Harmonized Reference Ranges for Circulating Testosterone Levels in Men of Four Cohort Studies in the United States and Europe

Thomas G. Travison,[1] Hubert W. Vesper,[3] Eric Orwoll,[4] Frederick Wu,[5] Jean Marc Kaufman,[6] Ying Wang,[4] Bruno Lapauw,[6] Tom Fiers,[7] Alvin M. Matsumoto,[8] and Shalender Bhasin[2]

[1]Institute for Aging Research, Hebrew Senior Life, and [2]Research Program in Men's Health, Aging, and Metabolism, Boston Claude D. Pepper Older Americans Independence Center, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115; [3]Centers for Disease Control and Prevention, Atlanta, Georgia 30341; [4]Oregon Health and Science University, Portland, Oregon 97239; [5]Andrology Research Unit, Centre for Endocrinology and Diabetes, Manchester Academic Health Science Centre, University of Manchester, Manchester M13 9WL, United Kingdom; [6]Department of Endocrinology and [7]Department of Clinical Chemistry, Ghent University Hospital, Ghent B-9000, Belgium; and [8]Geriatric Research, Education, and Clinical Center, VA Puget Sound Health Care System, Division of Gerontology and Geriatric Medicine, Department of Medicine, University of Washington School of Medicine, Seattle, Washington 98108

Background: Reference ranges for testosterone are essential for making a diagnosis of hypogonadism in men.

Objective: To establish harmonized reference ranges for total testosterone in men that can be applied across laboratories by cross-calibrating assays to a reference method and standard.

Population: The 9054 community-dwelling men in cohort studies in the United States and Europe: Framingham Heart Study; European Male Aging Study; Osteoporotic Fractures in Men Study; and Male Sibling Study of Osteoporosis.

Methods: Testosterone concentrations in 100 participants in each of the four cohorts were measured using a reference method at Centers for Disease Control and Prevention (CDC). Generalized additive models and Bland-Altman analyses supported the use of normalizing equations for transformation between cohort-specific and CDC values. Normalizing equations, generated using Passing-Bablok regression, were used to generate harmonized values, which were used to derive standardized, age-specific reference ranges.

Results: Harmonization procedure reduced intercohort variation between testosterone measurements in men of similar ages. In healthy nonobese men, 19 to 39 years, harmonized 2.5th, 5th, 50th, 95th, and 97.5th percentile values were 264, 303, 531, 852, and 916 ng/dL, respectively. Age-specific harmonized testosterone concentrations in nonobese men were similar across cohorts and greater than in all men.

Conclusion: Harmonized normal range in a healthy nonobese population of European and American men, 19 to 39 years, is 264 to 916 ng/dL. A substantial proportion of intercohort variation in testosterone levels is due to assay differences. These data demonstrate the feasibility of generating harmonized reference ranges for testosterone that can be applied to assays, which have been calibrated to a reference method and calibrator. (***J Clin Endocrinol Metab*** **102: 1161–1173, 2017)**

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in USA
Received 10 August 2016. Accepted 18 November 2016.
First Published Online 10 January 2017

Abbreviations: BMI, body mass index; CDC, Centers for Disease Control and Prevention; EMAS, European Male Aging Study; FHS, Framingham Heart Study; LC-MS/MS, liquid chromatography tandem mass spectrometry; MrOS, Osteoporotic Fractures in Men Study; SIBLOS, Sibling Study of Osteoporosis.

The reference range refers to the distribution of the circulating concentrations of a hormone or an analyte in a specific population (1, 2). Rigorously derived reference ranges are essential for distinguishing healthy from diseased individuals and constitute the foundation of our contemporary approach to making the diagnosis of clinical disorders.

Hypogonadism in men is a syndrome characterized by a set of symptoms and signs of androgen deficiency that occur in association with consistently low circulating testosterone levels (3). The reference ranges provide the basis for differentiating low from normal testosterone levels, and are, therefore, essential for making the diagnosis of hypogonadism. We have published reference ranges for circulating testosterone levels generated in healthy nonobese men who were participants in the Framingham Heart Study (FHS) (4); similar data have been published in other populations (5–11). However, an important unresolved question is whether the reference ranges generated in one population of men can be applied more broadly to men in other geographic regions and in other populations. The distribution of testosterone concentrations could vary in men from different regions due to interassay or interlaboratory differences, or biological or environmental factors.

The objective of this initiative of the Endocrine Society was to compare the distribution of total testosterone concentrations in epidemiologic studies that included men from different geographic regions of the United States and Europe and to generate consensus reference ranges for total testosterone levels in men. We anticipated that, notwithstanding the substantial interindividual variation in testosterone levels observed within each cohort, there would also be significant and correctable variation in mean testosterone levels between cohorts owing specifically to differences in measurement technology. We sought to minimize the influence of these systematic differences by harmonizing all measurements to a higher order standard prior to the estimation of reference ranges.

Accordingly, serum testosterone levels were measured in male participants of four epidemiologic studies: the FHS, the European Male Aging Study (EMAS), the Osteoporotic Fractures in Men Study (MrOS), and the Sibling Study of Osteoporosis (SIBLOS). Because different assays were used for measuring testosterone levels in these four epidemiologic studies and because these assays used different calibrators, the assays were cross-calibrated centrally by measuring testosterone levels in serum samples from a subset of men in each cohort in the Centers for Disease Control and Prevention (CDC) Clinical Reference Laboratory using an assay calibrated with higher order reference materials and using serum-based reference materials as additional accuracy controls. By comparing these new CDC-derived values with the original values obtained on these men from each cohort, we developed normalizing equations permitting translation from the original cohort-specific measurements to the CDC standard, and then applied them to the full sample of values in each cohort.

Because testosterone levels decline with advancing age, we first generated reference ranges in healthy nonobese young men, 19 to 39 years, as this approach based on limits derived in a healthy young population has been favored historically for analytes that exhibit clinically meaningful age-related trends, such as estradiol and bone mineral density. Because of the well-known effect of obesity on testosterone levels and on age-related change in testosterone levels, we present age-adjusted reference ranges in nonobese men, and additionally for all men, by decades of age.

Methods

Ethical approval for the study was obtained from institutional review boards for human subject research at each participating institution.

General approach

First, fasting morning serum samples obtained from 100 men from each of the four cohorts, in which testosterone levels had previously been assayed locally, were transported to the central laboratory at the CDC. These 100 men with previous assay results from the local laboratory were chosen at random to approximate the distribution of age and other factors within each of the four cohorts. At CDC, testosterone concentrations were measured on each sample using a higher order (a reference method against which other methods are compared) liquid chromatography tandem mass spectrometry (LC-MS/MS) method under the supervision of Dr. Hubert Vesper. We then developed transformational equations for each study describing the relationship between the 100 local and 100 central measurements, providing an estimate of the systematic variation in local measurements from the reference standard. These normalizing equations were applied to all testosterone levels measured in each of the four cohorts to generate harmonized values. These harmonized measurements were in turn used to derive standardized, age-specific reference ranges in each of the four cohorts and overall.

