadal concentrations of TT are associated with an increased risk of fragility fracture *(2, 3)*. In older men from the Dubbo Study, after adjustment for major risk factors, the risk of fracture was increased by 37% for every 199 ng/dL (6.9 nmol/L) decrease in TT concentrations *(3)*.

Serum testosterone can be measured by LC-MS/MS and by immunoassay, including RIA, chemiluminescent immunoassay, and enzyme-linked immunosorbent assay. Because of its lower cost and higher throughput, RIA has been widely used in research and routine clinical practice. However, the imprecision and accuracy of direct RIAs are usually suboptimal, in particular for measurements of low concentrations, which are clinically relevant (i.e., in women, children, and hypogonadal men) *(4)*. The correlation between RIA and LC-MS/MS for measuring testosterone concentrations is high, with the coefficient of correlation being >0.9 *(4–8)*. However, discordance between commercial immunoassays and mass spectrometry has been reported to be ≤5-fold at TT concentrations of ≤230 ng/dL *(7)*. Correlation is a population-level measure, not necessarily applicable to an individual. Thus it is not clear whether such a high

[1] Osteoporosis and Bone Biology Program and [4] Clinical Translation and Advanced Education, Garvan Institute of Medical Research, Sydney, Australia; [2] Clinical School, St. Vincent's Hospital, Sydney, Australia; [3] Bone Research Program, ANZAC Research Institute, The University of Sydney, Sydney, Australia; [5] School of Medicine, Sydney, University of Notre Dame Australia, Sydney, Australia; [6] ARUP Institute for Clinical and Experimental Pathology and Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT; [7] School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia; [8] Centre for Health Technologies, University of Technology, Sydney, Australia.

* Address correspondence to this author at: Professor Tuan V. Nguyen, Osteoporosis and Bone Biology Program, Garvan Institute of Medical Research, 384 Victoria St, Sydney, NSW 2010, Australia. Fax +612–9295-8241; e-mail t.nguyen@garvan.org.au.

Received April 15, 2015; accepted June 9, 2015.
Previously published online at DOI: 10.1373/clinchem.2015.242339

[9] Nonstandard abbreviations: TT, total testosterone; BMD, bone mineral density; SHBG, sex hormone–binding globulin.

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 32 000 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 ($TT_{LC\text{-}MS/MS}$) and RIA (TT_{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 14 000*g*. Serum samples were removed and stored in 2-mL Eppendorf tubes at −80 °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 μg/mL), mid-range (6.4 μg/mL), and low (2.2 μg/mL) concentrations, respectively. Estradiol concentrations were measured by LC-MS/MS *(10)* at ARUP Laboratories. The limit of quantification for estradiol was 1.5 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 κ 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_{RIA} and (*b*) fracture as a function of $TT_{LC\text{-}MS/MS}$. 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 $TT_{LC\text{-}MS/MS}$ and 0.35 for TT_{RIA}) and SHBG (0.35 for $TT_{LC\text{-}MS/MS}$ and 0.36 for TT_{RIA}). Serum $TT_{LC\text{-}MS/MS}$ and TT_{RIA} 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 $TT_{LC\text{-}MS/MS}$ was 398 (296 to 555) ng/dL [13.8 (10.3–19.2) nmol/L], and of serum TT_{RIA}, 386 (280 to 525) ng/dL [13.4 (9.7–18.2) nmol/L]. Mean TT_{RIA} was lower than that of $TT_{LC\text{-}MS/MS}$ 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 $\ln(TT_{LC\text{-}MS/MS} + 10) = 0.87 + 0.87 \times \ln(TT_{RIA} + 10)$. This equation suggested that the RIA method overestimated TT concentrations by 13% compared with the LC-MS/MS method (R^2 = 0.54). Free testosterone concentrations were estimated from TT and SHBG concentrations with the empirical algorithm proposed by Sartorius et al. *(15)*. Free testosterone calculated from TT_{RIA} was lower than that calculated from $TT_{LC\text{-}MS/MS}$ by 0.35 ng/dL [95% CI 0.17–0.53, P (paired t-test) = 0.0002].

