Testosterone Supplementation and Cognitive Functioning in Men—A Systematic Review and Meta-Analysis

Cecilie R. Buskbjerg,[1] Claus H. Gravholt,[2,3] Helene R. Dalby,[4] Ali Amidi,[1] and Robert Zachariae[1]

[1]Unit for Psychooncology and Health Psychology, Department of Oncology, Aarhus University Hospital and Department of Psychology and Behavioral Science, Aarhus University, 8000 Aarhus, Denmark; [2]Department of Endocrinology and Internal Medicine, Aarhus University Hospital, 8000 Aarhus, Denmark; [3]Department of Molecular Medicine, Aarhus University Hospital, 8000 Aarhus, Denmark; and [4]Department of Clinical Medicine, Aarhus University Hospital and Aarhus University, 8000 Aarhus, Denmark

ORCiD numbers: 0000-0002-5997-8973 (C. R. Buskbjerg).

Testosterone supplementation (TS) is assumed important for cognitive functioning in men, but conflicting results have prevented firm conclusions. The current study systematically reviewed available randomized controlled trials (RCTs) on effects of TS on cognitive functioning in men, subjected the findings to meta-analysis, and explored between-study differences as possible moderators of the effects. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, two authors independently searched for eligible records in the electronic databases of PubMed, PsycINFO, Web of Science, the Cochrane Library, Cumulative Index of Nursing and Allied Health, and Embase and determined eligibility using the following (population, intervention, comparison, outcome) criteria: population, male adults (>18 years); intervention, TS; comparison, placebo; and outcome, results of standardized neuropsychological tests. Following duplicate removal, 3873 records were screened with 92 remaining for full-text screening. Twenty-one papers reporting results of 23 independent RCTs were included, of which none treated samples of clinically hypogonadal men. The small improvement found in overall cognitive functioning (Hedges $g = 0.09$; CI 95%: −0.02 to 0.19) failed to reach statistical significance ($P = 0.108$) and approached zero when adjusting for possible publication bias ($g = 0.04$). The effects for the 11 individual cognitive domains did not reach statistical significance (g: −0.04 to 0.19, P: 0.061 to 0.989). Small statistically significant ($P < 0.05$) effects were found for five study subsets but failed to meet the fail-safe criterion. The available evidence indicates that effects of TS on cognitive functioning in men with testosterone levels within normal ranges are less robust and of insufficient magnitude to be of clinical relevance. The effects in clinically hypogonadal men remain to be investigated.

Freeform/Key Words: testosterone supplementation, androgens, cognitive functions, systematic review, meta-analysis

From 2001 to 2011, global testosterone prescription sales increased 12 fold from $150 million to $1.8 billion [1], and this increase is expected to continue and reach $3.8 billion by 2022 [2]. One reason for this boom in prescription sales is an increased recognition of hypogonadism [3], a syndrome affecting an estimated 25% of men over age 65 [4] and characterized by low physiological testosterone together with clinical symptoms of hypogonadism, such as decreased libido, impaired erectile function, and decreased energy [5].

Abbreviations: AAMI, age-associated memory impairment; RCT, randomized controlled trial; TS, testosterone supplementation.

Received 27 March 2019
Accepted 31 May 2019
First Published Online 6 June 2019

ISSN 2472-1972
August 2019 | Vol. 3, Iss. 8
doi: 10.1210/js.2019-00119 | Journal of the Endocrine Society | 1465–1484

Physiological testosterone is hypothesized to be important for cognition in men, and this has been supported by several lines of research. First, testosterone appears to influence neurobiological processes associated with cognitive aging and the development of neurodegenerative disorders, such as Alzheimer's disease. Testosterone has thus been found to delay neuronal apoptosis [6], accelerate the rate of nerve regeneration [7], modulate neuronal damage caused by oxidative stress [8], exert anti-inflammatory actions [9], and reduce beta amyloid peptide levels [10]. Second, cognitive functions deteriorate with advanced age [11], in parallel with an age-mediated decline in male testosterone levels, starting in the third decade [12]. Third, there is some evidence to suggest that prostate cancer patients receiving androgen-deprivation therapy are at increased risk of cognitive impairment and dementia compared with prostate cancer patients receiving other types of treatment [13, 14]. On this background, testosterone supplementation (TS) may possibly improve cognitive functioning in men.

Several trials have investigated the effect of TS on cognitive functioning in men, and whereas one previous systematic review [15] concluded that promising associations had been found between TS and cognitive functions in men with both normal and low levels of testosterone and in men with and without cognitive impairment, another systematic review [16] concluded that the use of TS to improve cognitive functioning was not supported by data from clinical trials. More recently, two nonsystematic reviews [17, 18] concluded that evidence indicates that TS has positive effects on cognitive functions, particularly in men with cognitive impairment and low testosterone levels. The conflicting conclusions from existing reviews reflect that the results of the existing trials vary considerably, which may possibly be a result of between-study differences in the included neuropsychological tests, as well as treatment modalities, *e.g.*, dosage, duration, type, and route of administration [15]. As all of the available systematic and nonsystematic reviews [15–20] to date have been nonquantitative, narrative reviews, a need for a systematic review with quantitative meta-analysis is indicated. Our aim was therefore to conduct a systematic review and meta-analysis to evaluate the effect of TS on cognitive functioning in men and to explore possible moderating effects of between-study differences in relevant study characteristics. Our primary hypothesis was that TS would have a positive effect on overall cognitive functioning, *i.e.*, the individual cognitive domains aggregated into one combined estimate of cognitive functioning. Furthermore, as existing findings indicate that men tend to outperform women on tasks that use visuospatial skills [21, 22] and that women's visuospatial skills improve when they receive TS [23], we expected to find the strongest effects for cognitive domains that use these skills (*i.e.*, visuospatial function, visuospatial learning, visuospatial memory, and visuomotor function). Finally, as a result of the suggested neuroprotective effect of testosterone against Alzheimer's disease pathology [7, 10], we hypothesized that the effect of TS on cognitive functioning would be stronger in studies administering testosterone to cognitively impaired men, *e.g.*, men with neurodegenerative disorders, such as Alzheimer's disease.

1. Methods

A. Search Strategy and Selection Criteria

Two authors (C.R.B. and H.R.D.) independently searched for reports on the effect of TS on cognitive functions in men in the electronic databases of PubMed, PsycINFO, Web of Science, the Cochrane Library, Cumulative Index of Nursing and Allied Health, and Embase. The final searches were repeated and updated on 6 December 2018. The search strategy included key words for men, testosterone, cognitive functions, and their synonyms (search strategy and results can be obtained by request from the corresponding author). In addition, a backward search (snowballing) of reference lists of identified articles and earlier systematic reviews was conducted, together with a forward search (citation tracking). Only English-language publications in peer-reviewed journals were included. Eligibility was determined using the population, intervention, comparison, outcome approach [24]: population, a male adult (>18 years) healthy or clinical sample; intervention, TS; comparison, placebo; and

outcome, standardized neuropsychological test results. The two authors independently screened the identified papers, excluded noneligible studies, retrieved and evaluated full texts of the remaining papers, excluded studies with registration of reasons, and extracted *a priori* specified data from eligible studies. The current study was preregistered with PROSPERO [25] (Number CRD42017060530) and is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses recommendations [26].

B. Data Analysis

Hedges *g*, which corrects for possible bias as a result of a small sample size [27], was chosen as the effect size. The data were combined using random-effects models. The effect of TS on overall cognitive functioning was chosen as the primary outcome and was calculated by pooling the effects for all cognitive domains across included studies. If a paper reported results of more than one neuropsychological test assessing the same cognitive domain, then a pooled, weighted effect size for that domain was calculated for the study in question to ensure independence. Secondary outcomes included pooled effect sizes calculated separately for individual cognitive domains reported in more than three studies. Effect sizes were categorized as small (0.2), medium (0.5), or large (0.8), respectively [28].

Studies were rated as having high, low, or unclear risk of bias using the Cochrane Collaborations tool for assessing risk of bias [29]. Industry sponsorship of pharmacotherapy trials has been associated with more favorable outcomes [30], and industry-sponsored studies were therefore rated as having a high risk of bias in the category of "other bias." To evaluate incomplete outcome reporting, records of trial registrations were reviewed and authors contacted to clarify study characteristics when needed.

Statistical heterogeneity was evaluated using I^2, with values of 0%, 25%, 50%, and 75% taken to indicate no, low, moderate, and high heterogeneity, respectively [31]. The Q statistic was used to evaluate the probability that results reflect systematic between-study differences [32]. As a result of the generally low statistical power of heterogeneity tests [33], $P < 0.10$ was taken to indicate heterogeneity.