The four cohort studies

The EMAS

The EMAS recruited 3369 men, aged 40 to 79 years, at eight European centers (12, 13). The men, randomly selected from the general population, were invited for study-related assessments, including an interviewer-assisted questionnaire, performance measures, and a fasting blood test before 10:00 AM. A total of 150 men were excluded because of pituitary, testicular, or adrenal disease, or use of medications that affect sex-steroid production or action, yielding an analytic sample of 3219 men.

FHS

The original FHS cohort was established in 1948 by recruiting 5209 men and women between the ages of 30 and 62 from Framingham, Massachusetts. In 1971, the study enrolled 5124 of the original participants' adult children and their spouses, who constituted the Second Generation Cohort (Generation 2). The Generation 2 examination 7 was attended by 1625 men between 1998 and 2002. Exclusion of men with prostate cancer undergoing androgen deprivation therapy (n = 8), men receiving testosterone therapy, and men with missing testosterone data (n = 158) resulted in a sample of 1459 for Generation 2.

A Third Generation Cohort (4095 children of Generation 2, referred to as Generation 3) was established in 2002 to 2005 (14) (http://nhbli.nih.gov/about/framingham). Of the 1912 men who attended the first Generation 3 examination in 2002 to 2005, 1893 had total testosterone measurements, and 962 were ≤40 years, among whom 456 men of Generation 3 were free of cancer, cardiovascular disease, diabetes mellitus, hypertension, hypercholesterolemia, and obesity [body mass index (BMI) >30 kg/m^2] and constituted the reference sample. The men who were receiving androgen deprivation therapy or had undergone orchiectomy for prostate cancer or were taking testosterone were excluded.

The FHS combined sample was created by combining Generation 2 and Generation 3 samples. Generation 2 examination 7 was attended by 1625 men between 1998 and 2002. Exclusion of men with prostate cancer undergoing androgen deprivation therapy (n = 8), men receiving testosterone therapy, and men with missing testosterone data (n = 158) resulted in a sample of 1459 for Generation 2. This sample of 3352 men (1459 men in Generation 2 plus 1893 men in Generation 3) constituted the FHS combined sample.

The MrOS Study

MrOS, an observational study of the determinants of fracture in older men, recruited 5994 community-dwelling men ≥65 years at six US centers (15,16). Total testosterone concentration was measured on fasting, morning specimens in 1583 randomly selected men. Among these, 95 were excluded because of androgen or antiandrogen use, or orchiectomy, resulting in an analytical sample of 1488 participants.

The Belgian SIBLOS

The SIBLOS is a population-based study of healthy young men sampled in sibling pairs, who were recruited from the population registries of three semirural or suburban communities around Ghent, Belgium (17, 18). A total of 1114 men, 25 to 45 years, was recruited over 24 months. A total of 113 men were excluded because they had used medications affecting androgen status, or had disorders affecting body composition or bone metabolism. The included population of 1001 men consisted of 424 pairs of brothers, 23 families with three brothers, and 84 single participants whose brothers could not participate in the study. Among these, testosterone measurements were available for 995 men, who constituted the analytical sample. The analytic sample included 729 men who were <40 years of age and had a BMI <30 kg/m^2.

Designation of analytic samples

We conducted three independent analyses of the harmonized data.

Generation of reference ranges in healthy, nonobese (BMI <30 kg/m^2) young men

First, we selected men, 19 to 39 years, who were nonobese (BMI <30 kg/m^2) and free of major comorbidities, as described (4). Because men <40 years were available only in the FHS and SIBLOS studies, data for 1185 men meeting these criteria from these cohorts were included in this analysis.

Age-specific reference ranges in nonobese men

We computed reference ranges for individuals with BMI <30 kg/m^2 by decades of age (19 to 39, 40 to 49, 50 to 59, 60 to 69, 70 to 79, and 80 to 99 years). There were 6933 men from the four cohorts meeting this BMI criterion. The analyses were first performed within each cohort, and then the cohorts were combined to derive model-based estimates of age trends in population quantiles.

Age-specific reference ranges in all men

We computed model-based estimates of population reference ranges for all men in each age range regardless of obesity status, using combined data from all cohorts.

Hormone assays

FHS samples were obtained in the morning, after an overnight fast of ~10 hours, typically between 7:30 and 8:30 AM. The samples were frozen immediately and stored at −80°C until the time of assay. We measured total testosterone in the FHS samples using a LC-MS/MS assay, which has been described (4, 19). The lower limit of quantitation was 2 ng/dL; no sample was outside the linear range of 2 to 2000 ng/dL. The interassay coefficient of variation was 15.8% at 12.0 ng/dL, 10.6% at 23.5 ng/dL, 7.9% at 48.6 ng/dL, 7.7% at 241 ng/dL, 4.4% at 532 ng/dL, and 3.3% at 1016 ng/dL, respectively. As part of the CDC Hormone Standardization Program, quality control samples provided by the CDC were run every 3 months; the coefficient of variation in quality control samples with testosterone concentrations in 100 to 1000 ng/dL range was consistently <6%.

Total testosterone levels in the EMAS (20) and MrOS (21) samples were measured using gas chromotography tandem mass spectrometry with sensitivities of 5 ng/dL and 2.5 ng/dL, respectively. The intra- and interassay coefficients of variation in the low, medium, and high pools were 4.3%, 5.5%, and 4.9%, and 2.4%, 8.1%, and 2.5%, respectively.

Testosterone levels in the SIBLOS were measured in serum samples that were obtained between 8:00 and 10:00 AM after overnight fasting, and stored at −80°C. Testosterone was measured by LC-MS/MS using an AB Sciex 5500 triple-quadrupole mass spectrometer (AB Sciex, Toronto, Canada) and Shimadzu liquid chromatography system, and validated against an isotope dilution mass spectrometry reference method (22). The interassay coefficient of variation was 6.5% at 3 ng/dL, lower limit of quantitation 1 ng/dL, and recovery between 96% and 104%.

Cross-calibration of assays in the CDC Clinical Reference Laboratory

Approximately 100 randomly selected samples from each cohort were shipped on dry ice to the CDC, where they were stored at −70°C until analysis. Serum total testosterone levels were measured using a reference LC-MS/MS method (23). In brief, testosterone was isolated from serum by two serial

liquid–liquid extraction steps and quantified with [^{13}C] stable isotope–labeled testosterone as the internal standard. The imprecision of the method at 15.2, 228, and 886 ng/dL was 4.8%, 3.7%, and 3.8%, respectively. Agreement with established quality control limits was assessed using standard procedures (24). In addition, two serum materials with reference values assigned by an internationally recognized reference laboratory (target values: 265 and 513 ng/dL) were analyzed to assess the accuracy of each analytical run. The difference from the target values averaged 0.31%.

