Testosterone, Dehydroepiandrosterone, and Physical Performance in Older Men: Results from the Massachusetts Male Aging Study

Amy B. O'Donnell, Thomas G. Travison, Susan S. Harris, J. Lisa Tenover, and John B. McKinlay

Objective: This manuscript examines the relationships of total testosterone (T), bioavailable T, dehydroepiandrosterone (DHEA), and DHEA sulfate (DHEAS) to measures of physical performance in a large, population-based, random sample of men.

Methods: In the most recent wave of the Massachusetts Male Aging Study, measures of strength and physical performance [seven-item physical performance test (PPT), timed chair stand test, and grip strength] were made in 684 men, aged 55–83 yr. Complete hormone data were also obtained. Initial graphical exploration of performance outcomes as a function of hormone levels showed linear increases in physical performance up to certain threshold hormone concentrations, beyond which the associations were diminished. Regression models were used to estimate threshold locations and standardized regression coefficients quantifying the association between hormones and strength.

Results: All hormones exhibited significant age-adjusted positive association with PPT score below, but not necessarily above, the thresholds. DHEA was positively associated with chair stand score below, but not above the threshold. None of the hormones studied was significantly associated with grip strength.

Conclusion: Up to certain critical concentrations, elevated levels of TT, total T, bioavailable T, DHEA, and DHEA sulfate are associated with increased physical performance, as indicated by the PPT. However, levels beyond those critical concentrations, as might be achieved through exogenous supplementation, do not appear to confer any additional benefit. In general, hormone concentrations do not appear to be meaningfully associated with grip strength or chair stand scores.

(J Clin Endocrinol Metab 91: 425–431, 2006)
physical performance, muscle, and strength. Although Nestler et al. (21) demonstrated an increase in muscle mass in healthy young men with DHEA therapy, these findings were not upheld in later work on obese (22) and nonobese (23) young men. Brown et al. (24) gave varying DHEA doses to 30 young men and found no difference in muscle strength between the groups during resistance training.

The Massachusetts Male Aging Study (MMAS) is the first study to examine associations of T and DHEA with a range of physical performance measures in a large, population-based, random sample of men.

**Subjects and Methods**

**Study sample and in-home visit**

The MMAS is an observational cohort study of health in aging men. The design and field protocol have been described previously (25). Briefly, communities in the Boston, Massachusetts, area were randomly selected with probabilities proportional to the population within each of six strata defined by community size and median income. Men born between 1917 and 1946 were drawn at random from the Massachusetts annual state census list, with sampling fractions adjusted to produce a uniform distribution between ages 40 and 70 yr. A total of 1709 respondents completed a baseline visit between 1987 and 1989. Follow-up visits were conducted after approximately 9 yr (1995–1997) and 15 yr (2002–2004; Fig. 1). Measures of physical strength and performance were newly introduced at the time of the second follow-up.

Of the 855 men who completed the second follow-up interview, 79 were interviewed by telephone because they had moved out of state, and one subject began, but did not complete, the interview. These subjects’ physical performance measurements are therefore unavailable. Of the remaining 775 men, 91 did not have valid hormone measures (they refused the blood draw, did not have a blood draw due to health risks, or were currently being treated with supplemental androgens or anrogen suppressor medications). The remaining 684 subjects are considered in this analysis.

The MMAS field protocol involved a trained field technician/phlebotomist visiting each subject in his home, obtaining written informed consent, and drawing two nonfasting blood samples. After administering an extensive set of questions concerning health status and behavior, the interviewer conducted a series of physical performance measures, including chair stands, grip strength, and a performance test.

**Measure of T**

Research design (e.g. time of day of blood sampling and method of analysis) and the study sample composition (e.g. whether it is made up of patients or volunteers) may impact T levels (25); accordingly, the design of the MMAS accounted for potential biases in measuring hormone levels. Nonfasting blood samples were drawn within 4 h of the subject’s awakening to control for diurnal variations in hormone levels. Two samples were drawn 30 min apart and pooled for analysis in equal aliquots to dampen the influence of episodic secretion. Blood was kept in an ice-cooled container for transport and was centrifuged within 6 h. Serum was stored in 5-ml scintillation vials at −70 C until the time of assay.