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

Variable	Value
Age, years	72.59 (5.68)
Weight, kg	78.35 (12.93)
Height, cm	172.37 (6.32)
Body mass index, kg/m^2	26.34 (3.79)
Dietary calcium, mg/day	635.62 (342.24)
Past or present smoker	356 (59.14)
Lumbar spine BMD, g/cm^2	1.26 (0.22)
Lumbar spine T-score	0.48 (1.87)
Femoral neck BMD, g/cm^2	0.91 (0.15)
Femoral neck T-score	−1.08 (1.28)
Serum TT measured by LC-MS/MS	
ng/dL	430.17 (206.18)
nmol/L	14.93 (7.15)
Serum free testosterone calculated from $TT_{LC\text{-}MS/MS}$, ng/dL[b]	
ng/dL	5.91 (2.53)
nmol/L	0.20 (0.09)
Serum TT measured by RIA, ng/dL	
ng/dL	403.01 (195.53)
nmol/L	13.98 (6.78)
Serum free testosterone calculated from TT_{RIA}, ng/dL[b]	
ng/dL	5.55 (2.45)
nmol/L	0.19 (0.08)
Serum estradiol	
pg/mL	20.27 (9.32)
pmol/L	74.41 (34.21)
Serum SHBG	
μg/mL	5.35 (2.30)
nmol/L	47.59 (20.46)
Previous fracture	112 (18.60)

[a] Data are mean (SD) or n (%).
[b] Free testosterone (pmol/L) = 24.00314 × TT/$\log_{10}$ SHBG − 0.04599 × TT^2 [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 $TT_{LC\text{-}MS/MS}$, 45 (28.8%) were classified as eugonadal by TT_{RIA}. On the other hand, of the 446 men classified as eugonadal by $TT_{LC\text{-}MS/MS}$, 65 (14.6%) were considered testosterone deficient by TT_{RIA}. Slightly more patients were classified as low testosterone if TT_{RIA} 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