Statistical analysis

Harmonization of testosterone measurements across cohorts

In exploratory analyses, we used Generalized Additive Models to assess the best functional form of association between the local and the central values within each group of 100 representative individuals. Bootstrapping was used to quantify uncertainty in the estimates of these associations. Additionally, Bland–Altman analyses were used to assess the degree to which differences between local and central measurements tracked with the level of testosterone concentrations. These assessments supported a model of linear correspondence between local and central measurements after log transformation to reduce the influence of outlying values, counteract modest heteroscedasticity on the natural scale, and to insure that transformations would yield no negative values. Normalizing equations for each of the four cohorts describing the local to central transformation were generated using orthogonal Passing-Bablok regression, which has superior performance to other methods in the presence of outliers and irregularities (25). Each of the normalizing transformations was then applied to local testosterone measurements obtained from the corresponding cohort.

Derivation of estimated quantiles of testosterone distributions

For each cohort and age range, we obtained simple sample quantiles as estimates of their population counterparts using a method that is median unbiased and robust to statistical transformation (25–27). Unified estimates of age group-specific population quantiles combining all data were obtained from the four cohorts using a semiparametric growth curve model under the assumption that testosterone levels are monotonically nonincreasing with age (28). All population centiles were estimated simultaneously, permitting restriction such that centile curves do not cross (*e.g.*, estimates of the 40th percentile are no greater than those of the 50th at all ages).

Reference ranges in healthy nonobese young men

We computed the population centiles among nonobese men, 19 to 39 years, without major comorbidities in the FHS Generation 3 and the SIBLOS samples, which had recruited men in this age range. Study and age subcohort–specific reference ranges were then derived by estimation of the relevant percentiles. Consistent with the approach used for defining reference limits for many other analytes, total testosterone values <2.5th percentile were deemed low (29).

All analyses were performed using R version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Subject characteristics

The characteristics of the 9054 qualifying participants from the four cohorts are summarized in Table 1.

Harmonization of testosterone concentrations across epidemiologic studies

Exploratory generalized additive models describing the association between local and central measurements supported linear transformation of local to central measurements (Fig. 1, left panels). Visual assessment of the bootstrapped smooths overlaid on estimated mean trends also suggested that such transformations were reasonable. Bland–Altman analyses (Fig. 1, right panels) indicated some association between testosterone concentrations and the magnitude of the difference between methods, with the absolute difference between local and centralized values being greater at higher concentrations in all but the SIBLOS study. The transformations derived from the use of the Passing-Bablok procedure (Supplemental Table 1) displayed good overall agreement between central and local testosterone values in all four cohorts. The harmonization resulted in some decrease in most measurements from FHS (Fig. 1): some increase in most measurements from MrOS, with less substantial shifts in the EMAS and SIBLOS. Although we have previously observed substantial intercohort variation in locally measured testosterone levels (4), the harmonization procedure was successful in substantially reducing intercohort variation between measurements, such that they resulted in greater similarity in the distribution and means of estimated testosterone levels in men of similar ages, as can be observed in Fig. 2. For instance, locally generated values for men of age 40 to 49 years yielded mean testosterone measurements of 501, 551, and 618 ng/dL in EMAS, SIBLOS, and FHS, respectively. After harmonization, the corresponding values were 487, 494, and 471 ng/dL, respectively. Similarly, local measurements on men of age 70 to 79 years yielded mean testosterone measurements of 397, 470, and 575 ng/dL in MrOS, EMAS, and FHS, respectively, whereas after harmonization the corresponding values were 489, 455, and 438, respectively. These observations provide empirical support for the hypothesis that measurement variation is a contributor to observations of variation in age-specific estimates of mean testosterone concentrations across cohorts, and lend support for the objective of establishing reference ranges after combining harmonized data from multiple cohorts.

Distribution of testosterone levels in the reference sample of young men, 19 to 39 years, in the FHS and SIBLOS studies

The distribution of harmonized total testosterone levels in the nonobese healthy young men in the FHS

Table 1. Characteristics of 9054 Participants From Each of the Cohorts; Mean (Standard Deviation) or N (%) Shown

	EMAS (N = 3219)	FHS (N = 3352)	SIBLOS (N = 995)	MrOS (N = 1488)
Age, years	60 (11)	49 (14)	34 (6)	74 (6)
<30	—	224 (7%)	225 (23)	—
30–39	—	660 (20%)	560 (56%)	—
40–49	785 (24%)	872 (26%)	210 (21%)	—
50–59	873 (27%)	788 (24%)	—	—
60–69	799 (25%)	493 (15%)	—	447 (30%)
70–79	746 (23%)	289 (9%)	—	782 (53%)
80+	16 (0.5%)	26 (0.8%)	—	259 (17%)
BMI, kg/m^2	27.7 (4.1)	28.3 (4.7)	25.1 (3.5)	27.4 (3.7)
Obese (BMI >30 kg/m^2)	773 (24%)	964 (29%)	80 (8%)	304 (20%)
Diabetes, %	236 (7%)	274 (8%)	—	165 (11%)
Glucose, mg/dL	102 (25)	103 (23)	85 (9)	106 (27)
Systolic BP, mm Hg	146 (21)	124 (15)	126 (14)	139 (19)
Diastolic BP, mm Hg	87 (12)	77 (10)	80 (10)	N/A
Total cholesterol, mg/dL	215 (49)	193 (36)	198 (38)	193 (33)

Proportions computed with respect to nonmissing records.

Abbreviations: BP, blood pressure; N/A, not available.

Generation 3 and the SIBLOS study was remarkably similar (Table 2, top). After harmonization, the median testosterone was ~530 ng/dL, and the mean 550 ng/dL in each of the two cohorts. The 2.5th percentile values for harmonized total testosterone concentrations in healthy nonobese men in the FHS and SIBLOS studies were 265 and 264 ng/dL, respectively, and the corresponding 97.5th percentile values were 923 and 916 ng/dL, respectively. Consistent with the findings of two other studies (30, 31), the total testosterone levels were higher in nonobese healthy men than in all men (Table 2, bottom). The greater difference observed between the overall young participant samples in FHS and SIBLOS than observed in the healthy nonobese young sample is consistent with the difference in the design of these two studies; whereas some FHS participants carry diagnoses of comorbid conditions, the SIBLOS participants were screened so that the eligible participants were free of comorbid conditions.

Age-specific distribution of testosterone levels and intercohort variation

Figure 2 illustrates age range–specific testosterone levels based on harmonized measurements, plotted separately for each of the four cohorts. As noted previously, the harmonization procedure reduces but does not entirely remove the intercohort variation in mean testosterone levels. The observed cross-sectional trend suggesting decrease in testosterone concentrations with age is of lesser magnitude than the corresponding within-individual trend observed previously (31).

Table 3 describes the distribution of harmonized reference ranges by decades of age in nonobese men. Although the 2.5th and 5th percentiles generally decreased with age, this was less the case for values at the upper end of the distribution, which tended to fluctuate across the age groups. However, as expected, values in nonobese men at all percentiles tended to be greater than those derived from all (obese and nonobese) participants.

