CLINICAL STUDY

Testosterone and cardiometabolic risk in the general population – the impact of measurement method on risk associations: a comparative study between immunoassay and mass spectrometry

Robin Haring¹,², Sebastian E Baumeister³, Matthias Nauck¹,², Henry Völzke²,³, Brian G Keevil⁴, Georg Brabant⁴ and Henri Wallaschofski¹,²

¹Institute of Clinical Chemistry and Laboratory Medicine, ²DZHK (German Centre for Cardiovascular Research), partner site, ³Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany and ⁴Department of Clinical Chemistry, University Hospital South Manchester, Manchester, UK

(Correspondence should be addressed to R Haring at Institute of Clinical Chemistry and Laboratory Medicine; Email: robin.haring@uni-greifswald.de)

Abstract

Objective: Low total testosterone (TT) serum concentrations in men have been associated with various cardiometabolic risk factors. But given error-prone immunoassays used for TT assessment, upcoming mass spectrometry methods question the validity of these risk associations. Thus, we performed the first comparative study quantifying potential differences in the association of TT with cardiometabolic risk factors between the two methods.

Methods: We used data from 1512 men aged 20–81 years, recruited for the cross-sectional population-based Study of Health in Pomerania (SHIP), Germany. TT concentrations were repeatedly measured by chemiluminescent immunoassay (CLIA, Immulite 2500) and liquid chromatography–tandem mass spectrometry (LC–MS/MS). We tested for significant differences between coefficients from CLIA- and LC–MS/MS-based multiple linear regression models associating TT with major cardiometabolic risk factors including adiposity, lipid metabolism, blood pressure, diabetic status, and inflammation.

Results: TT measurements by CLIA and LC–MS/MS yielded a Pearson correlation coefficient of 0.84. Only three of the ten tested associations for TT with cardiometabolic risk factor showed significant differences between the two measurement methods: in comparison to LC–MS/MS, CLIA-based TT assessment significantly underestimated risk associations of TT with waist circumference (β: 0.54 vs 0.63), BMI (β: 0.19 vs 0.22), and serum glucose levels (β: 0.006 vs 0.008).

Conclusion: In this comparative study, the CLIA platform showed a reasonable measurement error and yielded comparable risk associations, providing little support to measure TT concentrations in men from the general population exclusively by LC–MS/MS.

European Journal of Endocrinology

Introduction

There is a considerable interest in total testosterone (TT) serum concentrations as a predictive biomarker of cardiovascular disease (1) and mortality (2) in men. Low TT concentrations have shown predictive value in numerous observational studies with regard to incident metabolic syndrome (3), type 2 diabetes mellitus (4), and increased cardiometabolic risk factor burden (5). But key to the clinical implications of low TT concentrations in men is its measurement reliability (6, 7). Most commercially available chemiluminescent immunoassays (CLIAas), however, lack sufficient reliability in the low concentration range, which are commonly present in older and/or multi-morbid men (8). Consequently, the increased risk of misclassifying healthy men as TT deficient (false positive), as well as of missing TT-deficient men (false negative), hinders clinical decision making (9).

By contrast, the more precise mass spectroscopic procedures, including liquid chromatography–tandem mass spectrometry (LC–MS/MS), demonstrate considerably lower interassay variabil ity (10), for which reason they are increasingly considered as ‘gold standard’ for clinical and epidemiological TT assessment (11). However, the only available comparative study to date assessed serum TT concentrations of 3174 men aged 40–79 years from the European Male Aging Study, measured by a commercial immunoassay and an in-house gas chromatography (GC–MS/MS) method and proved the immunoassay sufficient for the diagnosis of male hypogonadism (12). But the effects of potential
differences between the two methods on associations of TT with cardiometabolic risk factors and outcomes are still unknown. Furthermore, there are still objections concerning the applicability of immunological TT measurements in epidemiological studies (13), despite several comparative studies proving their good reliability for observational research (14, 15). Therefore, we examined associations of TT concentrations, comparatively measured by CLIA and LC–MS/MS, with diverse cardiometabolic risk factors using a large, population-based sample of men.