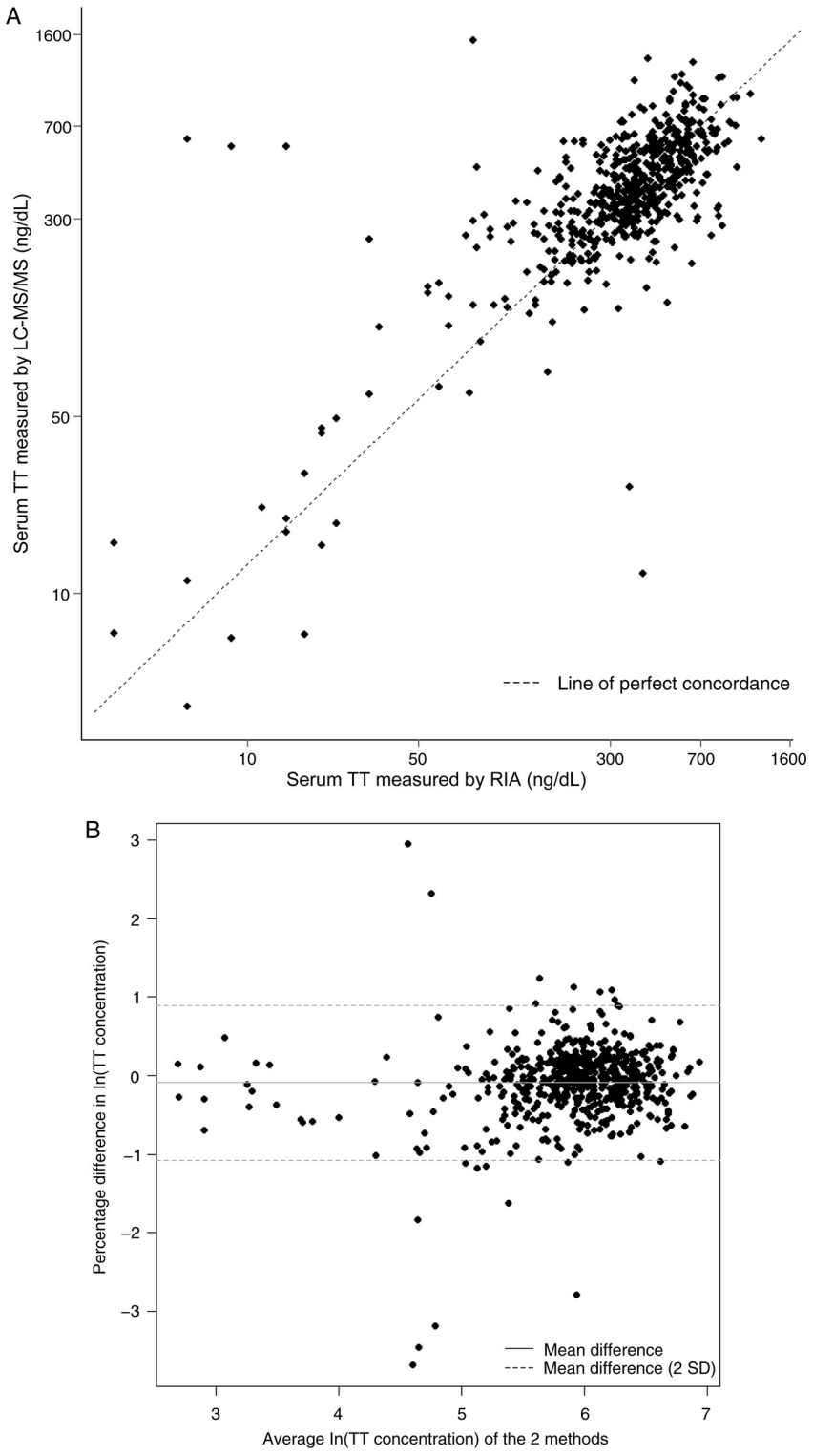

Fig. 1. TT measurement comparison between RIA and LC-MS/MS.

(A), Concordance correlation coefficient between TT_{RIA} concentration (*x* axis) and $TT_{LC\text{-}MS/MS}$ 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).

Table 2. Concordance in the diagnosis of hypogonadism between LC-MS/MS and RIA methods.[a]

Method	Actual $TT_{LC-MS/MS}$ ≤300 ng/dL	Actual $TT_{LC-MS/MS}$ >300 ng/dL
Actual TT_{RIA}		
≤300 ng/dL	111 (0.71)[a]	65 (0.15)
>300 ng/dL	45 (0.29)	381 (0.85)
Predicted $TT_{LC-MS/MS}$[b]		
≤300 ng/dL	96 (0.61)	34 (0.08)
>300 ng/dL	60 (0.39)	412 (0.92)

[a] Data are n (%).
[b] Predicted $TT_{LC-MS/MS}$ was estimated by Deming equation: $\ln(TT_{LC-MS/MS} + 10) = 0.87 + 0.87 \times \ln(TT_{RIA} + 10)$.

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 $TT_{LC-MS/MS}$ and TT_{RIA}, 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 TT_{RIA} (denoted by P_{RIA}) and $TT_{LC-MS/MS}$ ($P_{LC-MS/MS}$). Fig. 3 shows the correlation between P_{RIA} and $P_{LC-MS/MS}$. For each fracture site, the coefficient of correlation between P_{RIA} and $P_{LC-MS/MS}$ 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

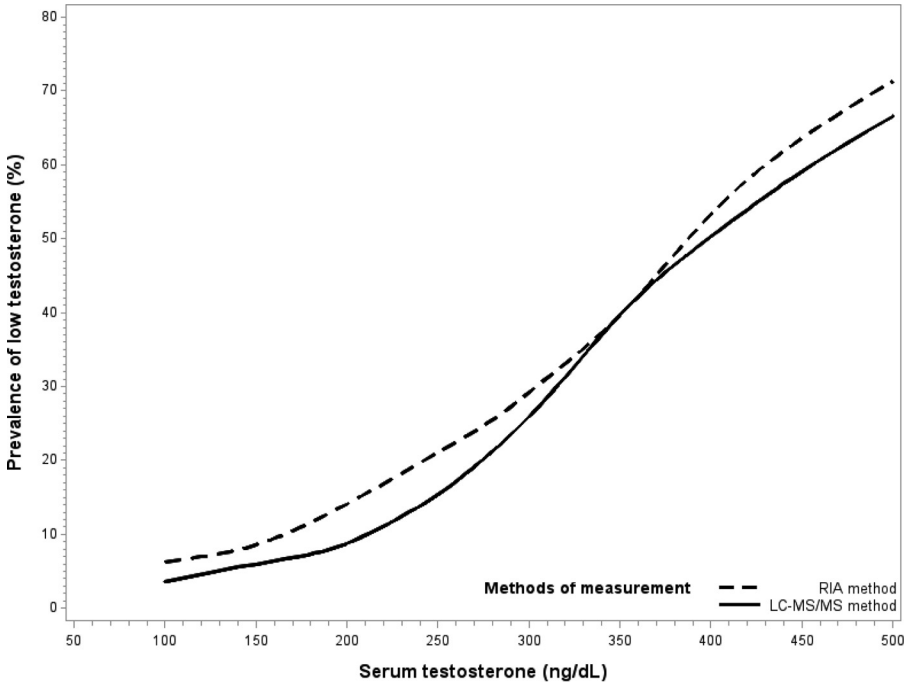

Fig. 2. Concordance in the diagnosis of low testosterone between RIA and LC-MS/MS methods by different serum testosterone thresholds.