Table 4 provides age-specific estimates of the percentiles of total testosterone distribution derived from all studies combined, after harmonization, using constrained quantile regression models. As is the case with the exploratory estimates described in Table 3, we observed age-related decreases in concentrations at the lower end of the distributions, whereas the upper centiles were largely stable across the age groups. Thus, among nonobese men, the age-specific 95th percentile estimates lie in a tight range (839 to 850 ng/dL), whereas the 5th percentile estimates vary more substantially, ranging from 304 in men 19 to 39 years of age to 252 in those 70 to 79 and 218 in those 80 and above.

Discussion

These data show that the cross-calibration of assays using a higher order standard and a higher order assay in a central laboratory provides substantial reduction in intercohort variation. This suggests that measurement variation contributes to the previously observed variation in mean testosterone levels among epidemiological cohorts from different geographic regions, the substantial interindividual variation in hormone levels within any cohort notwithstanding. The distribution of harmonized total testosterone values in healthy nonobese young men was very similar between the FHS Generation 3 and the

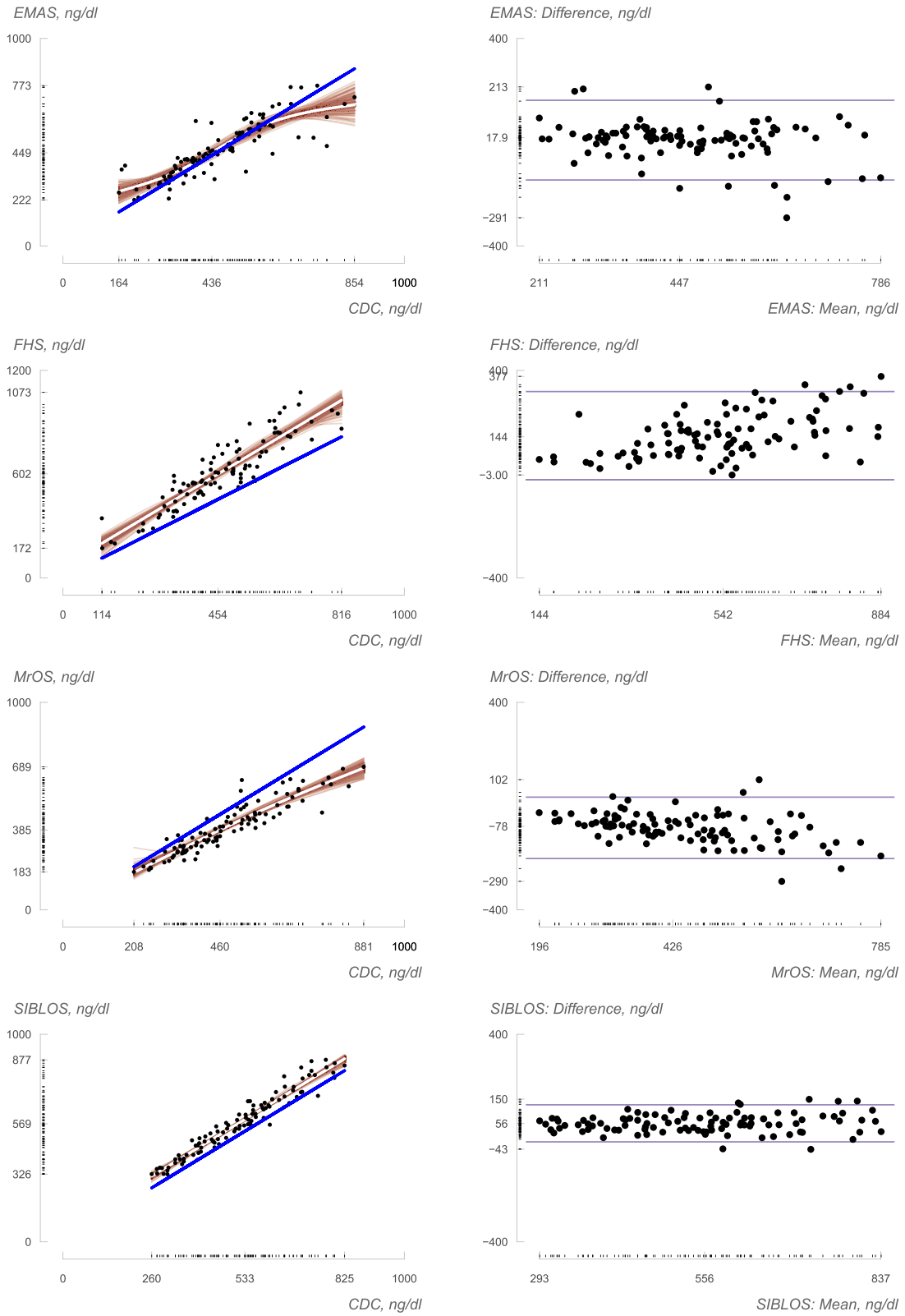

Figure 1. Relation of study-specific (local) measurements to reference standard (standardized) measurements. At left, local measurements are plotted as functions of standardized measurements (N = 100 for each study), and best fit line obtained via generalized additive model is plotted in white. The line of perfect agreement is shown in blue. Two hundred bootstrapped iterations of the generalized additive models fit (see *Methods*) are displayed in red, giving a sense of the uncertainty in the transformations. At right, Bland–Altman plots characterizing the difference (local minus central) in measurements as a function of the average of the two. Sample minimum and maximum are provided on each axis.