Subjects and methods

Study population

The Study of Health in Pomerania (SHIP) was conducted as a population-based cross-sectional health survey in West Pomerania (16). This northeastern area of Germany comprised the three cities of Greifswald, Stralsund, and Anklam and 29 surrounding communities, with a total of 212 157 residents. A sample of 7008 adults aged 20–79 years was invited to participate in the study. A two-stage cluster sampling method was adopted for this purpose from the WHO MONICA in Germany (Augsburg) and yielded twelve 5-year age strata for both genders, each including 292 individuals in a total of 34 towns or villages. The net sample drawn for the SHIP comprised 6265 eligible individuals (3105 men) with German citizenship and main residency in the study area, whereof 4308 (2116 men) aged 20–81 years finally participated in the baseline study between 1997 and 2001 (response 68.8%). All participants gave informed written consent. A review board of independent scientists monitored the study. The study protocol is consistent with the Declaration of Helsinki, as reflected by an a priori approval of the Ethics Committee of the University of Greifswald. From the 2116 men participating in the baseline study, complete TT measurements were available in 1678 men. We excluded another 75 men due to missing covariable information; five men due to intake of sex steroids (anatomic-therapeutic-chemical (ATC) code G03), testosterone 5α-reductase inhibitors (ATC code G04CB), or sex steroid antagonists (ATC code L02B); and five men with inconsistent TT measurements (i.e., CLIA < 1.5 nmol/l and LC–MS/MS-based TT > 8.0 nmol/l) resulting in a final study sample of 1512 men.

Measures

A non-fasting blood sample was drawn from the cubital vein in the supine position between 0800 and 1900 h; serum aliquots were prepared for storage at −80 °C. Serum TT concentrations were measured from frozen serum aliquots from December 2005 to January 2006 in one central laboratory (Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Germany) using CLIA on an Immulite 2500 analyzer (Siemens Healthcare Medical Diagnostics, Bad Nauheim, Germany) (17). To assess measurement bias, an aliquot of two alternating levels of a third-party commercial control material (Bio-Rad) was included in each series for single determination. The interassay coefficient of variation was 13.2% with a measurement bias of 2.3% in the low pool and 8.9% with a measurement bias of 0.24% in the high pool. The measurement range was 0.70–55.0 nmol/l with a lower limit of detection of 0.70 nmol/l.

Complementary serum TT concentrations were measured using LC–MS/MS in one central laboratory (Department of Clinical Chemistry, University Hospital South Manchester, UK) from frozen aliquots between November 2010 and February 2011, as described previously in detail (18). The interassay coefficient of variation was 11.6% with a measurement bias of +4.0% in the low pool and 4.8% with a measurement bias of 2.4% in the high pool. The measurement range was 0.3–35.0 nmol/l with a lower limit of detection of 0.25 nmol/l (19). This method was subsequently shown to give excellent agreement when compared with an isotope dilution gas chromatography (ID-GCMS) reference measurement procedure using 58 specimens from different subject categories, with a measurement bias of 0.4% from the ID-GCMS target value (20).