All hormone assays were performed in Dr. Christopher Longcope’s laboratory at the University of Massachusetts Medical School (Worcester, MA). Total T (TT) was measured by RIA kit (Diagnostic Products Corp., Los Angeles, CA). The intraassay coefficient of variation was 3.6%, and the interassay coefficient of variation was 8.3%. Bioavailable T (BT; T not bound to SHBG) was calculated as BT = TT × (% free T + percent albumin-bound T)/100. DHEA and DHEAS were also measured by RIA, with intraassay coefficients of variation of 2.1% and 4.2%, and interassay coefficients of variation of 10.3% and 6.4%, respectively. Full details regarding hormone measurements have been reported previously (26). As we have noted previously (25), the distribution of MMAS serum total T values is comparable to those reported by other major epidemiologic studies using similar RIA techniques and exhibits a similar rate of decline with age.

**Measures of physical strength and function**

The PPT was used to assess physical performance on functional tasks (27) for 684 men. The PPT has been shown to be reliable (Cronbach’s α = 0.79; interrater reliability = 0.93) (27) and has demonstrated concurrent validity with self-reported measures of physical function. We administered the seven-item version that included writing a sentence, scooping a bean from a container (simulated eating), putting on a jacket (simulated dressing), picking up a penny from the floor, placing a book on a high shelf, turning 360°, and walking 50 ft. The PPT is scored by awarding between 0 and 4 points for each of the seven tasks, yielding a range of 0–28 for the total score. If the respondent was unable to perform a specific task, 0 points were given. Otherwise, the score was based on the time taken to do the task, with 4 points for the fastest performance and 1 point for the slowest. For the 360° turn, points were awarded based on qualitative features of the turn: 2 points for continuous rather than discontinuous steps, and 2 points for steady rather than unsteady movement, giving possible scores of 0, 2, or 4 points. The PPT test has been previously used in other large epidemiological studies for community-dwelling and institutionalized older adults, demonstrating construct validity and internal consistency (28).

Grip strength was measured using the Jamar hydraulic hand dynamometer (Sammons Preston, Bolingbrook, IL), which measures isometric grip force from 0–200 lb. A peak-hold needle automatically retains the highest reading until reset. The instrument can be adjusted to five grip positions to accommodate different hand sizes. Reliability has been assessed using the Pearson product-moment correlation and a one-way ANOVA for repeated measures, such that all correlations were significant (r < 0.01), and the median and modal correlations were all 0.97 or 0.98 (29). For the grip strength measurement, respondents were seated comfortably in straight back chairs with their arms at their sides and the elbows of their dominant hands flexed at a 90° angle. The forearm and wrist were in neutral alignment. Each respondent was asked to squeeze the dynamometer for 3 sec once, rest 1 min, and then squeeze again for 3 sec. The higher of the two measurements was used in the analyses. Three people did not complete the screening, 53 had pain/arthritis in their hands, one had had surgery on the dominant hand, and four did not believe they could do the grip strength test safely, leaving grip strength measurements for 623 participants.

The repeated chair stand test, used to measure lower extremity strength, was conducted according to a standard protocol (30). Participants were asked whether they thought it was safe to make five attempts and, if they agreed, were asked to stand up straight as quickly as they could and then sit back down without stopping in between. Points were awarded based on the time needed to complete the chair stands. The chair stand task was added to the research protocol after 25 men had been interviewed, yielding 659 chair stand scores.

---

**Fig. 1. MMAS design.**

O'Donnell et al. • Androgens and Physical Performance in Older Men
Statistical analysis

Scatterplots and locally weighted regression smoothers were used to explore the relationship between hormone levels and performance. In several specific cases, exploratory data analysis elicited a so-called threshold effect as a feature of the association between hormone levels and measures of physical function; that is, physical performance was positively correlated with hormone levels up to certain concentrations, beyond which mean physical performance was relatively stable or might, in fact, appear to decrease with increasing hormone concentrations.