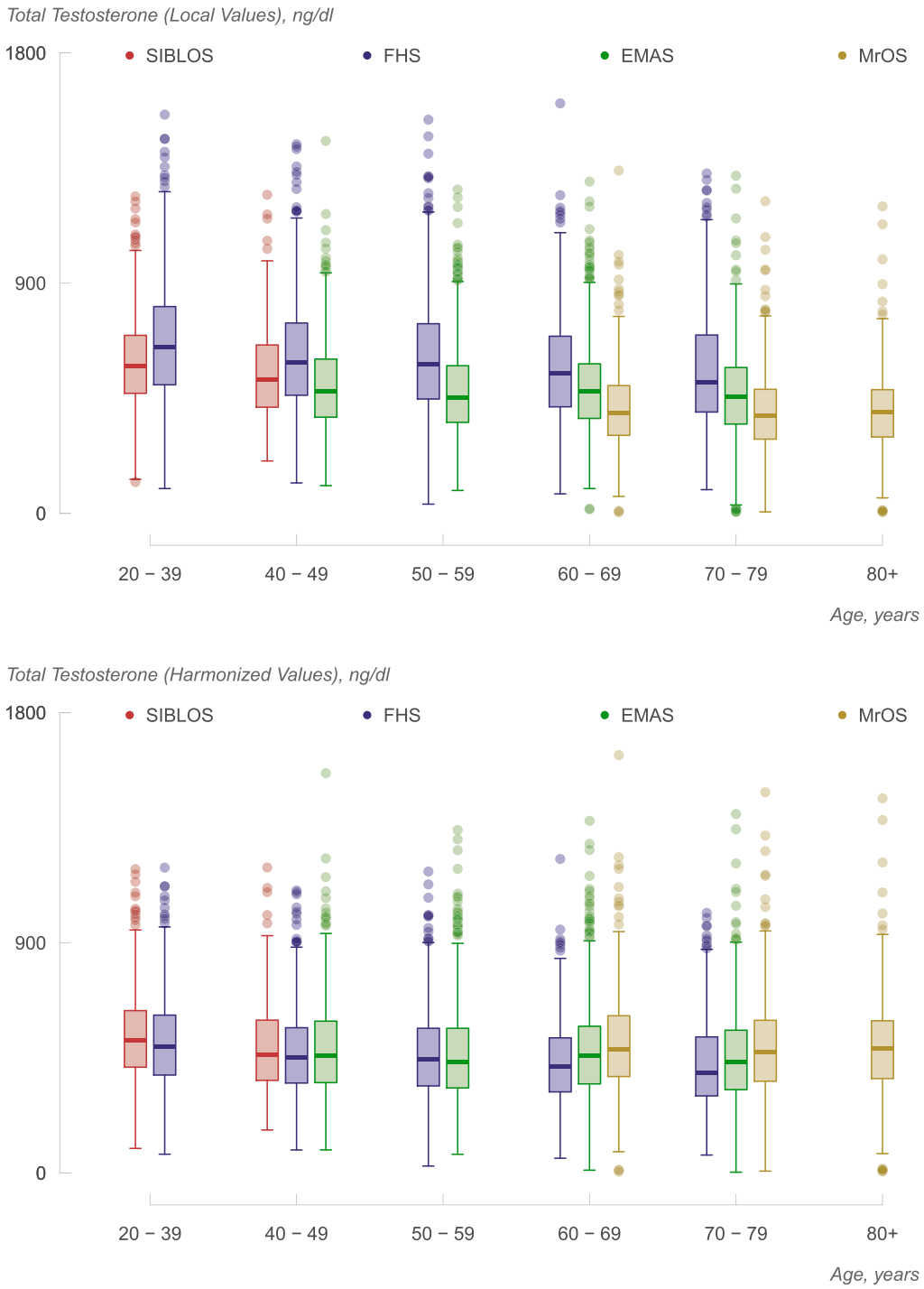

Figure 2. Box and whisker plots showing the distribution of total testosterone levels by decades of age in the four cohorts without harmonization (upper panel) and after harmonization (lower panel). The lower and upper boundaries of the box represent the 25th and 75th percentile values; the line inside the box represents the median. Independent adjustment of each study's measurements to the CDC (as shown in the lower panel) reduces interstudy variation substantially over that observed in unstandardized measurements (shown in the upper panel).

SIBLOS cohorts—2 geographically distinct cohorts. The 2.5th, 5th, 50th, 95th, and 97.5th percentile values in healthy nonobese young men were 264, 303, 531, 852, and 916 ng/dL, respectively (Table 2). We conclude that standardized hormone measurements calibrated to a higher order benchmark, such as that offered by the CDC Clinical Reference Laboratory, provide a rational and feasible approach to generating harmonized reference ranges for testosterone and possibly other analytes.

These reference ranges were derived from single morning samples and discounted the intraindividual variation in testosterone levels due to pulsatile, diurnal, and circannual secretory rhythms. Previous analyses by our groups and others have shown that early morning testosterone levels,

Table 2. Distribution of CDC-Standardized Circulating Total Testosterone Measurements Among Healthy Men of Age 19–39 Years, N = 1656

Healthy, Nonobese Young Men			
Percentile	**Framingham Heart Study (N = 456)**	**SIBLOS Study (N = 729)**	**Combined (N = 1185)**
2.5	265	264	264
5	309	301	303
10	357	344	349
25	430	426	428
50	533	529	531
75	657	639	645
90	772	775	773
95	858	846	852
97.5	923	916	916
All Young Men			
Percentile	**Framingham Heart Study (N = 871)**	**SIBLOS Study (N = 785)**	**Combined (N = 1656)**
2.5	209	250	228
5	263	280	273
10	312	332	318
25	383	413	396
50	495	518	507
75	617	635	626
90	745	767	755
95	833	844	834
97.5	883	906	895

obtained in a manner similar to that used by clinicians in practice, are associated cross-sectionally and longitudinally with symptoms and clinical outcomes (4, 13, 32, 33). The assays were performed in samples stored at −80°C. Although the stability of cholesterol levels has been demonstrated in FHS over a period of 15 years, the long-term effects of storage at −80°C on testosterone concentrations have not been clearly demonstrated.

Although the cohorts included in these analyses were diverse in morbidity, age, and geographic location, they were largely composed of men who identify as white within a US or European social context. Significant geographic and racial differences in sex-steroid levels, which have been reported in some studies (34, 35) but not in other studies (36), might have important implications for clinical decision making. It might therefore be important to develop larger investigations of multiracial, multiethnic, and more geographically diverse cohorts to confirm applicability of reference ranges to broader populations.

Several important conceptual issues remain unresolved. Should the reference range be based on a sample of healthy young men (the so-called T-score approach) or should the reference range be age adjusted (the Z-score approach)? We have provided reference ranges in a young healthy reference sample as well as by decades of age. The rationale for generating the reference range in healthy young men is similar to the use of bone mineral density T-scores for the diagnosis of osteoporosis. For analytes that exhibit substantial age-related change, such as testosterone and estradiol, it might arguably be more appropriate to derive the reference ranges in a healthy young population. Notably, results obtained in this study show a lesser age trend than that reported previously in cross-sectional analyses of men of different ages, underscoring the need for longitudinal studies of the effects of aging on sex steroid concentrations.

Another unresolved issue relates to whether the reference sample should include only the healthy nonobese men or whether it should include the entire population of men 19 to 39 years. Obesity and comorbid conditions affect circulating total testosterone concentrations (31, 37); therefore, inclusion of obese and men with comorbid conditions could distort the reference ranges. Whether the reference ranges generated in nonobese men are appropriate for use in obese men deserves further investigation. Even though men with known diagnoses of conditions or diseases associated with hypogonadism were excluded, it is possible a small percentage of individuals in these cohorts may be hypogonadal.