Positive and negative findings may not be equally likely to get published, thereby introducing risk of publication bias. The distribution of effect sizes was visually inspected by means of funnel plots and tested with Egger test [34]. If results were suggestive of publication bias, an adjusted effect size was estimated using Duval and Tweedie trim and fill method [35]. The file-drawer problem, *i.e.*, the possibility that unidentified or unpublished studies with null findings could alter statistically significant meta-analysis results, was evaluated using Rosenthal fail-safe n [36]. If the fail-safe number exceeded the criterion of 5K + 10, with K being the number of studies included in the meta-analysis, then the results were considered relatively robust in case of unpublished null findings.

Two types of sensitivity analyses were conducted. First, the meta-analyses were repeated with "0" imputed as the effect size when authors of included studies stated that they had obtained nonsignificant results for neuropsychological outcomes without presenting the data. Second, the meta-analyses were repeated with "winsorizing", *i.e.*, by truncating outliers so that they did not differ >2 SD from the original pooled effect size [37].

When available in a sufficient number of studies ($K > 8$), a number of categorical and continuous moderators were explored with unadjusted meta-regression analyses (random-effects model, maximum-likelihood method) [38]. These included treatment characteristics [TS type, duration, testosterone measurement characteristics (time of day, assay type), effect of TS on physiological testosterone levels from pre- to post-treatment in the intervention group compared with placebo (Hedges *g*), participant characteristics (mean sample age, cognitive status, gonadal status), and study characteristics (publication year, industry sponsoring].

When associations between the included moderators reached statistical significance ($P < 0.05$) or were of medium magnitude ($r > 0.30$), moderators were analyzed together in adjusted models [27], taking the variance inflation factor into consideration [39, 40]. Between-study differences were further explored by pooling effect sizes for subgroups of studies according to the categorical moderators when available in at least two studies.

Comprehensive Meta-Analysis, version 3.3 [41], and IBM SPSS statistics, version 25 [42], were used for all analyses.

2. Results

The study selection process is shown in Fig. 1. Following duplicate removal, a total of 3873 records were screened with the two authors agreeing on 98.5% of the inclusion/exclusion decisions. After solving disagreement by negotiation, 92 records remained for full-text screening. The authors agreed on 82.4% of decisions. After solving disagreement by negotiation, a total of 71 records were excluded (data not shown; can be obtained by request from the corresponding author), resulting in 21 papers. One paper [43] presented separate data on Alzheimer's patients and healthy men, and another [44] reported separate data on young and old men, yielding a total number of 23 independent randomized controlled trials (RCTs) to be included and subjected to meta-analysis. The characteristics of the included studies are shown in Table 1. The mean sample age was 64.9 years (SD = 13.0), and mean treatment duration was 33.4 weeks (SD = 42.1). The neuropsychological tests used in the 23 RCTs corresponded to 11 distinct cognitive domains (data not shown; can be obtained by request from the corresponding author). Seventeen studies included men who had testosterone levels within the normal range (mean total testosterone between 321 and 865 ng/dL),

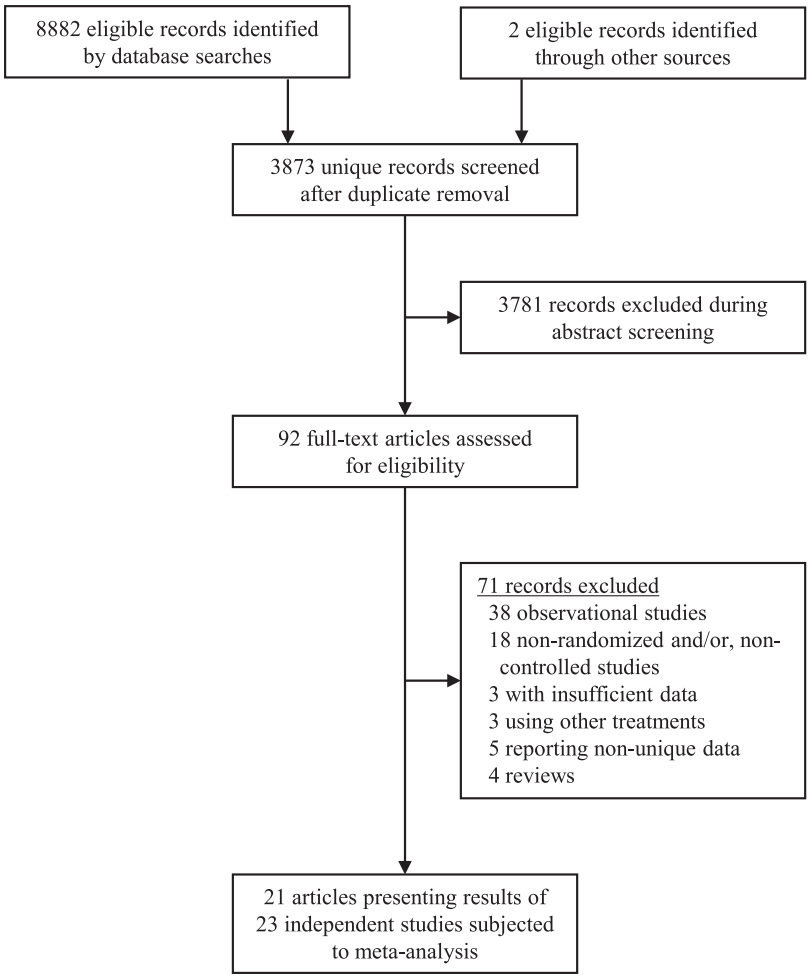


Figure 1. Flowchart of study selection.

Table 1. Description of Included Studies

Study	Cognitive Outcome	n[a] Pre	n[b] Post	Duration[c]	Type[d]	Doses	Mean Age (Range/SD)[e]	Cognitive Status	Gonadal Status Pre[f]	Gonadal Status Post[f]	TS Effect[g]	T Medium	Assay	Time
Janowsky *et al.* [45]	Executive function; verbal memory; visuomotor function	56	56	12.00	Patch	15 mg/d	67.40 (60–75)	Normal	Normal 555.2 (132.6) pM	High 850.2 (397.3) pM	1.11	Serum	Direct RIA	Morning/ NR
Janowsky *et al.* [46]	Attention/working memory	19	19	4.00	Injection	150 mg/wk	67.45 (61–75)	Normal	Normal 423.3 (128.7) pM	High 1568.4 (1086.1) pM	0.12	Serum	Direct RIA	Morning/ NR
Cherrier *et al.* [47]	Executive function; language; verbal memory; visuospatial memory; attention/working memory	28	25	6.00	Injection	100 mg/wk	67.40 (50–80)	Normal	Normal 576.4 (66.9) ng/dL	High 1239.2 (53.0) ng/dL	7.68	Serum	Direct RIA	Random/ not fasting
O'Connor *et al.* [48]	Executive function; language; verbal memory; visuomotor function; attention/working memory	30	29	8.00	Injection	200 mg/wk	28.30 (19–45)	Normal	Normal 625.4 (66.9) ng/dL	High 1107.6 (111.2) ng/dL	2.43	Serum	Direct CIA	Random/ not fasting
Kenny *et al.* [49]	Executive function; language; attention/working memory	67	44	52.00	Patch	2–2.5 mg/d	76.00 (65–87)	Normal	Normal 93.0 (34.0) ng/dL	Normal 162.0 (100) ng/dL	1.00	Serum	Direct RIA	Morning/ fasting
Kenny *et al.* [50]	Cognitive status; executive function; visuomotor function; attention/working memory	11	11	12.00	Injection	200 mg/ 2 wk	80.00 (73–87)	Impaired (MCI)	Normal 410.0 (112.0) ng/dL	High 1211.0 (360.0) ng/dL	2.59	Serum	Direct RIA	NR/ fasting
Cherrier *et al.* [51]	Executive function; language; visuospatial memory; visuomotor function; attention/ working memory	41	38	6.00	Injection	100 mg/wk	65.00 (50–85)	Normal	Normal 403.5 (230.6) ng/dL	High 1476.5 (734.9) ng/dL	1.73	Serum	Direct CIA	Random/ not fasting
Cherrier *et al.* [52]	Executive function; language; verbal memory; visuospatial memory; verbal learning; visuomotor function	32	28	6.00	Injection	100 mg/wk	76.00 (63–85)	Impaired (AD + MCI)	Normal 403.5 (172.9) ng/dL	High 1441.9 (230.6) ng/dL	5.11	Serum	Direct CIA	Random/ not fasting
Haren *et al.* [53]	Cognitive status; visuomotor function; attention/working memory	76	58	52.00	Pellets	80 mg/d	68.50 (60–86)	Normal	Normal 482.4 (130.6) ng/dL	Normal 449.3 (236.6) ng/dL	0.02	Serum	Direct CIA	Morning/ fasting
Lu *et al.* [43][h]	Cognitive status; verbal memory; visuomotor function; visuospatial function	20	14	24.00	Gel (1%)	75 mg/d	69.80 (8.7)	Impaired (AD)	Normal 387.7 (76.6) ng/dL	Normal 597.1 (554.3) ng/dL	0.64	Serum	Direct RIA	NR/NR
Lu *et al.* [43][h]	Cognitive status; verbal memory; visuomotor function; visuospatial function	29	20	24.00	Gel (1%)	75 mg/d	62.36 (6.7)	Normal	Normal 385.8 (170.1) ng/dL	Normal 737.5 (241.9) ng/dL	1.77	Serum	Direct RIA	NR/NR