A computer-assisted personal interview was conducted to collect information about age, physical inactivity (at least 1 h/week), smoking habits, alcohol consumption (mean daily alcohol consumption calculated using beverage-specific pure ethanol volume proportions), and medication use. Waist circumference (WC) was measured to the nearest 0.1 cm using an inelastic tape midway between the lower rib margin and the iliac crest in the horizontal plane, with the subject standing comfortably with weight distributed evenly on both feet. Height and weight were measured for the calculation of BMI (weight (kg)/height (m)2). After a resting period of at least 5 min, systolic and diastolic blood pressure was measured three times on the right arm of seated participants with an oscillometric digital blood pressure monitor (HEM-705CP, Omron Corporation, Tokyo, Japan). The interval between the readings was 3 min. The mean systolic and diastolic blood pressure was calculated from the second and third measurement. Serum total cholesterol (TC) and HDL concentrations were measured photometrically (Hitachi 704, Roche). Serum glucose concentrations were determined enzymatically using reagents from Roche Diagnostics (Hitachi 717, Roche Diagnostics). HbA1c was measured by HPLC (Bio-Rad Diamat). Plasma fibrinogen concentrations were determined according to Clauss using an Electra 1600 analyzer (Instrumentation Laboratory, Barcelona, Spain), and high-sensitive C-reactive protein (hs-CRP) using a
Behring Nephelometer (Dade Behring Instrumentation, Eschborn, Germany). Serum γ-glutamyltransferase (GGT) levels were measured photometrically (Hitachi 717; Roche Diagnostics).

**Statistical analyses**

Continuous variables were expressed as median (25th and 75th percentile) and categorical variables as proportions. We regressed LC–MS/MS on CLIA-based TT measurements taking into account the impression in the measurements of both the independent and dependent variables using a Deming regression model (21). Associations between TT concentrations, measured using two assays (CLIA and LC–MS/MS), and cardiometabolic risk factors (WC, BMI, TC:HDL ratio, diastolic and systolic blood pressure, glucose, HbA1c, hs-CRP, GGT, and fibrinogen) were investigated using multiple linear regression models with continuously distributed independent and dependent variables and robust (Huber/White) standard errors (22). We transformed serum glucose, GGT, and hs-CRP levels using a natural logarithm to meet the linear model assumption of normally distributed error terms. Estimates were presented as β-coefficients and corresponding 95% CIs. Differences between CLIA- and LC–MS/MS-based coefficients were tested using seemingly unrelated regressions (22).

We investigated two correction methods for linear (non-differential) random measurement error with constant variance. If covariables are measured with non-differential error, point and interval estimates are often underestimated (attenuated), wherefore this effect is referred to as regression dilution (or attenuation) bias. The first correction method fitted a classical (moment-based) measurement error model, which corrects the error-prone CLIA-based coefficients and standard errors using the estimated reliability as attenuation factor (23, 24, 25). The intra-class correlation or reliability coefficient (corresponding 95% CI) was estimated using a linear random intercept model. Because the long data format (two rows per individual with two correlated TT measurements) induces correlation within an individual, we applied random effects models to overcome this violation (23). For the second correction method (i.e., regression calibration), we fitted a linear random intercept model, saving the predicted TT values (TT) and a second linear regression of each continuous cardiometabolic risk factor on TT, and covariables (23). We used 100 bootstrap replications to derive standard errors for this two-stage procedure. Two-sided P values of <0.05 were considered statistically significant. SigmaPlot 12.0 (San Jose, CA, USA) was used for the Deming regression and Stata/MP 12.1 (Stata Corp., College Station, TX, USA) for all other statistical analyses.

**Results**

The study population of 1512 men had a median age of 51.0 years (25th–75th percentiles: 37.0–64.0 years). Further baseline characteristics of the study population are presented in Table 1. The high correlation (Pearson correlation coefficient of 0.84) between TT measurements by CLIA and LC–MS/MS is illustrated in Fig. 1. Comparing cross-sectional associations between TT and cardiometabolic risk factors including obesity, lipid metabolism, blood pressure, diabetic status, and inflammation by measurement method (Table 2), we found no significant differences between the regression coefficients of CLIA and LC–MS/MS-based measurements.