Historical experience with cholesterol, hemoglobin A1C, and vitamin D assays indicates that the application of reference ranges across laboratories and across geographic regions is a challenging process that requires mechanisms for standardizing assays and an understanding of biological as well as social differences in the distribution of the analyte (38, 39). The CDC Hormone

Table 3. Distribution of Total Testosterone (ng/dL) Levels by Age Among Nonobese Individuals (N = 6933) and Among All Men in Each of the Four Cohorts (N = 9050)

		Age, Years				
	Percentile	40–49	50–59	60–69	70–79	80–89
Nonobese men (N = 6933)						
EMAS nonobese						
	2.5	232	210	234	167	
	5	272	246	260	215	
	10	300	289	299	269	
	25	382	363	373	350	
	50	483	467	481	456	
	75	615	598	602	585	
	90	755	733	768	721	
	95	830	827	893	831	
	97.5	928	962	990	895	
FHS nonobese						
	2.5	213	214	214	192	
	5	263	255	241	210	
	10	304	295	264	247	
	25	382	384	345	326	
	50	473	485	445	437	
	75	600	614	555	560	
	90	719	747	678	699	
	95	807	840	781	882	
	97.5	863	900	863	948	
MrOS nonobese						
	2.5			243	196	14
	5			275	252	188
	10			331	309	282
	25			410	388	374
	50			506	494	493
	75			657	623	620
	90			789	761	786
	95			919	850	897
	97.5			1044	929	964
SIBLOS nonobese						
	2.5	244				
	5	275				
	10	323				
	25	370				
	50	483				
	75	625				
	90	760				
	95	835				
	97.5	976				
All men (N = 9050)						
EMAS all men						
	2.5	204	198	180	79	
	5	236	220	224	193	
	10	278	255	264	248	
	25	353	332	348	326	
	50	459	433	457	435	
	75	594	566	574	558	
	90	743	706	728	702	
	95	812	803	842	807	
	97.5	904	942	952	876	
FHS all men						
	2.5	203	181	189	155	
	5	232	218	211	195	
	10	270	257	248	229	
	25	351	340	316	300	
	50	451	445	416	390	
	75	568	566	528	533	
	90	689	709	647	684	

(*Continued*)

Table 3. Continued

		Age, Years				
	Percentile	**40–49**	**50–59**	**60–69**	**70–79**	**80–89**
	95	766	796	741	863	
	97.5	850	880	822	916	
MrOS all men						
	2.5			203	177	15
	5			260	228	171
	10			313	279	264
	25			377	358	369
	50			485	472	488
	75			615	597	595
	90			762	743	778
	95			887	835	851
	97.5			957	909	955
SIBLOS all men						
	2.5	208				
	5	244				
	10	294				
	25	362				
	50	464				
	75	597				
	90	752				
	95	827				
	97.5	947				

Standardization Program for testosterone is an important step to address this challenge, which will facilitate the application of these reference ranges across laboratories. The harmonized references ranges can be helpful to clinicians in facilitating clinical decision making and in improving patient care. The data reported in this work illustrate the promise and feasibility of generating reference ranges using harmonized values that can be applied across different geographic regions of the world to CDC-certified laboratories that use a common calibrator. Such calibrators for testosterone and some other analytes are now available from the National Institute of Standards and Technologies.

Table 4. Model-Based Estimates of Population Centiles for Total Testosterone Concentrations (ng/dL) Based on Data From Nonobese Men (N = 6933) and in All Men (N = 9054) in the Four Harmonized Cohorts

	Age, Years					
Percentile	**19–39**	**40–49**	**50–59**	**60–69**	**70–79**	**80–99**
All nonobese men						
2.5	267	235	219	218	218	157
5.0	304	273	256	254	252	218
10.0	344	310	297	296	292	278
25.0	424	386	374	374	372	362
50.0	531	481	477	477	477	476
75.0	643	608	605	604	604	604
90.0	774	749	749	749	749	749
95.0	850	839	839	839	839	839
97.5	929	929	929	929	926	913
All men						
2.5	229	208	192	190	190	119
5.0	273	243	222	221	220	203
10.0	318	283	262	260	259	256
25.0	396	358	341	340	340	338
50.0	507	461	446	446	446	446
75.0	626	588	573	572	572	572
90.0	755	729	720	720	720	720
95.0	834	813	812	812	812	812
97.5	902	902	902	902	902	902

Further validation of these harmonized reference ranges using outcomes data from longitudinal studies and randomized trials is an essential next step. Validation of reference ranges is a complex multistep process, which should include evaluation of the relation of varying degrees of deviation from the reference range with androgen-dependent outcomes (*e.g.*, sexual symptoms, hemoglobin, bone mineral density) in epidemiologic cohorts. Men with varying degree of deviation from the harmonized threshold would be expected to be substantially more likely to have symptoms/conditions typical of androgen deficiency. Furthermore, in randomized testosterone trials, the men with testosterone levels below the harmonized threshold would be more likely to respond to testosterone therapy than those with testosterone levels above the harmonized threshold. Indeed, recent data from the Testosterone Trials demonstrated that men with an average of two morning total testosterone levels <275 ng/dL exhibited improvements in

sexual activity and several domains of sexual function (40). In contrast, in the Testosterone Effects on Atherosclerosis Progression in Aging Men Trial (41) of men, 60 and older, whose mean testosterone level was >300 ng/dL, testosterone administration did not improve sexual function. Eventually, the specificity, sensitivity, and predictive value of these harmonized reference ranges should be evaluated in clinical populations of men seeking medical care.

In summary, these data demonstrate the feasibility and potential value of generating harmonized reference ranges for testosterone concentrations, whose serum total testosterone concentrations have been measured in a CDC-certified laboratory. There was a remarkable concordance in age-adjusted harmonized testosterone levels among men in four geographically distinct cohorts, suggesting that intercohort variation may be influenced by interassay variation. Further studies of the distribution of testosterone concentrations in other racial and ethnic groups and in populations in other regions of the world are needed to demonstrate the applicability of these ranges to broader populations of men in different regions of the United States and the world.

Acknowledgments

Address all correspondence and requests for reprints to: Shalender Bhasin, MB, BS, Brigham and Women's Hospital, 221 Longwood Avenue, Boston, MA 02115. E-mail: sbhasin@partners.org.

This work was supported primarily by National Institutes of Health Grant 1RO1AG31206 to S.B. Additional support was provided by the Endocrine Society and Boston Claude D. Pepper Older Americans Independence Center Grant 5P30AG031679 from the National Institute on Aging. The Framingham Heart Study was supported by National Heart, Lung, and Blood Institute Framingham Heart Study Contract N01-HC-25195. The European Male Aging Study (EMAS) was supported by Commission of the European Communities Fifth Framework Programme "Quality of Life and Management of Living Resources" Grant QLK6-CT-2001-00258. The Osteoporotic Fractures in Men Study (MrOS) was supported by National Institutes of Health. The following institutes provided support: National Institute on Aging, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Center for Advancing Translational Sciences, and National Institutes of Health Roadmap for Medical Research under Grants U01 AG027810, U01 AG042124, U01 AG042139, U01 AG042140, U01 AG042143, U01 AG042145, U01 AG042168, U01 AR066160, and UL1 TR000128. The Belgian Sibling Study of Osteoporosis was supported by a grant from the Fund for Scientific Research–Flanders (FWO–Vlaanderen Grant G.0662.08) and by a grant from the Hercules Foundation, Flanders.