Because of this pattern, subsequent analyses conceptualized the presence of the threshold effects as nonlinearities in the relationship between hormone levels and measures of physical performance. For each model, the location of the nonlinearity was chosen from those consistent with the exploratory graphical analysis in such a way as to minimize the residual sum of squares of a regression of physical performance on hormone concentration when controlling for age effects.

The resulting models allow the estimated slope of mean performance on hormone concentration below the threshold to differ from that above. In many cases these differences appeared to be both statistically and clinically significant. For all models, age and age squared were included as covariates. Height was also included for the grip strength model, because height has been shown to have a positive correlation to grip strength (31). Nonparametric rank-sum tests were employed to compare physical performance among men with hormone levels below the relevant threshold to that among men with hormone levels above that threshold. These comparisons were supplemented with comparisons based upon a linear regression analysis controlling for age and body mass index (BMI). Results were considered statistically significant if the null hypothesis of no difference could be rejected at the 0.05 level.

Results

Table 1 presents the characteristics of the full analysis sample (n = 684; mean ± sd). The PPT score was 22.8 ± 4.3 of a possible 28 points, indicating that our population possessed a level of performance comparable to other epidemiological studies of similar populations (32). The mean age of the subjects considered in this analysis was 68 ± 8.0 yr at the time PPT measurements were made. Two hundred eighty-six subjects (42%) were between the ages of 55 and 64 yr, 238 subjects (35%) were between the ages of 65 and 74 yr, and 160 subjects (23%) were between the ages of 75 and 85 yr.

As described above, threshold values were determined for the association between each hormone and performance outcome (Table 2). In most cases threshold values were between the median and the 75th percentile of the hormone distribution (to accommodate distributional skew, log-transformed DHEA and DHEAS were considered, with little substantive impact on results; for ease of presentation, we focus on untransformed hormone values here).

The standardized regression coefficients above and below the estimated threshold for each of the hormones are listed in Table 2. These coefficients may be interpreted as standardized slopes. For example, in Table 2 we note that the model-estimated slope of PPT on BT below the relevant threshold concentration, 141 ng/dl (4.89 nmol/liter), is 0.224 and is highly statistically significant (P < 0.001). This indicates that, on the average, a typical subject with a BT less than or equal to 141 ng/dl would score 0.224 points higher on the PPT than would another subject of the same age whose BT concentration was 1 sd, roughly 52 ng/dl (1.80 nmol/liter), lower.

Below their respective thresholds, both TT and BT were positively associated with age-adjusted scores on the PPT. DHEA and DHEAS measures followed the same trend below their respective thresholds, but were inversely associated with the PPT at a statistically significant level above their thresholds. Neither TT nor BT was positively associated with age-adjusted scores on the timed chair stand test (controlling for age) or with grip strength (controlling for age and height). Figure 2 depicts the predicted PPT score over a range of possible TT values for 68-yr-old, 28 kg/m² men (these are the mean age and BMI for the analysis sample, respectively). A typical such man with a total T concentration of 300 ng/dl (10.40 nmol/liter) would be expected to have a PPT score of approximately 23.0.

In terms of statistical significance, the strongest positive association between hormones and physical measures when controlling for age was that between DHEAS and a PPT score below the DHEAS threshold of 97 µg/dl (2614 nmol/liter). The only significant negative associations appear above the threshold for DHEA and DHEAS with the PPT score and above the threshold for DHEAS with chair stand test score. DHEAS followed a similar pattern as DHEA below the thresholds, showing a significant positive association with the PPT score and no significant association with grip strength. Only DHEA was significantly associated with chair stand score, and no hormone concentrations were associated with grip strength.

Components of PPT

Analyses revealed that hormone concentrations appear to be more strongly associated with the PPT than with the chair stand score or grip strength. When we examined the relationship between T and DHEA/DHEAS with individual components of the total PPT, we observed similar relationships to those observed with the overall PPT.