**Table 1** Baseline characteristics of the study population (n=1512 men). Values shown are expressed as median (Q1, Q3) or percentage. Total testosterone may be converted to ng/dl by multiplying these given values by 28.82.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median (Q1, Q3) or Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>51 (37, 64)</td>
</tr>
<tr>
<td>Total testosterone by CLIA (nmol/l)</td>
<td>14.8 (11.5, 18.9)</td>
</tr>
<tr>
<td>Total testosterone by LC–MS/MS (nmol/l)</td>
<td>15.8 (12.4, 20.0)</td>
</tr>
<tr>
<td>Current smoker (%)</td>
<td>34.3</td>
</tr>
<tr>
<td>Physically inactive (%)</td>
<td>59.6</td>
</tr>
<tr>
<td>Alcohol consumption (g/day)</td>
<td>10.9 (2.9, 25.8)</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>95.1 (87.5, 102.6)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>27.4 (24.9, 29.9)</td>
</tr>
<tr>
<td>Total cholesterol (mmol/l)</td>
<td>5.7 (4.9, 6.4)</td>
</tr>
<tr>
<td>HDL cholesterol (mmol/l)</td>
<td>1.3 (1.1, 1.5)</td>
</tr>
<tr>
<td>Total:HDL cholesterol ratio</td>
<td>4.5 (3.6, 5.8)</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg)</td>
<td>85.0 (78.0, 93.5)</td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg)</td>
<td>140.5 (129.0, 153.0)</td>
</tr>
<tr>
<td>Non-fasting glucose (mmol/l)</td>
<td>5.4 (5.0, 6.0)</td>
</tr>
<tr>
<td>HbA1c (%)</td>
<td>5.4 (5.0, 5.8)</td>
</tr>
<tr>
<td>γ-Glutamyltransferase (μmol/l)</td>
<td>0.46 (0.30, 0.75)</td>
</tr>
<tr>
<td>Highly sensitive C-reactive protein (mg/l)</td>
<td>1.3 (0.7, 2.8)</td>
</tr>
<tr>
<td>Plasma fibrinogen (g/l)</td>
<td>2.8 (2.5, 3.3)</td>
</tr>
</tbody>
</table>

The Deming graph for total testosterone measured by CLIA vs LC–MS/MS is illustrated in Fig. 1. We investigated two correction methods for linear (non-differential) random measurement error with constant variance. If covariables are measured with non-differential error, point and interval estimates are often underestimated (attenuated), wherefore this effect is referred to as regression dilution (or attenuation) bias. The first correction method fitted a classical (moment-based) measurement error model, which corrects the error-prone CLIA-based coefficients and standard errors using the estimated reliability as attenuation factor (23, 24, 25). The intra-class correlation or reliability coefficient (corresponding 95% CI) was estimated using a linear random intercept model. Because the long data format (two rows per individual with two correlated TT measurements) induces correlation within an individual, we applied random effects models to overcome this violation (23). For the second correction method (i.e., regression calibration), we fitted a linear random intercept model, saving the predicted TT values (TT) and a second linear regression of each continuous cardiometabolic risk factor on TT, and covariables (23). We used 100 bootstrap replications to derive standard errors for this two-stage procedure. Two-sided P values of <0.05 were considered statistically significant. SigmaPlot 12.0 (San Jose, CA, USA) was used for the Deming regression and Stata/MP 12.1 (Stata Corp., College Station, TX, USA) for all other statistical analyses.

![Figure 1 Deming graph for total testosterone measured by CLIA vs LC–MS/MS. CLIA, chemiluminescent immunoassays; LC–MS/MS, liquid chromatography–tandem mass spectrometry.](www.eje-online.org)
coefficients of LC–MS/MS- or CLIA-based TT measurements, except for WC, BMI, and glucose levels. In comparison with LC–MS/MS, CLIA-based TT measurements significantly underestimated point and interval estimates for the inverse associations of TT with WC (β = -0.54 vs -0.63), BMI (β = -0.19 vs -0.22), and glucose (β = -0.006 vs -0.008). CIs around the regression coefficients for the LC–MS/MS-based associations were also wider reflecting the small loss in power using the error-prone CLIA-based TT measurements. The association grid in Fig. 2 illustrates the level of statistical significance for differences in phenotype associations dependent on TT measurement method.