Disclaimers: The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official views or positions of the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry. All individuals listed as authors agreed to be co-authors.

Disclosure Summary: S.B. has received research grant support from AbbVie Pharmaceuticals, Transition Therapeutics, Takeda Pharmaceuticals, and Eli Lilly for investigator-initiated research unrelated to this study. S.B. has served as a consultant to AbbVie, Regeneron, Novartis, and Eli Lilly. S.B. has a financial interest in Function Promoting Therapies, a company aiming to develop innovative solutions that enhance precision and accuracy in clinical decision making and facilitate personalized therapeutic choices in reproductive health. S.B.'s interests were reviewed and are managed by Brigham and Women's Hospital and Partners HealthCare in accordance with their conflict of interest policies. A.M.M. has received research grant support from AbbVie and GlaxoSmithKline and has served as a consultant to AbbVie, Endo, Lilly, and Lipocine. F.W. has received research grant support from Besins Healthcare and Eli Lilly and has served as a consultant to Besins Healthcare and Repro Therapeutics. Other authors have nothing to disclose.

References

1. Bhasin S, Zhang A, Coviello A, Jasuja R, Ulloor J, Singh R, Vesper H, Vasan RS. The impact of assay quality and reference ranges on clinical decision making in the diagnosis of androgen disorders. *Steroids*. 2008;**73**(13):1311–1317.
2. PetitClerk C, Solberg HE. Approved recommendations (1987) on the theory of reference values, II: selection of individuals for the production of reference values. *Clin Chim Acta*. 1987;**170**:S1–S11.
3. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, Montori VM. Testosterone therapy in adult men with androgen deficiency syndromes: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2006;**91**(6):1995–2010.
4. Bhasin S, Pencina M, Jasuja GK, Travison TG, Coviello A, Orwoll E, Wang PY, Nielson C, Wu F, Tajar A, Labrie F, Vesper H, Zhang A, Ulloor J, Singh R, D'Agostino R, Vasan RS. Reference ranges for testosterone in men generated using liquid chromatography tandem mass spectrometry in a community-based sample of healthy nonobese young men in the Framingham Heart Study and applied to three geographically distinct cohorts. *J Clin Endocrinol Metab*. 2011;**96**(8):2430–2439.
5. Eskelinen S, Vahlberg T, Isoaho R, Kivelä SL, Irjala K. Biochemical reference intervals for sex hormones with a new AutoDelfia method in aged men. *Clin Chem Lab Med*. 2007;**45**(2):249–253.
6. Sikaris K, McLachlan RI, Kazlauskas R, de Kretser D, Holden CA, Handelsman DJ. Reproductive hormone reference intervals for healthy fertile young men: evaluation of automated platform assays. *J Clin Endocrinol Metab*. 2005;**90**(11):5928–5936.
7. Tennekoon KH, Karunanayake EH. Serum FSH, LH, and testosterone concentrations in presumably fertile men: effect of age. *Int J Fertil*. 1993;**38**(2):108–112.
8. Boyce MJ, Baisley KJ, Clark EV, Warrington SJ. Are published normal ranges of serum testosterone too high? Results of a cross-sectional survey of serum testosterone and luteinizing hormone in healthy men. *BJU Int*. 2004;**94**(6):881–885.
9. Salameh WA, Redor-Goldman MM, Clarke NJ, Reitz RE, Caulfield MP. Validation of a total testosterone assay using high-turbulence liquid chromatography tandem mass spectrometry: total and free testosterone reference ranges. *Steroids*. 2010;**75**(2):169–175.