(*Continued*)

doi: 10.1210/js.2019-00119 | Journal of the Endocrine Society | 1469

Table 1. Description of Included Studies (Continued)

Study	Cognitive Outcome	n[a] Pre	n[b] Post	Duration[c]	Type[d]	Doses	Mean Age (Range/SD)[e]	Cognitive Status	Gonadal Status Pre[f]	Gonadal Status Post[f]	TS Effect[g]	T Medium	Assay	Time
Cherrier *et al.* [54]	Verbal memory; visuospatial memory	57	50	6.00	Injection	50, 100, or 300 mg/6 wk	67.00 (56–78)	Normal	Normal 396.3 (160.5) ng/dL	High 1915.5 (1230.8) ng/dL	2.50	Serum	Direct CIA	Random/not fasting
Maki *et al.* [55]	Executive function; verbal memory; verbal learning; visuospatial learning; attention/working memory	15	15	12.90	Injection	200 mg/2 wk	73.90 (66–86)	Normal	Normal 10.2 ± 3.2 pg/mL	High 970.21 (359.1) ng/dL	1.82	Serum	Direct RIA and CIA	NR/NR
Vaughan *et al.* [56]	Executive function; verbal memory; verbal learning; visuospatial learning; visuospatial function; attention/working memory	47	32	156.00	Injection	200 mg/2 wk	70.80 (65–83)	Normal	Low–normal 285.3 (46.1) ng/dL	Normal 587.9 (279.5) ng/dL	1.10	Serum	NR	Morning/NR
Emmelot-Vonk *et al.* [57]	Executive function; verbal memory; verbal learning; visuospatial function	237	223	26.00	Pellets	160 mg/d	67.25 (60–80)	Normal	Low–normal 317.0 (54.8) ng/dL	Low–normal "unchanged"	−0.56	Serum	Direct CIA	Morning/fasting
Young *et al.* [44][i]	Executive function; language; verbal memory; verbal learning; visuospatial learning; visuospatial function; attention/working memory	13	13	6.00	Gel	100 mg/d	29.31 (3.3)	Normal	Normal 411 (125.8) pM	Normal 541.9 (310.2) pM	0.23	Serum	Direct RIA	NR/NR
Young *et al.* [44][i]	Executive function; language; verbal memory; verbal learning; visuospatial learning; visuospatial function; attention/working memory	15	15	6.00	Gel	75 mg/d	67.40 (5.5)	Normal	Normal 241 (65.5) pM	Normal 347.6 (155.2) pM	1.06	Serum	Direct RIA	NR/NR
Borst *et al.* [58]	Executive function; visuospatial memory; visuospatial learning; visuospatial function; attention/working memory	30	19	52.00	Injection	125 mg/wk	70.00 (8.9)	Normal	Low–normal 245.0 (73.0) ng/dL	Normal 474.0 (193.5) ng/dL	1.46	Serum	Direct CIA	NR/NR
Cherrier *et al.* [59]	Executive function; language; verbal memory; visuospatial memory; verbal learning; visuospatial learning; visuospatial function; reaction time	22	19	24.00	Derma gel	50–100 mg/d[j]	70.50 (60–88)	Impaired (MCI)	Low–normal 308.2 (92.1) ng/dL	Normal 600.7 (19.7) ng/dL	1.91	Serum	LC-MS/MS	Random/not fasting
Huang *et al.* [60]	Executive function; language; verbal memory; visuospatial memory; verbal learning; visuospatial learning	308	240	156.00	Gel (1%)	7.5 g 1% T gel/d	67.55 (5.10)	Normal	Low–normal 305.5 (63.4)	Normal 567.7 (265.1) ng/dL	1.40	Serum	Direct IA	Morning/fasting
Melehan *et al.* [61]	Executive function; reaction time	67	54	18.00	Injection	1000 mg/6 wk	49.00 (1.6)	Normal	Normal 352.7 (161.4) ng/dL	Normal 539.04 (115.16) ng/dL	1.40	Serum	LC-MS/MS	Morning/NR

(*Continued*)

Table 1. Description of Included Studies (Continued)

Study	Cognitive Outcome	n[a] Pre	n[b] Post	Duration[c]	Type[d]	Doses	Mean Age (Range/SD)[e]	Cognitive Status	Gonadal Status Pre[f]	Gonadal Status Post[f]	TS Effect[g]	T Medium	Assay	Time
Wahjoepramono *et al.* [62]	Cognitive status; verbal memory; verbal learning	50	44	24.00	Cream (5%)	50 mg/d	61.05 (7.7)	Normal	Normal 474.4 (126.8) ng/dL	Normal 769.5 (348.7) ng/dL	1.40	Serum	LC-MS/MS	NR/NR
Resnick *et al.* [63]	Cognitive status; executive function; verbal memory; verbal learning; visuospatial learning; reaction time	493	438	52.00	Gel (1%)	5.00 g 1% T gel/d	72.20 (6.0)	Impaired (AAMI)[k]	Low–normal 234.4 (65.2) ng/dL	Normal 490 (86.2) ng/dL	1.80	Serum	LC-MS/MS	NR/NR

Abbreviations: AAMI, age-associated memory impairment; AD, Alzheimer's disease patients; CIA, chemiluminescent immunoassay; IA, immunoassay; LC, liquid chromatography; MCI, mild cognitively impaired patients; MS/MS, tandem mass spectrometry; NR, not reported; Post, post-treatment assessment; Pre, pretreatment assessment; RIA, radioimmunoassay; T, testosterone.

[a]Number of participants randomized to the study.

[b]Number of participants completing the study.

[c]Duration of treatment is presented in weeks.

[d]Type refers to the type of treatment used to administer TS.

[e]Total sample mean age and associated range or SD depending on the available data.

[f]The treatment groups' gonadal status before and after treatment was categorized depending on mean total testosterone (TT) levels as follows: (i) low, TT ≤ 231 ng/dL; (ii) low–normal, TT between 232 and 320 ng/dL; (iii) normal, TT between 321 and 865 ng/dL; and (iv) supraphysiological (high), TT > 865 ng/dL. When only mean free testosterone (FT) levels were available, treatment groups' gonadal status were categorized as the following: (i) low, FT ≤ 174 pM; (ii) low–normal, FT between 175 and 220 pM; (iii) normal, FT between 221 and 846 pM; and (iv) supraphysiological (high), FT > 846 pM. In one case [49], only mean bioavailable testosterone (BT) levels were provided, and these were categorized as normal (BT between 72 and 235 ng/dL). Mean TT, FT, or BT and associated SDs are shown below the gonadal category.

[g]The effect of TS on testosterone levels was calculated as an effect size (Hedges *g*) for the increase in serum testosterone in the active vs control group from pre- to post-treatment. Positive values indicate an effect in the expected direction, *i.e.*, largest increase in the active groups' testosterone levels compared with the control group.

[h]Lu *et al.* [43] presented separate data on Alzheimer's disease patients and healthy men, and the reference is thus represented with two different studies in the table.

[i]Young *et al.* [44] presented separate data on young men and older men, and the reference is thus represented with two different studies in the table.

[j]Treatment doses were adjusted to keep testosterone concentrations within specified ranges.

[k]Resnick *et al.* [63] classified patients as having AAMI when they had both subjective memory complaints and relative impairment on objective tests of memory performance.

doi: 10.1210/js.2019-00119 | Journal of the Endocrine Society | 1471

whereas six studies included men who had testosterone levels in the low–normal range (mean total testosterone between 232 and 320 ng/dL).