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

	$TT_{LC\text{-}MS/MS}$		TT_{RIA}	
	Hazard ratio (95% CI)[a]	*P*	Hazard ratio (95% CI)	*P*
Unadjusted	1.41 (1.19–1.64)	<0.001	1.23 (1.06–1.43)	0.006
Adjusted for SHBG	1.43 (1.23–1.67)	<0.001	1.37 (1.20–1.54)	<0.001
Adjusted for SHBG and age	1.28 (1.10–1.49)	0.002	1.12 (0.98–1.30)	0.1
Adjusted for SHBG, age, and weight	1.30 (1.11–1.52)	0.001	1.18 (1.02–1.37)	0.02
Adjusted for age, BMD, SHBG, weight, fracture history, calcium intake, and smoking status	1.32 (1.12–1.54)	0.001	1.23 (1.06–1.43)	0.005

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

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 *(1)*. 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 *(1)*, 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 *(15)*. Therefore, 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 2 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 *(5–8)*. At the individual level, this correlation translated into some inconsistency in the classification of male hypogonadism between the 2 methods of measurement, regardless of the serum TT thresholds used. Nevertheless, at the population level, this had little effect on the predictive ability of TT in terms of fracture risk.

The discordance between RIA and mass spectrometry in the measurement of serum testosterone could be related to matrix effects and the functional sensitivity of each method. Interfering substances or matrix effects could substantially affect the imprecision of immunoassays without an extraction and chromatography component. Because 98% of circulating testosterone binds to serum proteins *(16)*, certain serum compounds that are not removed from serum, especially SHBG, could interfere with the no-extraction immunoassays *(17, 18)*. The discordance could also result from antibody cross-reactivity, inadequate limit of detection, or poor functional sensitivity of the immunoassays *(7)*. Moreover, the current testosterone immunoassays that use testosterone analogs as standards have not been fully validated or standardized *(19)*.

Excellent agreement between TT_{RIA} and $TT_{LC\text{-}MS/MS}$ has been reported *(8)*, with correlation coefficients ranging 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. Wang et al. *(8)* recruited 122 men, ages 18–68 years; meanwhile, all 602 elderly men in our study were age ≥60 years old, with a mean age of 73 years. Aging is known to be associated with decline in serum testosterone *(20, 21)*. More importantly, whereas 29% of our participants had TT concentrations of ≤300 ng/dL, only 25 hypogonadal men in Wang's study had sera collected before testosterone therapy, a rate of 20%. The higher proportion of low TT concentrations contributes positively to the between-method discordance, since imprecision of RIAs increases at lower TT concentrations *(4, 6, 7)*. The mean ratio of the concentrations of TT_{RIA} to $TT_{LC\text{-}MS/MS}$ was 1.06 but varied between 0.5 and 2.5 at low TT concentrations *(6)*. Similarly, disagreement between commercial immunoassays and mass spectrometry has been reported to be ≤5-fold at TT concentrations <230 ng/dL *(7)*. The other possible 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.