We applied two correction methods for the biased CLIA-based coefficients (Table 2, columns 4 and 5). For the classical measurement error model, we estimated a reliability coefficient for CLIA-based TT measurements of 0.8227 (95% CI: 0.806–0.839) and divided point estimates and standard errors by this reliability estimate. This simple strategy already corrected coefficients from six of ten investigated associations with continuous cardiometabolic risk factors. It also increased the precision and power, which helped to detect potential associations. The remaining coefficients, namely TCHDL ratio, systolic blood pressure, HbA1c, and fibrinogen, were slightly overcorrected (Table 2, column 4). Regression calibration, the second correction method, overcorrected the coefficients and standard errors for most outcomes, for TCHDL ratio and systolic blood pressure even substantially (Table 2, column 5).

### Discussion

This is the first study investigating the differences in associations of TT with cardiometabolic risk factors depending on the measurement method (CLIA vs LC–MS/MS). Our study offers three principal findings: first, we found a high correlation between CLIA- and LC–MS/MS-based TT measurements. Secondly, we observed no consistent differences with regard to risk associations between the two methods, with CLIA-based TT measurements underestimating only three of the ten assessed outcomes. Finally, we identified a simple correction method for the slightly underestimated CLIA-based TT measurements.

Putting these findings into context, the present correlation coefficient of 0.84 is well in the range between 0.77 and 0.87, reported from previous comparative studies between multiple immunoassays, and LC–MS/MS (26) measured serum TT in 50 male adult patients from an endocrinology unit (mean age: 44.7 years, range: 19–71 years) using eight non-isotopic immunoassays, two isotopic immunoassays, and GC–MS/MS. Although all the immunoassays showed correlation coefficients with GC–MS/MS between 0.86 and 0.97, the immunoassays
underestimated TT giving mean results 12% below those obtained by GC–MS/MS (26). Another comparative study of six commonly applied immunoassays with LC–MS/MS among 101 adult men (age range: 18–60 years) reported high interclass correlation coefficients between 0.92 and 0.97 and overestimation as well as underestimation bias from immunoassays (27). Finally, between 0.92 and 0.97 and overestimation as well as underestimation bias from immunoassays (27). Finally, the comparison of the classical measurement error model and regression calibration suggests that the former better corrects for biases in the coefficients (i.e., strength of association) and precision (power to detect significant associations) than the latter. Thus, the measurement error model, using the estimated reliability as a correction factor, provides reasonable bias corrections for CLIA-based regression coefficients. However, the potential application of the reliability ratio for the correction of CLIA-based TT measurements warrants further validation in additional, independent epidemiological studies. But apart from analytical bias, potentially even more important are the various pre-analytical factors influencing TT serum concentrations, including fasting status, smoking, blood collection technique, or sample storage, and their occasionally considerable impact on measurement variability (6, 7).

These findings have important implications for clinical practice and future epidemiological research. First, our findings from the first population-based comparative study of risk associations between TT measurement methods show that immunoassays are sufficiently reliable for epidemiological research in men from the general population. But as a prerequisite for properly validated immunoassay platforms, we would like to reiterate the importance of the recommendation to establish adult male reference ranges in each individual laboratory (27). Therefore, we previously published according to in-house TT reference ranges for the present population-based sample of healthy men from the general population (17). And recently published LC–MS/MS-based TT reference ranges from three geographically distinct populations indicate another step to provide potential reference ranges for healthy men from the general population (32). However, in light of the age-related physiological decline in TT concentrations and interlaboratory differences in measurement accuracy, the application of arbitrarily defined, fixed TT cutoffs for risk assessment (e.g., <10.4 nmol/l) across different epidemiological studies lead to misclassification and erroneous findings. Thus, others and we advocate the application of age-specific
percentile cutoffs (e.g., <10th percentile in each 10-year age group) (33, 34, 35).