Can the association between physical performance and hormone levels be summarized using the thresholds only? We sought to determine whether the observed differences persisted when subjects with hormone concentrations above the thresholds were compared, as a group, to those subjects whose concentrations were below the thresholds. Table 3 presents the characteristics of the full analysis sample.
presents the mean time to perform PPT tasks for subjects whose hormone concentrations were above the thresholds for TT and DHEAS vs. those whose concentrations were not. Subjects whose hormone levels were above thresholds consistently performed all tasks more rapidly than those whose levels were below the thresholds, but in most cases the differences were small and did not rise to the level of statistical significance. Whether TT levels were above or below the relevant threshold was independently associated with time taken to walk 50 ft, to simulate eating, and to picking up a penny. Being above vs. below the DHEAS threshold was independently associated with writing a sentence, simulated eating, and walking 50 ft.

Age was more strongly associated with DHEAS than with TT, whereas BMI was more strongly associated with TT than with DHEAS. The mean age (±sd) of subjects whose TT levels were below the threshold 451 ng/dl was 68.01 (±8.19) yr, whereas the mean age of those whose TT levels were above the threshold was 68.05 (±7.80) yr. By contrast, the mean (±sd) age of subjects with DHEAS below the threshold of 97 μg/dl was 69.73 (±8.05) yr, whereas that of subjects whose DHEAS values were higher were younger, on the average, was 65.50 (±7.34) yr. In regression analyses with age and BMI effects controlled, being above the TT threshold remained associated with decreased time to pick up a penny and walk 50 ft, but differences with respect to the DHEAS threshold were no longer statistically significant.

Turning in a circle was omitted from this analysis, because it was not a timed task.

### TABLE 2. Age-adjusted associations between hormones and PPT scores (n = 684), chair stand test scores (n = 659), and grip strength scores (n = 623), Massachusetts Male Aging Study, Third Wave 2002–2004

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Threshold (SI)</th>
<th>Standardized regression coefficient for hormone levels below threshold (P values)</th>
<th>Standardized regression coefficient for hormone levels above threshold (P values)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PPT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>390 ng/dl</td>
<td>451 ng/dl</td>
<td>0.141 (0.037)</td>
<td>0.023 (0.742)</td>
</tr>
<tr>
<td>BT</td>
<td>128 ng/dl</td>
<td>141 ng/dl</td>
<td>0.224 (0.001)</td>
<td>-0.050 (0.432)</td>
</tr>
<tr>
<td>DHEA</td>
<td>280 ng/dl</td>
<td>464 ng/dl</td>
<td>0.222 (0.000)</td>
<td>-0.280 (0.004)</td>
</tr>
<tr>
<td>DHEAS</td>
<td>80 μg/dl</td>
<td>97 μg/dl</td>
<td>0.684 (0.000)</td>
<td>-0.109 (0.013)</td>
</tr>
<tr>
<td><strong>Chair stand test scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>390 ng/dl</td>
<td>537 ng/dl</td>
<td>0.086 (0.129)</td>
<td>0.024 (0.812)</td>
</tr>
<tr>
<td>BT</td>
<td>128 ng/dl</td>
<td>140 ng/dl</td>
<td>0.118 (0.115)</td>
<td>-0.033 (0.629)</td>
</tr>
<tr>
<td>DHEA</td>
<td>280 ng/dl</td>
<td>551 ng/dl</td>
<td>0.164 (0.001)</td>
<td>-0.057 (0.688)</td>
</tr>
<tr>
<td>DHEAS</td>
<td>80 μg/dl</td>
<td>117 μg/dl</td>
<td>0.288 (0.082)</td>
<td>-0.105 (0.036)</td>
</tr>
<tr>
<td><strong>Grip strength</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>390 ng/dl</td>
<td>433 ng/dl</td>
<td>0.054 (0.450)</td>
<td>-0.033 (0.611)</td>
</tr>
<tr>
<td>BT</td>
<td>128 ng/dl</td>
<td>142 ng/dl</td>
<td>0.126 (0.066)</td>
<td>-0.031 (0.621)</td>
</tr>
<tr>
<td>DHEA</td>
<td>280 ng/dl</td>
<td>361 ng/dl</td>
<td>0.103 (0.150)</td>
<td>-0.076 (0.275)</td>
</tr>
<tr>
<td>DHEAS</td>
<td>80 μg/dl</td>
<td>266 μg/dl</td>
<td>0.121 (0.089)</td>
<td>0.067 (0.374)</td>
</tr>
</tbody>
</table>

Nanograms per deciliter can be converted to nmol/liter by multiplying by 0.03467; μg/dl can be converted to nmol/liter by multiplying by 26.9485.