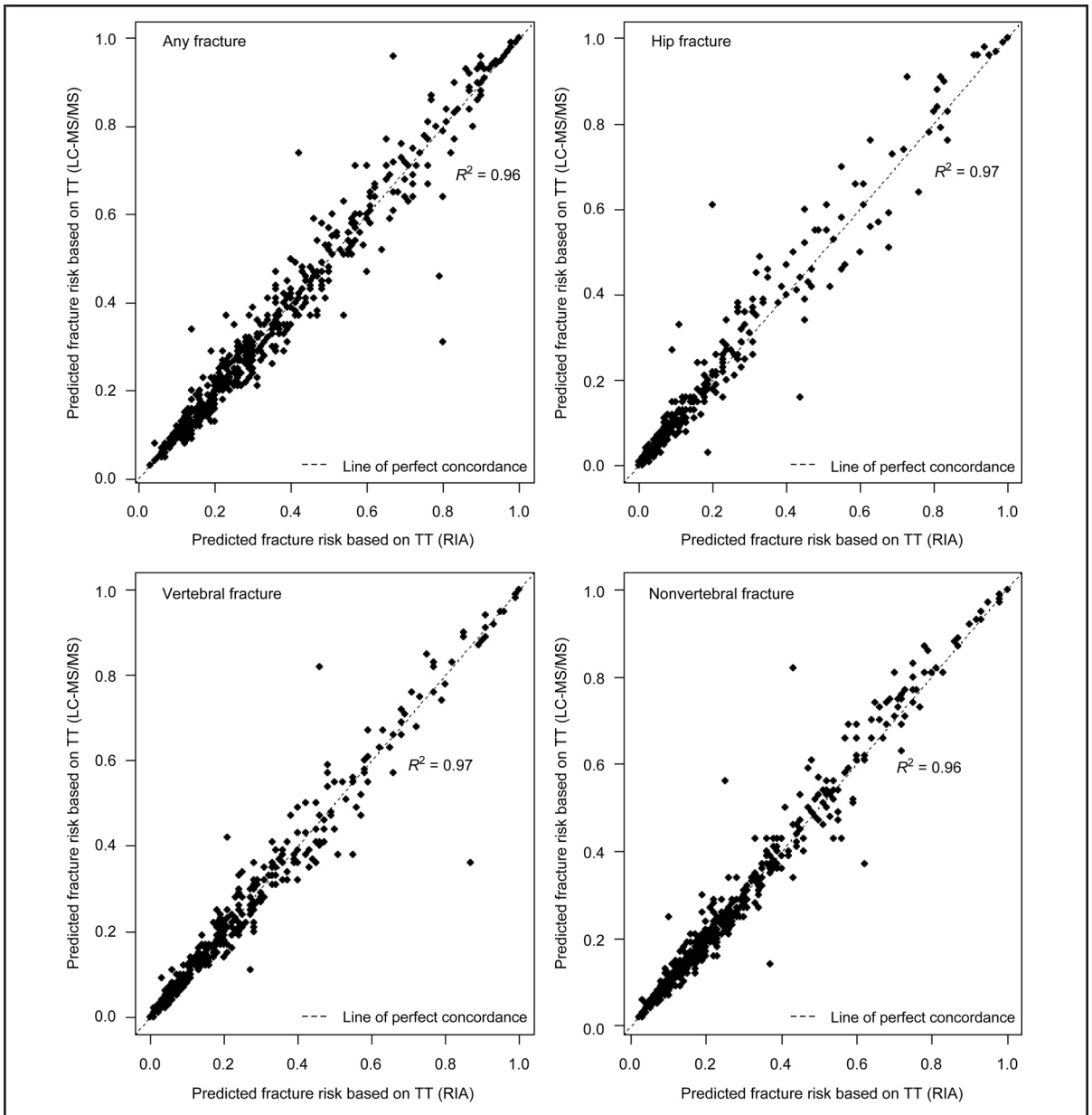

Fig. 3. Correlation between predicted probabilities of fracture by $TT_{LC\text{-}MS/MS}$ and TT_{RIA} concentration.

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 κ 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 ≤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 ≥346 ng/dL is reported as healthy in

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 *(1)*, 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 *(1)*.

An important contribution of this study is that the discordance between $TT_{LC\text{-}MS/MS}$ and TT_{RIA} 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 TT_{RIA} and $TT_{LC\text{-}MS/MS}$ was excellent ($R \geq 0.96$). More importantly, an absolute difference between TT_{RIA} and $TT_{LC\text{-}MS/MS}$ (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 $TT_{LC\text{-}MS/MS}$ and 0.35 for TT_{RIA}) and SHBG (0.35 for $TT_{LC\text{-}MS/MS}$ and 0.36 for TT_{RIA}).

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.
Honoraria: 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

1. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder JP, Swerdloff RS, Montori VM. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2010;95:2536–59.
2. Cawthon PM, Ensrud KE, Laughlin GA, Cauley JA, Dam TT, Barrett-Connor E, et al. Sex hormones and frailty in older men: the osteoporotic fractures in men (MrOS) study. J Clin Endocrinol Metab 2009;94:3806–15.
3. Meier C, Nguyen TV, Handelsman DJ, Schindler C, Kushnir MM, Rockwood AL, et al. Endogenous sex hormones and incident fracture risk in older men: the Dubbo Osteoporosis Epidemiology Study. Arch Intern Med 2008;168:47–54.
4. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. J Clin Endocrinol Metab 2007;92:405–13.
5. Hsing AW, Stanczyk FZ, Belanger A, Schroeder P, Chang L, Falk RT, Fears TR. Reproducibility of serum sex steroid assays in men by RIA and mass spectrometry. Cancer Epidemiol Biomarkers Prev 2007;16:1004–8.
6. Shiraishi S, Lee PW, Leung A, Goh VH, Swerdloff RS, Wang C. Simultaneous measurement of serum testosterone and dihydrotestosterone by liquid chromatography-tandem mass spectrometry. Clin Chem 2008;54:1855–63.
7. Taieb J, Mathian B, Millot F, Patricot MC, Mathieu E, Queyrel N, et al. Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women, and children. Clin Chem 2003;49:1381–95.
8. Wang C, Shiraishi S, Leung A, Baravarian S, Hull L, Goh V, Lee PW, et al. Validation of a testosterone and dihydrotestosterone liquid chromatography tandem mass spectrometry assay: interference and comparison with established methods. Steroids 2008;73:1345–52.
9. Kushnir MM, Rockwood AL, Roberts WL, Pattison EG, Bunker AM, Fitzgerald RL, Meikle AW. Performance