The combined effect of TS on the primary outcome of overall cognitive functioning was small and failed to reach statistical significance (Hedges $g = 0.09$; CI 95%: −0.02 to 0.19, K = 23; see Table 2 and Fig. 2). The pooled effect sizes for all individual cognitive domains were small and statistically nonsignificant ($g = -0.04$ to 0.19, $P = 0.061$ to 0.989). When studies were grouped according to the proposed categorical moderators, small and statistically significant pooled effects were found for the following: (i) studies assessing men with mean testosterone levels within the normal range at baseline ($g = 0.17$, CI 95%: 0.00 to 0.33, K = 17), (ii) studies assessing younger men (mean age <68 years; $g = 0.20$, CI 95%: 0.02 to 0.36, K = 12), (iii) studies administering testosterone with injection ($g = 0.25$; CI 95%: 0.04 to 0.57, K = 11), (iv) studies assessing testosterone at random times ($g = 0.32$; CI 95%: 0.02 to 0.61, K = 6), and (v) nonindustry-sponsored studies ($g = 0.29$; CI 95%: 0.06 to 0.52, K = 10). Heterogeneity generally appeared to be low to moderate ($I^2 = 0.0$ to 59.3).

The risk of bias assessment was challenged by insufficient reporting, particularly of details related to methods of randomization and allocation concealment (see Table 3). Only five studies were preregistered online, limiting the assessment of selective reporting. Industry sponsorship was reported in 10 studies. The authors of 12 papers were contacted by E-mail to clarify study characteristics, of which four responded. Although the result for the primary outcome of overall cognitive functioning did not reach statistical significance, a funnel plot suggested possible publication bias in the direction of positive outcomes (Egger test, $P = 0.053$). The effects of four "missing" studies were imputed, resulting in an adjusted effect approaching zero ($g = 0.04$; Fig. 3). Although all fail-safe ns for the five statistically significant results failed to meet the criterion (Table 2), there were no indications of publication bias.

When explored with meta-regression in unadjusted models, the effect of TS on testosterone emerged as a statistically significant moderator of the overall pooled effect, with studies with larger effects of TS on testosterone showing stronger effects on cognitive functioning (B = 0.09). None of the other suggested moderators of the overall pooled effect reached statistical significance (Table 4). The results for three moderators showed a statistical trend ($P < 0.10$). This included TS type, with studies using gel/cream and studies using pellets, both showing weaker effects than studies using injection (B = −0.27 and −0.36). A trend toward larger effects was found for studies assessing testosterone at random times compared with those assessing testosterone in the morning (B = 0.27), and a trend for weaker effects was found for industry-sponsored studies compared with studies not sponsored by the industry (B = −0.23). When the effect of TS on testosterone was explored as a moderator of the overall pooled effects in studies with younger and older men, respectively, a trend toward studies with larger effects of TS on testosterone, showing stronger effects on cognitive functioning (B = 0.14, K = 11), was found in the subgroup of studies assessing older men. When intercorrelated moderators were explored in adjusted models, the effect of TS on testosterone levels was no longer a statistically significant moderator of the overall pooled effect after adjusting for TS type and time of day of testosterone measurement, respectively (B = 0.08 and 0.07; see Table 4).

When imputing 0 as the effect size for five neuropsychological outcomes not reported in two included studies, the effects were reduced to statistical nonsignificance in four out of the five study subgroups with statistically significant effects, *i.e.*, (i) studies assessing men with normal testosterone levels, (ii) studies assessing younger men, (iii) studies using injection, and (iv) studies measuring testosterone at random times during the day. When winsorizing the effects of two outlier studies, the results were generally similar to those obtained with nonwinsorized data. The exception was in the study subgroup investigating men with normal testosterone levels, where the effect failed to reach significance when outliers were winsorized (see Table 5).

3. Discussion

Our meta-analysis of 23 RCTs of the effect of TS on cognitive functioning in men revealed a negligible (Hedges $g = 0.09$) and statistically nonsignificant effect on the primary outcome of

Table 2. Pooled Effect Sizes Across Outcomes

	Sample Size		Heterogeneity[a]				Effect Size[b]			Fail-Safe n[c]	Criterion[d]
Outcome	**K**	**n**	***Q***	***df***	***P***	***I*²**	***g***	**95% CI**	***P***		
Overall combined effect	23	1638	20.34	22	0.562	0.00	0.09	−0.02 to 0.19	0.108	—	—
Cognitive domain											
Attention/working memory	11	488	7.028	10	0.723	0.00	0.16	−0.07 to 0.33	0.061	—	—
Cognitive status	6	657	1.24	5	0.624	0.00	−0.04	−0.19 to 0.11	0.624	—	—
Executive function	16	1447	16.15	15	0.372	7.11	0.02	−0.07 to 0.12	0.611	—	—
Language	7	423	9.64	6	0.141	37.77	0.07	−0.11 to 0.24	0.442	—	—
Verbal learning	10	1201	5.51	9	0.788	0.00	0.03	−0.62 to 0.11	0.572	—	—
Verbal memory	16	1378	23.11	15	0.082	35.10	0.08	−0.06 to 0.22	0.256	—	—
Visuomotor function	10	563	22.12	9	**0.009**	59.31	0.15	−0.14 to 0.43	0.311	—	—
Visuospatial function	10	878	3.99	9	0.912	0.00	0.00	−0.11 to 0.12	0.989	—	—
Visuospatial learning	7	902	10.82	6	0.094	44.56	0.19	−0.12 to 0.49	0.234	—	—
Visuospatial memory	7	454	6.81	6	0.339	11.92	0.05	−0.11 to 0.22	0.525	—	—
Visuospatial skills (combined)[e]	16	1322	12.25	15	0.660	0.00	0.04	−0.06 to 0.15	0.420	—	—
Participants' cognitive status											
Cognitively normal	18	1072	15.21	17	0.580	0.00	0.14	−0.01 to 0.28	*0.060*	—	—
Cognitively impaired	5	566	4.09	4	0.394	2.26	0.03	−0.15 to 0.21	0.739	—	—
Participants gonadal status											
Normal	17	571	18.06	16	0.320	11.42	0.17	0.00 to 0.33	**0.048**	7	95
Low/normal	6	1067	0.69	5	0.983	0.00	0.02	−0.13 to 0.17	0.785	—	—
Age dichotomized											
Young age (<68 y)	12	591	10.50	11	0.486	0.00	0.20	0.02 to 0.36	**0.032**	7	70
Old age (≥68 y)	11	1048	7.65	10	0.663	0.00	0.03	−0.10 to 0.16	0.676	—	—
Administration type											
Gel/cream	8	914	2.40	7	0.934	0.00	0.02	−0.13 to 0.16	0.839	—	—
Pellets	2	281	1.26	1	0.263	20.34	−0.11	−0.50 to 0.28	0.593	—	—
Injection	11	325	11.18	10	0.343	10.59	0.25	0.04 to 0.57	**0.021**	10	65
Time of measure											
Morning	9	1277	9.86	8	0.275	18.86	0.07	−0.09 to 0.23	0.378	—	—
Not stated	8	181	3.00	7	0.941	0.00	0.07	−0.20 to 0.34	0.608	—	—
Random	6	180	5.37	5	0.372	6.92	0.32	0.02 to 0.61	**0.037**	3	40
T measurement assay											
LC-MS/MS	4	611	2.18	3	0.536	0.00	0.03	−0.12 to 0.18	0.691	—	—
Direct CIA/RIA	19	1027	17.15	18	0.513	0.00	0.14	−0.01 to 0.29	*0.064*	—	—

(*Continued*)

doi: 10.1210/js.2019-00119 | Journal of the Endocrine Society | 1473

Table 2. Pooled Effect Sizes Across Outcomes (Continued)

	Sample Size		Heterogeneity[a]				Effect Size[b]			Fail-Safe n[c]	Criterion[d]
Outcome	**K**	**n**	***Q***	***df***	***P***	***I²***	***g***	**95% CI**	***P***		
Publication year											
Early studies	11	360	15.63	10	0.111	36.01	0.24	−0.02 to 0.49	*0.069*	—	—
Late studies	12	1278	3.02	11	0.990	0.00	0.04	−0.09 to 0.17	0.524	—	—
Industry sponsoring											
Sponsored	12	1254	3.91	11	0.972	0.00	0.05	−0.08 to 0.18	0.470	—	—
Not sponsored	10	308	10.37	9	0.322	13.18	0.29	0.06 to 0.52	**0.014**	12	60

Abbreviations: *df*, degrees of freedom; K, number of studies in the analysis; n, number of subjects in the analysis.