Secondly, given the limited resources in epidemiological research, the gold standard MS-based TT measurement could be reliably approximated by the reliability coefficient derived from our study, lessening the need to perform the expensive MS method. And finally, it should be realized that a well-validated automated immunoassay platform is capable of proving accurate, cost-effective, convenient, rapid, and as a result easily available routine clinical TT measurements. In a comparative study between seven LC–MS/MS and one GC–MS/MS assays, the mean percent difference to the reference assay was 11%, and the between-assay variability was the major contributor to the overall variability (10). This study shows that, although the variability of MS-based TT measurements is substantially smaller than that reported for immunoassays, interlaboratory standardization of MS assays remains a major effort to reduce the variability of TT measurement results. Accordingly, the need for standardization/harmonization of immunological as well as MS-based steroid assays has been recognized (13, 36). The recently started ‘Steroid Hormone Standardization Project’ performs various activities to standardize TT measurements to establish reference ranges and to improve the comparability of TT measurements independently of method, time, and place (37). However, because of the substantial imprecision of available automated immunoassays in the low concentration range of TT, a diagnosis of androgen deficiency should not be made based on a single TT measurement, it rather requires at least two low TT measures as the hormone levels often fluctuate into the hypogonadal range in eugonadal men. Thus, in accordance with the latest Endocrine Society’s guideline, low TT concentrations should be, independent of the measurement method, verified by repeated testing in an early morning blood sample (11).

Strengths and limitations

Strengths of this study include the large and population-based study sample of 1517 men covering a broad age range and state-of-the-art LC–MS/MS-based TT assessment. Potential limitations may arise from the cross-sectional nature of this study, an exclusive study sample of Caucasian, healthy, adult men, restricting the generalizability of our results to other ethnicities, age or population subgroups, and our finding that the risk associations of the CLIA-based TT measurements in some cases overestimated the risk of the MS-based TT measurements (which may be explained by measurement error in outcome variables, covariables, or differential measurement error (23)). However, despite the detected differences in outcome associations between the two methods, statistical significance does not necessarily imply clinical relevance. Thus, future clinical comparison studies should assess the prediction of hard clinical endpoints and performance measures for T therapy including diagnosis, treatment, and monitoring among male hypogonadal patients.

Conclusion

In summary, this is the first study to assess the comparative impact of different measurement methods on the associations between TT and cardiometabolic risk factors. Using a large population-based sample of men, our study showed that the CLIA platform not only yielded comparable risk associations but also a reasonable measurement error compared with LC–MS/MS. We conclude that assay performance in terms of accuracy and precision, rather than assay technology (CLIA vs LC–MS/MS), should be the criterion for choice of assay and its quality assessment. Thus, a well-validated automated immunoassay platform is sufficient for TT measurement in epidemiological research among healthy adult men.

Declaraton of interest

S E Baumeister, M Nauck, B G Keesee, and G Brabant have nothing to disclose. R Haring has received honorarium for lectures by Bayer Pharma AG. H Wallascheksofski has received research grants from Bayer Pharma AG. Novartis, Novo Nordisk, and Pfizer for research unrelated to the contents of this manuscript and serves as a member of the International and German KIMS board.

Funding

The work is part of the Community Medicine Research (CMR) net of the University of Greifswald, Germany, which is funded by the BMBF (German Ministry of Education and Research) (grant no. ZZ9603), the Ministry of Cultural Affairs, as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. The CMR encompasses several research projects that are sharing data of the population-based Study of Health in Pomerania (SHIP; http://www.medizin.uni-greifswald.de/icm/). This work is also part of the research project Greifswald Approach to Individualized Medicine (GANI_MED) funded by the BMBF and the Ministry of Cultural Affairs of the Federal State of Mecklenburg – West Pomerania (03IS2061A). This study was supported by the DZHK (German Centre for Cardiovascular Research) and by the BMBF.

References

levels are associated with increased risk of mortality in a population-based cohort of men aged 20–79. European Heart Journal 2010 31 1494–1501. (doi:10.1093/eurheartj/ehq009)


Received 19 March 2013
Revised version received 9 July 2013
Accepted 29 July 2013