characteristics of a novel tandem mass spectrometry assay for serum testosterone. Clin Chem 2006; 52:120–8.
10. Kushnir MM, Rockwood AL, Bergquist J, Varshavsky M, Roberts WL, Yue B, et al. High-sensitivity tandem mass spectrometry assay for serum estrone and estradiol. Am J Clin Pathol 2008;129:530–9.
11. Linnet K. Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. Clin Chem 1998;44:1024–31.
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
13. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213–20.
14. R Development Core Team (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. IRBN 3–900051-07–0.
15. Sartorius G, Ly LP, Sikaris K, McLachlan R, Handelsman DJ. Predictive accuracy and sources of variability in calculated free testosterone estimates. Ann Clin Biochem 2009;46:137–43.
16. Dunn JF, Nisula BC, Rodbard D. Transport of steroid hormones: binding of 21 endogenous steroids to both testosterone-binding globulin and corticosteroid-binding globulin in human plasma. J Clin Endocrinol Metab 1981;53:58–68.
17. Masters AM, Hahnel R. Investigation of sex-hormone binding globulin interference in direct radioimmunoassays for testosterone and estradiol. Clin Chem 1989;35:979–84.
18. Slaats EH, Kennedy JC, Kruijswijk H. Interference of sex-hormone binding globulin in the "Coat-A-Count" testosterone no-extraction radioimmunoassay. Clin Chem 1987;33:300–2.
19. Matsumoto AM, Bremner WJ. Serum testosterone assays–accuracy matters. J Clin Endocrinol Metab 2004;89:520–4.
20. Feldman HA, Longcope C, Derby CA, Johannes CB, Araujo AB, Coviello AD, et al. Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. J Clin Endocrinol Metab 2002;87:589–98.
21. Harman SM, Metter EJ, Tobin JD, Pearson J, Blackman MR. Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. J Clin Endocrinol Metab 2001;86:724–31.
22. Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. Pharm Stat 2009;8:50–61.
23. Altman DG. Statistics in medical journals: some recent trends. Statistics Med 2000;19:3275–89.
24. Nieschlag E, Swerdloff R, Behre HM, Gooren LJ, Kaufman JM, Legros JJ, et al. Investigation, treatment and monitoring of late-onset hypogonadism in males. Aging Male 2005;8:56–8.
25. Araujo AB, Esche GR, Kupelian V, O'Donnell AB, Travison TG, Williams RE, et al. Prevalence of symptomatic androgen deficiency in men. J Clin Endocrinol Metab 2007;92:4241–7.
26. Zitzmann M, Faber S, Nieschlag E. Association of specific symptoms and metabolic risks with serum testosterone in older men. J Clin Endocrinol Metab 2006;91:4335–43.
27. Bilezikian JP, Morishima A, Bell J, Grumbach MM. Increased bone mass as a result of estrogen therapy in a man with aromatase deficiency. N Engl J Med 1998;339:599–603.
28. Carani C, Qin K, Simoni M, Faustini-Fustini M, Serpente S, Boyd J, Korach KS, et al. Effect of testosterone and estradiol in a man with aromatase deficiency. N Engl J Med 1997;337:91–5.
29. Smith EP, Boyd J, Frank GR, Takahashi H, Cohen RM, Specker B, Williams TC, et al. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. N Engl J Med 1994;331:1056–61.
30. Kaufman JM, Vermeulen A. The decline of androgen levels in elderly men and its clinical and therapeutic implications. Endocr Rev 2005;26:833–76.
31. Bremner WJ, Vitiello MV, Prinz PN. Loss of circadian rhythmicity in blood testosterone levels with aging in normal men. J Clin Endocrinol Metab 1983;56:1278–81.