10. Haring R, Hannemann A, John U, Radke D, Nauck M, Wallaschofski H, Owen L, Adaway J, Keevil BG, Brabant G. Age-specific reference ranges for serum testosterone and androstenedione concentrations in women measured by liquid chromatography-tandem mass spectrometry. *J Clin Endocrinol Metab*. 2012;**97**(2):408–415.
11. Yeap BB, Alfonso H, Chubb SA, Handelsman DJ, Hankey GJ, Norman PE, Flicker L. Reference ranges and determinants of testosterone, dihydrotestosterone, and estradiol levels measured using liquid chromatography-tandem mass spectrometry in a population-based cohort of older men. *J Clin Endocrinol Metab*. 2012;**97**(11):4030–4039.
12. Lee DM, O'Neill TW, Pye SR, Silman AJ, Finn JD, Pendleton N, Tajar A, Bartfai G, Casanueva F, Forti G, Giwercman A, Huhtaniemi IT, Kula K, Punab M, Boonen S, Vanderschueren D, Wu FC; EMAS Study Group. The European Male Ageing Study (EMAS): design, methods and recruitment. *Int J Androl*. 2009;**32**(1):11–24.
13. Wu FC, Tajar A, Beynon JM, Pye SR, Silman AJ, Finn JD, O'Neill TW, Bartfai G, Casanueva FF, Forti G, Giwercman A, Han TS, Kula K, Lean ME, Pendleton N, Punab M, Boonen S, Vanderschueren D, Labrie F, Huhtaniemi IT; EMAS Group. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med*. 2010; **363**(2):123–135.
14. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Sr, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*. 2007;**165**(11):1328–1335.
15. Orwoll E, Blank JB, Barrett-Connor E, Cauley J, Cummings S, Ensrud K, Lewis C, Cawthon PM, Marcus R, Marshall LM, McGowan J, Phipps K, Sherman S, Stefanick ML, Stone K. Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study: a large observational study of the determinants of fracture in older men. *Contemp Clin Trials*. 2005;**26**(5):569–585.
16. Blank JB, Cawthon PM, Carrion-Petersen ML, Harper L, Johnson JP, Mitson E, Delay RR. Overview of recruitment for the osteoporotic fractures in men study (MrOS). *Contemp Clin Trials*. 2005; **26**(5):557–568.
17. Lapauw BM, Taes Y, Bogaert V, Vanbillemont G, Goemaere S, Zmierczak H-G, De Bacquer D, Kaufman JM. Serum estradiol is associated with volumetric BMD and modulates the impact of physical activity on bone size at the age of peak bone mass: a study in healthy male siblings. *J Bone Miner Res*. 2009;**24**(6):1075–1085.
18. Roef G, Taes Y, Toye K, Goemaere S, Fiers T, Verstraete A, Kaufman JM. Heredity and lifestyle in the determination of between-subject variation in thyroid hormone levels in euthyroid men. *Eur J Endocrinol*. 2013;**169**(6):835–844.
19. Sir-Petermann T, Codner E, Pérez V, Echiburú B, Maliqueo M, Ladrón de Guevara A, Preisler J, Crisosto N, Sánchez F, Cassorla F, Bhasin S. Metabolic and reproductive features before and during puberty in daughters of women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2009;**94**(6):1923–1930.
20. Labrie F, Bélanger A, Bélanger P, Bérubé R, Martel C, Cusan L, Gomez J, Candas B, Castiel I, Chaussade V, Deloche C, Leclaire J. Androgen glucuronides, instead of testosterone, as the new markers of androgenic activity in women. *J Steroid Biochem Mol Biol*. 2006; **99**(4-5):182–188.
21. LeBlanc ES, Nielson CM, Marshall LM, Lapidus JA, Barrett-Connor E, Ensrud KE, Hoffman AR, Laughlin G, Ohlsson C, Orwoll ES; Osteoporotic Fractures in Men Study Group. The effects of serum testosterone, estradiol, and sex hormone binding globulin levels on fracture risk in older men. *J Clin Endocrinol Metab*. 2009; **94**(9):3337–3346.
22. Van Uytfanghe K, Stöckl D, Kaufman JM, Fiers T, Ross HA, De Leenheer AP, Thienpont LM. Evaluation of a candidate reference measurement procedure for serum free testosterone based on ultrafiltration and isotope dilution-gas chromatography-mass spectrometry. *Clin Chem*. 2004;**50**(11):2101–2110.
23. Wang Y, Gay GD, Botelho JC, Caudill SP, Vesper HW. Total testosterone quantitative measurement in serum by LC-MS/MS. *Clin Chim Acta*. 2014;**436**:263–267.
24. Caudill SP, Schleicher RL, Pirkle JL. Multi-rule quality control for the age-related eye disease study. *Stat Med*. 2008;**27**(20): 4094–4106.
25. Therneau T. Deming, Thiel-Sen and Passing-Bablock regression, 2014. Available at: https://CRAN.R-project.org/package=deming. Accessed 15 July 2016.
26. Hyndman RJ, Fan Y. Sample quantiles in statistical packages. *Am Stat*. 1996;**50**(4):361–365.
27. Reiss RD. *Approximate Distributions of Order Statistics With Applications to Nonparametric Statistics*. New York: Springer-Verlag; 1989.
28. Muggeo VMR, Sciandra M, Tomasello A, Calvo S. Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. *Environ Ecol Stat*. 2013; **20**:519–531.
29. Elveback L. The population of healthy persons as a source of reference information. *Hum Pathol*. 1973;**4**(1):9–16.
30. Wu FC, Tajar A, Pye SR, Silman AJ, Finn JD, O'Neill TW, Bartfai G, Casanueva F, Forti G, Giwercman A, Huhtaniemi IT, Kula K, Punab M, Boonen S, Vanderschueren D; European Male Aging Study Group. Hypothalamic-pituitary-testicular axis disruptions in older men are differentially linked to age and modifiable risk factors: the European Male Aging Study. *J Clin Endocrinol Metab*. 2008;**93**(7):2737–2745.
31. Mohr BA, Bhasin S, Link CL, O'Donnell AB, McKinlay JB. The effect of changes in adiposity on testosterone levels in older men: longitudinal results from the Massachusetts Male Aging Study. *Eur J Endocrinol*. 2006;**155**(3):443–452.
32. Rastrelli G, Carter EL, Ahern T, Finn JD, Antonio L, O'Neill TW, Bartfai G, Casanueva FF, Forti G, Keevil B, Maggi M, Giwercman A, Han TS, Huhtaniemi IT, Kula K, Lean MEJ, Pendleton N, Punab M, Vanderschueren D, Wu FCW; EMAS Study Group. Development of and recovery from secondary hypogonadism in aging men: prospective results from the EMAS. *J Clin Endocrinol Metab*. 2015; **100**(8):3172–3182.
33. Ahern T, Swiecicka A, Eendebak RJ, Carter EL, Finn JD, Pye SR, O'Neill TW, Antonio L, Keevil B, Bartfai G, Casanueva FF, Forti G, Giwercman A, Han TS, Kula K, Lean ME, Pendleton N, Punab M, Rastrelli G, Rutter MK, Vanderschueren D, Huhtaniemi IT, Wu FC; EMAS Study Group. Natural history, risk factors and clinical features of primary hypogonadism in ageing men: longitudinal data from the European Male Ageing Study. *Clin Endocrinol*. 2016; **85**(6):891-901.
34. Orwoll ES, Nielson CM, Labrie F, Barrett-Connor E, Cauley JA, Cummings SR, Ensrud K, Karlsson M, Lau E, Leung PC, Lunggren O, Mellström D, Patrick AL, Stefanick ML, Nakamura K, Yoshimura N, Zmuda J, Vandenput L, Ohlsson C; Osteoporotic Fractures in Men (MrOS) Research Group. Evidence for geographical and racial variation in serum sex steroid levels in older men. *J Clin Endocrinol Metab*. 2010;**95**(10):E151–E160.
35. Vesper HW, Wang Y, Vidal M, Botelho JC, Caudill SP. Serum total testosterone concentrations in the US household population from the NHANES 2011-2012 study population. *Clin Chem*. 2015; **61**(12):1495–1504.
36. Litman HJ, Bhasin S, Link CL, Araujo AB, McKinlay JB. Serum androgen levels in black, Hispanic, and white men. *J Clin Endocrinol Metab*. 2006;**91**(11):4326–4334.
37. Travison TG, Araujo AB, Kupelian V, O'Donnell AB, McKinlay JB. The relative contributions of aging, health, and lifestyle factors to serum testosterone decline in men. *J Clin Endocrinol Metab*. 2007; **92**(2):549–555.
38. Little RR, Rohlfing CL, Wiedmeyer H-M, Myers GL, Sacks DB, Goldstein DE; NGSP Steering Committee. The national glycohemoglobin standardization program: a five-year progress report. *Clin Chem*. 2001;**47**(11):1985–1992.

39. Myers GL, Cooper GR, Winn CL, Smith SJ. The Centers for Disease Control-National Heart, Lung and Blood Institute Lipid Standardization Program: an approach to accurate and precise lipid measurements. *Clin Lab Med*. 1989;**9**(1):105–135.
40. Snyder PJ, Bhasin S, Cunningham GR, Matsumoto AM, Stephens-Shields AJ, Cauley JA, Gill TM, Barrett-Connor E, Swerdloff RS, Wang C, Ensrud KE, Lewis CE, Farrar JT, Cella D, Rosen RC, Pahor M, Crandall JP, Molitch ME, Cifelli D, Dougar D, Fluharty L, Resnick SM, Storer TW, Anton S, Basaria S, Diem SJ, Hou X, Mohler ER III, Parsons JK, Wenger NK, Zeldow B, Landis JR, Ellenberg SS; Testosterone Trials Investigators. Effects of testosterone treatment in older men. *N Engl J Med*. 2016;**374**(7):611–624.
41. Basaria S, Harman SM, Travison TG, Hodis H, Tsitouras P, Budoff M, Pencina KM, Vita J, Dzekov C, Mazer NA, Coviello AD, Knapp PE, Hally K, Pinjic E, Yan M, Storer TW, Bhasin S. Effects of testosterone administration for 3 years on subclinical atherosclerosis progression in older men with low or low-normal testosterone levels: a randomized clinical trial. *JAMA*. 2015;**314**(6):570–581.