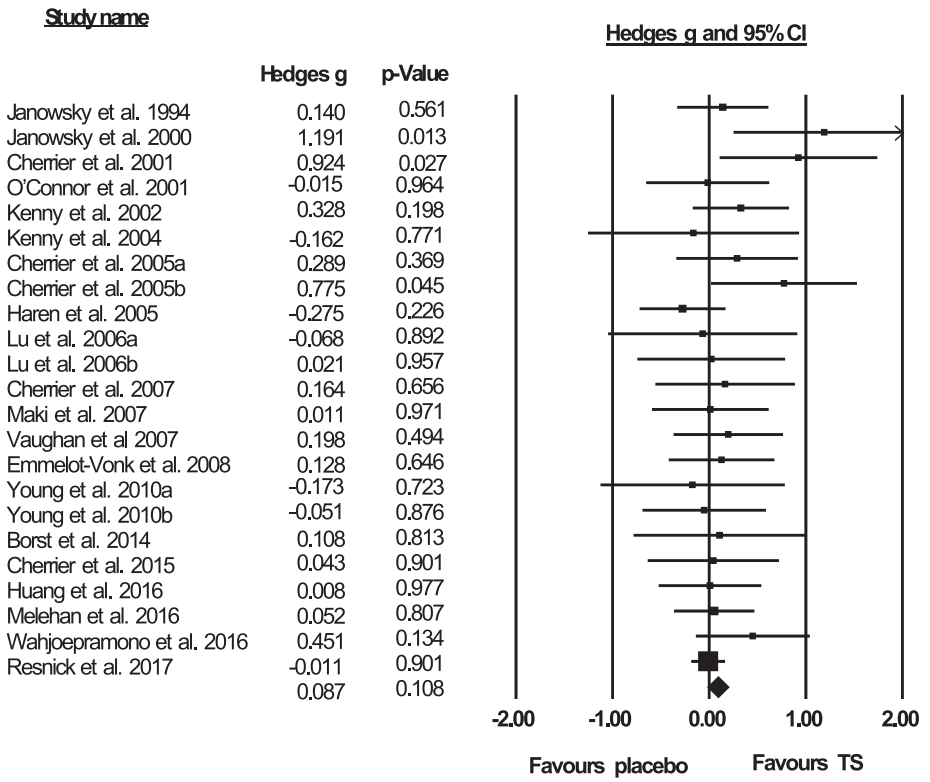
[a]*Q* statistic, $P < 0.1$ taken to suggest heterogeneity (bold) [33]; I^2 statistic, 0% (no heterogeneity), 25% (low heterogeneity), 50% (moderate heterogeneity), 75% (high heterogeneity) [31].

[b]Effect size = Hedges *g*. A positive value indicates an effect size in the hypothesized direction, *i.e.*, improvement in the active groups' cognitive functions compared with the control group. Conventions: small (0.2), medium (0.5), and large (0.8) [28]. Statistically significant *P* (<0.05) is in bold. Statistically trending *P* (>0.05 <0.10) is in italics.

[c]In case of statistically significant effects, the robustness of findings was examined by calculation of the fail-safe n (number of nonsignificant studies that would bring *P* to >0.05) [36].

[d]Fail-safe n exceeding the criterion (5K + 10) indicates a robust result [36].

[e]Cognitive domains using visuospatial skills (visuospatial function, visuospatial learning, and visuospatial memory) were pooled.



Meta Analysis

Figure 2. Individual effect sizes and forest plots. Individual studies' overall effect sizes (Hedges *g*) are shown in the second column and are the unit of analysis. The summary pooled effect size (Hedges *g*) computed with a random effect model is shown in the bottom row of the second column. Notation: Forest plot. Each square represents the effect size of one individual study. Size of the square indicates the relative weight assigned to the study, with more weight assigned to more precise studies, as indicated by larger squares. Lines: Each line represents a 95% CI of the effect size of the study. The diamond represents the summary pooled *g* of all of the individual *g*s. Precision of the summary effect is indicated by the width of the diamond.

overall cognitive functioning. When further taking possible publication bias into consideration, the pooled effect approached zero. Five (17%) of the 29 secondary analyses of various study subsets reached statistical significance, including studies assessing men with normal testosterone levels, studies assessing younger men (age <68 years), studies administering testosterone with injection, nonindustry-sponsored studies, and studies assessing testosterone at random times. The latter subgroup is of interest, as this suboptimal assessment of testosterone [64] may affect the estimate of TS effects on testosterone levels, which emerged as the only statistically significant moderator of the overall pooled effect when explored with meta-regression. The statistically significant effects all failed to meet the fail-safe number criterion, indicating less than robust results. The current study highlights several questions that remain unanswered and should be investigated further.

First, it may have seemed optimistic to expect an effect of TS on cognitive functions in men with sufficient physiological testosterone at the outset. Rather than the examination of the effects of TS in men with physiological levels within normal ranges, it would assumingly be of most interest to clinicians to know whether testosterone replacement has an effect in men with clinical hypogonadism. Unfortunately, we were unable to examine this, as none of the included studies treated samples of men with clinically low testosterone levels. However, six

Table 3. Risk of Bias of Included Studies

	Random Sequence Generation	Allocation Concealment	Blinding of Participants and Personal	Incomplete Outcome Data	Selective Reporting	Pharmaceutical Industry Sponsor
Janowsky *et al.* (45)	🟡	🟡	🟢	🟡	🟡	🟢
Janowsky *et al.* (46)	🟡	🟡	🟢	🟡	🟡	🟢
Cherrier *et al.* (47)	🟢	🔴	🔴	🔴	🔴	🟢
O'Connor *et al.* (48)	🟢	🟢	🔴	🟢	🟡	🟢
Kenny *et al.* (49)	🟡	🟢	🟢	🔴	🟡	🔴
Kenny *et al.* (50)	🟡	🟡	🟡	🟢	🟡	🟢
Cherrier *et al.* (51)	🟢	🔴	🟢	🟢	🔴	🟢
Cherrier *et al.* (52)	🟢	🔴	🟢	🔴	🔴	🟢
Haren *et al.* (53)	🟢	🟢	🟡	🔴	🟡	🟡
Lu *et al.* (43)	🟡	🟡	🟡	🔴	🟡	🔴
Cherrier *et al.* (54)	🟢	🔴	🟢	🟢	🔴	🟢
Maki *et al.* (55)	🟢	🔴	🟢	🟢	🟡	🟢
Vaughan *et al.* (56)	🟢	🟡	🟢	🔴	🟡	🟢
Emmelot-Vonk *et al.* (57)	🟢	🟢	🟢	🔴	🟡	🔴
Young *et al.* (44)	🟡	🟡	🟢	🟡	🟡	🔴
Borst *et al.* (58)	🟡	🟡	🟡	🟡	🟢	🔴
Cherrier *et al.* (59)	🟢	🟢	🟢	🔴	🔴	🔴
Huang *et al.* (60)	🟢	🟢	🟢	🟢	🟢	🔴
Melehan *et al.* (61)	🟢	🟢	🟢	🟢	🟢	🔴
Wahjoepramono *et al.* (62)	🟢	🔴	🟢	🔴	🟢	🔴
Resnick *et al.* (63)	🟢	🟢	🟢	🟢	🟢	🔴

Studies were rated as having 🟢 low risk, 🔴 high risk, or 🟡 unclear risk of bias using the Cochrane Collaboration tool for assessing risk of bias [29].

studies examined men with testosterone in the low end of the normal range. Surprisingly, the pooled effect of these studies were negligible and statistically nonsignificant ($g = 0.02$), opposed to the small and statistically significant effect ($g = 0.17$; $P = 0.048$) found for the 17 studies of men with mean testosterone levels within the normal range. However, there are several reasons this could be a chance finding. First, the results were less robust, *i.e.*, did not reach the fail-safe number criterion. Second, the effect no longer reached statistical