* Obtained from a linear spline regression of PPT score on hormone, height, age and age-squared. Standardized regression coefficient indicates the PPT increase, in units of sd, associated with 1 sd increase in hormone.
TABLE 3. Time to complete physical performance tasks (n = 684), by hormone level

<table>
<thead>
<tr>
<th>Task</th>
<th>Below 451 ng/dl&lt;sup&gt;a&lt;/sup&gt;</th>
<th>At or above 451 ng/dl&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Below 97 μg/dl&lt;sup&gt;c&lt;/sup&gt;</th>
<th>At or above 97 μg/dl&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean time (sec)</td>
<td>SD</td>
<td>Mean time (sec)</td>
<td>SD</td>
</tr>
<tr>
<td>Write a sentence</td>
<td>12.63</td>
<td>4.62</td>
<td>12.18</td>
<td>4.22</td>
</tr>
<tr>
<td>Simulate eating</td>
<td>12.46</td>
<td>3.32</td>
<td>12.00</td>
<td>3.09</td>
</tr>
<tr>
<td>Walk 50 ft&lt;sup&gt;e&lt;/sup&gt;</td>
<td>16.77</td>
<td>6.19</td>
<td>15.47</td>
<td>4.13</td>
</tr>
<tr>
<td>Pick up penny&lt;sup&gt;f&lt;/sup&gt;</td>
<td>2.85</td>
<td>1.59</td>
<td>2.53</td>
<td>1.38</td>
</tr>
<tr>
<td>Lift book</td>
<td>2.14</td>
<td>1.64</td>
<td>2.13</td>
<td>1.43</td>
</tr>
</tbody>
</table>

Nanograms per deciliter can be converted to nmol/liter by multiplying by 0.03467; μg/dl can be converted to nmol/liter by multiplying by 26.9485.

<sup>a</sup> n = 433 (63%); mean (SD) age (yr), 68.0 (8.2); mean (SD) BMI, 28.9 (5.0).
<sup>b</sup> n = 251 (37%); mean (SD) age (yr), 68.0 (7.8); mean (SD) BMI, 26.7 (3.8).
<sup>c</sup> n = 409 (60%); mean (SD) age (yr), 69.7 (8.1); mean (SD) BMI, 28.1 (4.5).
<sup>d</sup> n = 275 (40%); mean (SD) age (yr), 65.5 (7.3); mean (SD) BMI, 28.1 (4.9).
<sup>e</sup> Difference between subjects below and above thresholds, with age and BMI effects controlled, is statistically significant at the 0.05 level (result obtained via linear regression analysis).

Discussion

We found positive associations in older men between hormone concentrations and physical performance as measured by the PPT, although there were minimal, if any, significant relationships between these hormones and physical performance as measured by chair stands or grip strength. Positive associations between hormone concentrations and PPT performance were present only with hormone levels below certain threshold values. These threshold values are above the median in all cases. The TT, BT, and DHEAS thresholds fall in the mid-high range, whereas the DHEA threshold is in the high range.

We examined the possibility that the apparent nonlinearities in the associations we observed were the result of outcome ceiling effects, which might occur if, for instance, a large percentage of the dependent variable values were clustered at the high end of the scale. This was not the case; for instance, only 4% of the PPT scores were 28, the highest possible value for that scale.