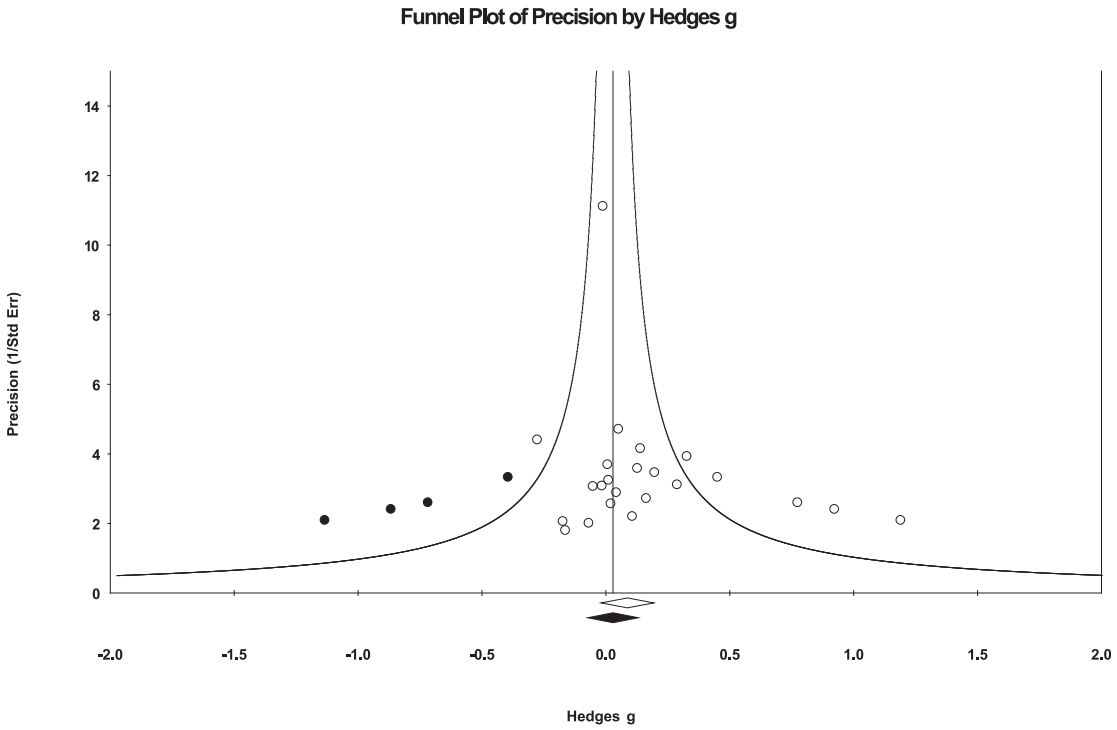


Figure 3. Funnel plot of the overall pooled effect with missing studies imputed. Funnel plot of precision [1/SE (Std Err)] by effect sizes (Hedges *g*) with imputed studies. Each open circle represents precision as a measure of the sample size on the *y*-axis as a function of the effect size of each independent study on the *x*-axis. Closed circles represent imputed studies. The open diamond in the bottom of the plot indicates the summary effect size of the analysis. The closed diamond indicates the summary effect size with imputed studies included. A 95% CI is indicated by the lines. Within these lines, 95% of the circles are expected to be located. The overall pooled effect size is indicated by a vertical line in the middle of the plot.

significance when we imputed zero as the effect size for five neuropsychological outcomes not reported in two studies. Finally, the result failed to reach statistical significance when winsorizing the outlier effect sizes of two studies. Furthermore, we have no clear explanation for our findings of statistically significant effects in the two subsets of studies investigating younger men (age <68 years) and studies assessing testosterone at random times during the day. Regarding men of younger age, we would expect these to benefit less from TS than men of older age who are more likely to have low levels of physiological testosterone and reduced cognitive reserve as a result of aging. Likewise, whereas the assessment time may affect the estimate of the effect of TS on testosterone levels, we see no apparent reason for a direct moderating effect of testosterone assessment time on the effect of TS on overall cognitive functioning. Given this lack of theoretical support for the findings, in addition to the lack of robustness of the results, we are inclined to believe these are chance findings.

Second, it remains to be explored in detail whether TS may actually be beneficial to cognitive functioning in the context of neurodegenerative disease. Whereas we found a smaller effect in subgroup analyses of studies with cognitively impaired men (g = 0.03) compared with studies of cognitively normal men (g = 0.14), only five studies assessed cognitively impaired men (n = 566) with the largest of these studies [63], including men with "age-associated memory impairment" (AAMI) (n = 493). As AAMI excludes men with neuropsychological test scores 2 SDs below the scores of age-matched men on tests of paragraph recall or visual memory [63], these participants were unlikely to suffer from neurodegenerative diseases. Thus, rather than the representation of the effect of TS on cognitive functioning in men with neurodegenerative diseases, the subgroup analysis of cognitively

Table 4. Results From Meta-Regression Analyses

		Unadjusted Models[a]			Adjusted Models[b]			
Variable	**K**	**B[c]**	**95% CI**	***P***	**B[c]**	**95% CI**	***P***	**VIF[d]**
Treatment characteristics								
TS Type	23							
Gel/cream (vs injection)		−0.27	−0.47 to 0.02	*0.073*				
Patches (vs injection)		−0.01	−0.41 to 0.38	0.949				
Pellets (vs injection)		−0.36	−0.76 to 0.04	*0.080*				
Treatment duration, wk	23	−0.00	−0.00 to 0.00	0.336				
TS effect (Hedges *g*)	23	0.09	0.01 to 0.19	**0.044**	0.08	−0.03 to 0.19	0.140	1.47
						Adj. TS type		
					0.07	−0.33 to 0.18	0.180	1.37
						Adj. for time of measurement		
TS effect, young age (<68 y)	12	0.08	−0.04 to 0.19	0.183				
TS effect, old age (≥68 y)	11	0.14	−0.02 to 0.30	*0.077*				
T measurement characteristics								
Time of measurement	23							
Random times (vs morning)		0.27	−0.05 to 0.58	*0.095*				
Not stated (vs morning)		0.02	−0.27 to 0.32	0.877				
Measurement assay	23							
Direct CIA/RIA (vs LC-MS/MS)		0.11	−0.10 to 0.32	0.317				
Participant characteristics								
Mean age	23	0.00	−0.01 to 0.01	0.758				
Cognitive status	23							
Impaired (vs normal)		−0.11	−0.33 to 0.10	0.309				
Gonadal status	23							
Low–normal (vs normal)		−0.14	−0.35 to 0.08	0.208				
Study characteristics								
Early publication (vs late)	23	−0.15	−0.39 to 0.08	0.194	−0.19	−0.49 to 0.11	0.205	1.68
						Adj. for sponsorship		
Sponsored (vs not sponsored)	22	−0.23	−0.48 to 0.02	*0.067*	−0.13	−0.42 to 0.18	0.407	1.66
						Adj. for publication y		

Statistically significant P (<0.05) is in bold. Statistically trending P (>0.05 <0.10) is in italic.
Abbreviations: *Adj.*, adjusted; VIF, variance inflation factor.
[a]Variables were explored individually in unadjusted models.
[b]When theoretically sound intercorrelations between variables reached statistical significance ($P < 0.05$), they were explored together in adjusted models.
[c]The association between moderators and the magnitude of the effect is expressed in unstandardized regression coefficients (B).
[d]The variance inflation factor (VIF), a measure of multicollinearity, was calculated when variables were explored together in adjusted models. Conventions: VIF > 10 indicates a serious problem with bias [39, 40]. If VIF is substantially greater than one, then the regression may be biased [40].

impaired men in the present meta-analysis represents an estimate of the effect of TS on men with various degrees of cognitive impairment ranging from AAMI and mild cognitive impairment to Alzheimer's disease. Given the evidence that points to the protective effect of testosterone against Alzheimer's neuropathology, it would be of interest to examine the effect of TS in a larger and more homogeneous group of patients with neurodegenerative diseases.

Third, whether TS will have a beneficial effect on cognitive functioning may also depend on how it is administered. The most recent trial [63] administered TS using gel and found no effect on cognitive functions. It has been speculated that an injectable type of TS that leads to spikes of testosterone, rather than a gel-based type that releases a steady dose, could have a transient effect on cognitive functioning [65]. This is supported by research indicating that transdermal TS is not always adequately absorbed [66], which is further supported by the present meta-analysis showing the largest, statistically significant effect in the subgroup of studies that used injection ($g = 0.25$), in addition to the statistically trending meta-regression findings of weaker effects in studies using gel/cream and pellets compared with injection

Table 5. Sensitivity Analyses of Statistically Significant Effects With Original and Revised Pooled Effects

Outcome	Analysis	Sample Size		Heterogeneity[a]				Effect Size[b]			Fail-Safe n[c]	Criterion[d]
		K	n	*Q*	*df*	*P*	I^2	*g*	95% CI	*P*		
Overall cognitive functioning	Original	23	1638	20.34	22	0.562	0.00	0.09	−0.02 to 0.19	0.108	—	—
	Imputed[e]	23	1638	14.71	22	0.874	0.00	0.07	−0.04 to 0.18	0.189	—	—
	Winsorized[f]	23	1638	18.24	22	0.692	0.00	0.07	−0.03 to 0.16	0.169	—	—
Normal gonadal status	Original	17	571	18.06	16	0.320	11.42	0.17	0.00 to 0.33	**0.048**	7	95
	Imputed[e]	17	571	16.17	16	0.441	1.079	0.14	−0.01 to 0.30	*0.065*	—	—
	Winsorized[f]	17	571	15.33	16	0.501	0.00	0.16	−0.00 to 0.31	*0.054*	—	—
Young age	Original	12	591	10.50	11	0.486	0.00	0.20	0.02 to 0.36	**0.032**	7	70
	Imputed[e]	12	591	7.71	11	0.739	0.00	0.17	−0.01 to 0.35	*0.062*	—	—
	Winsorized[f]	12	591	9.71	11	0.557	0.00	0.18	0.02 to 0.35	**0.033**	6	70
Using injection	Original	11	325	11.18	10	0.343	10.59	0.25	0.04 to 0.57	**0.021**	11	70
	Imputed[e]	11	355	6.97	10	0.728	0.00	0.19	−0.01 to 0.38	*0.068*	—	—
	Winsorized[f]	11	355	11.19	10	0.342	10.69	0.24	0.04 to 0.45	**0.021**	11	70
Random times	Original	6	180	5.37	5	0.372	6.92	0.32	0.02 to 0.61	**0.037**	3	40
	Imputed[e]	6	180	1.60	5	0.902	0.00	0.21	−0.09 to 0.50	0.169	—	—
Not sponsored	Original	10	308	10.37	9	0.322	13.18	0.29	0.06 to 0.52	**0.014**	12	60
	Imputed[e]	10	308	6.49	9	0.690	0.00	0.22	0.01 to 0.43	**0.044**	3	60

Sensitivity analyses of the primary outcome, *i.e.*, the effect of TS on overall cognitive functioning, defined as the individual cognitive domain scores aggregated into one combined estimate, and of five study subsets showing statistically significant improvements: (i) studies of men with mean testosterone levels within normal range (total testosterone between 321 and 865 ng/dL), *i.e.*, normal gonadal status; (ii) studies of younger men (age <68 y); (iii) studies administering testosterone with injection; (iv) studies assessing testosterone at random times; and (v) trials not sponsored by the industry.

[a] *Q* statistic, $P < 0.1$ taken to suggest heterogeneity (bold) [33]; I^2 statistic, 0% (no heterogeneity), 25% (low heterogeneity), 50% (moderate heterogeneity), 75% (high heterogeneity) [31].

[b] Effect size = Hedges *g*. A positive value indicates an effect size in the hypothesized direction, *i.e.*, improvement in the active groups' cognitive functions compared with the control group. Conventions: small (0.2); medium (0.5); large (0.8) [28]. Statistically significant *P* (<0.05) is highlighted in bold. Statistically trending *P* (>0.05 <0.10) is highlighted in italic.

[c] In case of statistically significant effects, the robustness of findings was examined by calculation of the fail-safe n (number of nonsignificant studies that would bring *P* to >0.05) [36].

[d] Fail-safe n exceeding the criterion (5K + 10) indicates a robust result [36].

[e] Imputed = imputed effect size of zero in two cases [47, 52], when included publications stated they had obtained nonsignificant results on one or more neuropsychological tests without presenting the data.

[f] A range of 2 SDs below and above the global effect size was used to truncate outliers. In two study subsets, *i.e.*, studies assessing testosterone at random times and studies not sponsored by the industry, no studies were below or above this range; thus, no sensitivity analyses with truncated outliers were conducted.

(B = −0.27 and −0.36). However, the difference between routes of administration may be a result of the effect of treatment on circulating testosterone and thus, the effect of testosterone on different organs, and it may also be a matter of differing doses given. In other words, this issue may only be resolved by head-to-head studies examining injection vs transdermal administration aiming at dose equivalency.

Fourth, yet another unanswered question relates to the concept that whereas industry sponsorship of pharmacotherapy trials is associated with more favorable outcomes [30], the effects reported in industry-sponsored trials were of a smaller magnitude ($g = 0.05$) than the effects found in nonindustry-sponsored trials ($g = 0.29$); the difference approaching statistical significance ($P = 0.075$). Whereas we have no clear explanation, a *post hoc* analysis revealed that studies published in 2007 and later were more likely to be industry sponsored than the early studies published before 2007. As the effects found in early studies ($g = 0.24$) were larger than in later studies ($g = 0.04$), this could be a partial explanation for the unexpected finding.

Finally, the hypothesis that TS should benefit cognitive functioning may be overly simplistic. It has been suggested that the association between low testosterone and increased

risk for Alzheimer's may not stem from testosterone depletion *per se* but rather from an increase in serum gonadotropins as a result of loss of negative regulation of testosterone on the hypothalamus and pituitary [67]. Consequently, rather than the focus solely on TS, future studies may need to focus on attempts to balance the dynamics of the hypothalamic-pituitary axis and thereby all sex hormones [65]. It also seems relevant to examine the role of estradiol for cognitive functioning in men, as estradiol primarily stems from the aromatization of testosterone and plays an important role in bone metabolism, body composition, and sexual function [68, 69].

A. Strengths and Limitations

We conducted this review based on an *a priori*-defined protocol and used a rigorous methodological approach. Among additional strengths are the comprehensive search strategy; the relatively homogenous studies available, as indicated by heterogeneity statistics; and our detailed examination of the possible role of between-study methodological differences. Some limitations should also be noted. First, our search was limited to English-language publications and did not include the "grey literature." English language as an inclusion criterion could possibly introduce a risk of overlooking important results reported in other languages. However, as we would be unable to include all languages as a result of restricted language competencies in the group of authors, this could in itself introduce additional bias. Furthermore, there is evidence to suggest that English-language restriction does not introduce systematic bias. A systematic review of reviews examining a total of 361 meta-analyses of studies thus found no evidence of a systematic bias from the use of language restrictions in systematic review-based meta-analyses in conventional medicine [70]. It is also worth noting that in the present review, none of the 220 non-English records identified in the databases that were excluded during the initial abstract screening met the remaining inclusion criteria. Concerning the grey literature, the inclusion of data from unpublished studies can itself introduce bias, as the studies that can be identified may be an unrepresentative sample of all unpublished studies [29]. Second, we categorized patients in included studies as having "low," "low–normal," "normal," or "supraphysiological" testosterone, respectively, depending on reported mean total testosterone levels. This approach has several drawbacks, given that reported ranges for testosterone concentrations vary among laboratories and assays and that testosterone levels should not be evaluated in isolation without taking important confounding factors, such as sexual hormone-binding globulin levels, into account [64]. However, the included papers did not present sufficient information to enable us to evaluate variations in reference ranges, and whereas some papers presented sexual hormone-binding globulin levels, this was as mean statistics only, thus not allowing for individually based evaluations. Despite these challenges, we attempted to categorize studies depending on testosterone levels to give an indication of the hypogonadal status of the participants. For the sake of transparency, we also reported the mean testosterone levels and associated SDs together with the hypogonadal categories in Table 1. Additional limitations relating to the available literature itself included the small number of available studies included in some of the subgroup analyses, making it difficult to draw robust conclusions regarding subgroup differences. Our ability to evaluate the strength of the evidence was generally limited, as many reports, in particular, the early studies, had not been preregistered online and failed to provide sufficient information concerning randomization and allocation concealment.

B. Strengths and Weaknesses in Relation to Other Studies

Six previous narrative reviews of the relationship between testosterone and cognitive functioning in men are relevant for comparison [15–20]. Of these, only one [16] included only RCTs and thus, appears to be the most comparable with the present review, which was also restricted to RCTs. That review [16] evaluated 156 RCTs assessing the effect of TS on various outcomes, including cognition (K = 23). In accordance with our findings, the authors

concluded that prescription of TS for improving cognitive functioning is without support from the available evidence. However, the review may be limited in several ways. First, in addition to studies assessing cognitive functioning by means of objective neuropsychological testing, the review also included studies assessing cognitive functioning by means of self-reports only. This may be problematic as self-reported cognitive functioning is more likely to reflect emotional distress rather than objective cognitive impairment. Often, self-reported measures of cognition tend not to be highly correlated with objective neuropsychological outcomes [71–73]. Second, as the remaining five previous narrative reviews [15, 17–20], that systematic review [16] was limited to simple vote counting of statistically significant results when evaluating the effect of TS on cognitive functioning in men. In contrast, in the present review, we have evaluated the effect of TS on cognitive functioning by means of meta-analysis, which more accurately estimates the overall effect of TS on cognitive functioning. The present systematic review with meta-analysis addresses the efficacy of TS on cognitive functioning in men, thus, addressing a gap in the existing knowledge within the field.