We observed differences in the linear relationship between performance and hormone levels below relevant thresholds vs. above. However, in an additional comparison of the collapsed group of subjects whose levels were above thresholds to those below (without regard to their hormone level, except in relation to the threshold cutoff), differences in time to perform specific components of the PPT were not usually significant when age effects were controlled, even though subjects with higher hormone levels completed every task more rapidly than their counterparts below the thresholds. The crude comparison of subjects above vs. below the threshold appears to diminish the strength of the relationships observed when the full spectrum of hormone values are employed in analyses; generally speaking, the differences observed do not persist when the subject groups above and below the thresholds are collapsed, and the threshold itself does not appear sufficient to account for the association between androgens and performance.

In particular, although the relationship between PPT scores and DHEAS levels was similar to that between PPT scores and T concentrations, we observed that when age effects were controlled, the crude differences between DHEAS groups (>97 μg/dl vs. below) in time taken to perform specific tasks were no longer statistically significant. The fact that in our sample those subjects with low DHEAS levels are somewhat older than those whose DHEAS levels are high suggests that the small, but consistent, differences between collapsed groups may be explained by age, and that the relationship between physical performance and DHEAS must be viewed across the spectrum of possible concentrations, rather than as a simple difference between those subjects below some threshold to those above.

One of the major advantages of this study is that we were able to measure overall physical performance, whereas much of the earlier work has focused solely on strength. The PPT (measuring physical performance) coupled with the grip strength (measuring upper extremity strength) and chair stands (measuring balance and lower extremity strength) presents a more complete picture of performance and ability than strength alone and may be more closely related to activities of everyday functioning and daily living, although tests such as the repeated chair stand may be somewhat ineffectual in discriminating among relatively healthy subjects. A second advantage of this study is that it was conducted in a large, community-based, random sample, rather than a clinical-based population, thus making our results more applicable to aging men in general.

Our results suggesting that beyond a certain threshold there may be little association between androgens and physical performance stand somewhat in contrast to those of Bhasin et al. (20), who found substantial gains in muscle mass and strength among older men at high-normal and supraphysiological doses of T and stated that the “best trade-off” between increased function and likelihood of adverse events is obtained in the high-normal range. Because these researchers employed the leg press as a general measure of muscle strength, with graduated resistance and weight-dependent repetition of the exercise, and measure the per time clinical effect of an induced change in T levels (whereas we measured a cross-sectional association between T and strength in an observational setting), some variance with respect to our results is to be expected.
Conversely, our results parallel those reported by Synder et al. (33), who demonstrated that the effectiveness of T treatment for increasing bone mineral density is inversely related to the level of pretreatment serum T.

An implication of our findings may be that men whose T and DHEA levels are high are likely to receive little benefit in physical performance from increased levels of either hormone. In addition, they may incur unnecessary risk. In a recent review, Tan and Culberson (34) list the following as potential risks for using T replacement therapy: increase in hematocrit level, increase in sleep apnea, sterility, erythrocytosis, edema, gynecomastia, and prostate stimulation. Others have suggested that potential risks for T replacement therapy include increased prostate-specific antigen levels (35) and overdosage for lifestyle reasons (36). Thus, we concur with others who have suggested that T replacement in men should be approached with caution (37). This caution should be applied to DHEA and its sulfate form as well. Although DHEA is currently taken by some in an effort to restore youth and vitality, there is very little research to support an antiaging effect (38, 39); in fact, the effects of oral administration of DHEA on blood T levels are uncertain (21, 40). When administering T or DHEA to improve physical performance, physicians should take the patient’s baseline hormone levels into account. If they are in the mid- to high-normal range, additional hormone supplementation may be useless and even potentially harmful.

Acknowledgments

We thank Sharon Tennstedt for helpful discussions, and William B. Simpson for statistical assistance.

Received June 2, 2005. Accepted November 23, 2005.

Address all correspondence and requests for reprints to: Dr. John B. McKinlay, New England Research Institutes, 9 Galen Street, Watertown, Massachusetts 02472. E-mail: jmckinlay@neriscience.com.

This work was supported by the National Institute of Diabetes and Digestive and Kidney Disorders (Grant DK-44995).

References


O’Donnell et al. • Androgens and Physical Performance in Older Men

JCEM is published monthly by The Endocrine Society (http://www.endo-society.org), the foremost professional society serving the endocrine community.