C. Meaning of the Study—Explanations and Implications

Taken together, the available RCTs do not support a beneficial effect of TS on cognitive functioning in men with testosterone levels within normal ranges. There is, thus, no evidence for prescribing TS for improving cognitive functioning in men with no clinical signs of hypogonadism, in particular, when considering the potential adverse effects and the inadequately understood risk for cardiovascular events associated with TS [64, 74]. Of five statistically significant subgroup analyses, four were unexplainable and appear to be chance findings given the lacking robustness of the results. Future studies could investigate whether TS may have positive effects on cognitive functioning in cases where testosterone replacement is indicated to induce and maintain secondary sex characteristics and correct symptoms of hypogonadism and where treatment is continued for a sufficient length of time for it to have a detectable effect on cognitive functions.

Acknowledgments

Correspondence: Cecilie R. Buskbjerg, MSc, Aarhus University, Bartholins Alle 9, Aarhus C, 8000, Denmark. E-mail: cdrc@psy.au.dk.

Disclosure Summary: The authors have nothing to disclose.

Data Availability: The datasets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author by reasonable request.

References and Notes

1. Handelsman DJ. Global trends in testosterone prescribing, 2000-2011: expanding the spectrum of prescription drug misuse. *Med J Aust*. 2013;**199**(8):548–551.
2. Jacobsen R. The hidden dangers of testosterone replacement therapy you should know about. CNBC. 3 October 2018. Available at: www.cnbc.com/2018/10/03/the-hidden-dangers-of-testosterone-replacement-therapy-can-be-deadly.html. Accessed 14 February 2019.
3. Jasuja GK, Bhasin S, Rose AJ. Patterns of testosterone prescription overuse. *Curr Opin Endocrinol Diabetes Obes*. 2017;**24**(3):240–245.
4. Surampudi PN, Wang C, Swerdloff R. Hypogonadism in the aging male diagnosis, potential benefits, and risks of testosterone replacement therapy. *Int J Endocrin*. 2012;**2012**:625434.
5. Aversa A, Morgentaler A. The practical management of testosterone deficiency in men. *Nat Rev Urol*. 2015;**12**(11):641–650.
6. Hammond J, Le Q, Goodyer C, Gelfand M, Trifiro M, LeBlanc A. Testosterone-mediated neuroprotection through the androgen receptor in human primary neurons. *J Neurochem*. 2001;**77**(5):1319–1326.
7. Pike CJ, Carroll JC, Rosario ER, Barron AM. Protective actions of sex steroid hormones in Alzheimer's disease. *Front Neuroendocrinol*. 2009;**30**(2):239–258.

8. Ahlbom E, Prins GS, Ceccatelli S. Testosterone protects cerebellar granule cells from oxidative stress-induced cell death through a receptor mediated mechanism. *Brain Res*. 2001;**892**(2):255–262.
9. Brown CM, Mulcahey TA, Filipek NC, Wise PM. Production of proinflammatory cytokines and chemokines during neuroinflammation: novel roles for estrogen receptors alpha and beta. *Endocrinology*. 2010;**151**(10):4916–4925.
10. Wahjoepramono EJ, Wijaya LK, Taddei K, Martins G, Howard M, de Ruyck K, Bates K, Dhaliwal SS, Verdile G, Carruthers M, Martins RN. Distinct effects of testosterone on plasma and cerebrospinal fluid amyloid-beta levels. *J Alzheimers Dis*. 2008;**15**(1):129–137.
11. Schaie KW, Willis SL. The Seattle Longitudinal Study of Adult Cognitive Development. *ISSBD Bull*. 2010;**57**(1):24–29.
12. Akinola OB, Gabriel MO. Neuroanatomical and molecular correlates of cognitive and behavioural outcomes in hypogonadal males. *Metab Brain Dis*. 2018;**33**(2):491–505.
13. McGinty HL, Phillips KM, Jim HS, Cessna JM, Asvat Y, Cases MG, Small BJ, Jacobsen PB. Cognitive functioning in men receiving androgen deprivation therapy for prostate cancer: a systematic review and meta-analysis. *Support Care Cancer*. 2014;**22**(8):2271–2280.
14. McHugh DJ, Root JC, Nelson CJ, Morris MJ. Androgen-deprivation therapy, dementia, and cognitive dysfunction in men with prostate cancer: how much smoke and how much fire? *Cancer*. 2018;**124**(7):1326–1334.
15. Hua JT, Hildreth KL, Pelak VS. Effects of testosterone therapy on cognitive function in aging: a systematic review. *Cogn Behav Neurol*. 2016;**29**(3):122–138.
16. Huo S, Scialli AR, McGarvey S, Hill E, Tügertimur B, Hogenmiller A, Hirsch AI, Fugh-Berman A. Treatment of men for "low testosterone": a systematic review. *PLoS One*. 2016;**11**(9):e0162480.
17. Mohamad NV, Ima-Nirwana S, Chin K-Y. A review on the effects of testosterone supplementation in hypogonadal men with cognitive impairment. *Curr Drug Targets*. 2018;**19**(8):898–906.
18. Giagulli VA, Guastamacchia E, Licchelli B, Triggiani V. Serum testosterone and cognitive function in ageing male: updating the evidence. *Recent Pat Endocr Metab Immune Drug Discov*. 2016;**10**(1):22–30.
19. Holland J, Bandelow S, Hogervorst E. Testosterone levels and cognition in elderly men: a review. *Maturitas*. 2011;**69**(4):322–337.
20. Yalamanchi S, Dobs A. Debate position: cognition and mood are not improved in men administered exogenous testosterone therapy. *Curr Opin Urol*. 2017;**27**(6):525–531.
21. Linn MC, Petersen AC. Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Dev*. 1985;**56**(6):1479–1498.
22. Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol Bull*. 1995;**117**(2):250–270.
23. Aleman A, Bronk E, Kessels RPC, Koppeschaar HPF, van Honk J. A single administration of testosterone improves visuospatial ability in young women. *Psychoneuroendocrinology*. 2004;**29**(5):612–617.
24. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM*. New York: Churhill Livingstone; 1997.
25. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, Stewart L.The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Syst Rev*. 2012;**1**:2.
26. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*. 2009;**6**(7):e1000100.
27. Hedges L, Olkin I. *Statistical Methods for Meta-Analysis*. New York: Academic Press; 1985.
28. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New York: Lawrence Earlbaum Associates; 1988.
29. Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0*. Chichester, West Sussex, United Kingdom: The Cochrane Collaboration; 2011.
30. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev*. 2017;**2**:MR000033.
31. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;**327**(7414):557–560.
32. Sterne JAC, Egger M, Moher D. Adressing reporting biases. In: Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Intervention*. Chichester, United Kingdom: Wiley-Blackwell; 2008:297–333.
33. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol*. 1999;**150**(5):469–475.

34. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;**315**(7109):629–634.
35. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000;**56**(2):455–463.
36. Rosenthal. The file drawer problem and tolerance for null results. *Psychol Bull*. 1979;**86**(3):638–641.
37. Lipsey MW, Wilson DB. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage; 2001.
38. Borenstein M, Hedges L, Higgins JP, Rothstein H. *Introduction to Meta-Analysis*. West Sussex, UK; John Wiley & Sons; 2009.
39. Bowerman B, O'Conell R. *Linear Statistical Models: An Applied Approach*. Belmont, CA: Duxbury; 1990.
40. Myers R. *Classical and Modern Regression With Applications*. Boston: Duxbury; 1990.
41. Eaglewood N. *Comprehensive Meta-Analysis*. Englewood, NJ: Biostat; 2014.
42. Field A. *Discovering Statistics Using IBM SPSS Statistics*, 5th ed. London: SAGE; 2017.
43. Lu PH, Masterman DA, Mulnard R, Cotman C, Miller B, Yaffe K, Reback E, Porter V, Swerdloff R, Cummings JL. Effects of testosterone on cognition and mood in male patients with mild Alzheimer disease and healthy elderly men. *Arch Neurol*. 2006;**63**(2):177–185.
44. Young LA, Neiss MB, Samuels MH, Roselli CE, Janowsky JS. Cognition is not modified by large but temporary changes in sex hormones in men. *J Clin Endocrinol Metab*. 2010;**95**(1):280–288.
45. Janowsky JS, Oviatt SK, Orwoll ES. Testosterone influences spatial cognition in older men. *Behav Neurosci*. 1994;**108**(2):325–332.
46. Janowsky JS, Chavez B, Orwoll E. Sex steroids modify working memory. *J Cogn Neurosci*. 2000;**12**(3): 407–414.
47. Cherrier MM, Asthana S, Plymate S, Baker L, Matsumoto AM, Peskind E, Raskind MA, Brodkin K, Bremner W, Petrova A, LaTendresse S, Craft S. Testosterone supplementation improves spatial and verbal memory in healthy older men. *Neurology*. 2001;**57**(1):80–88.
48. O'Connor DB, Archer J, Hair WM, Wu FCW. Activational effects of testosterone on cognitive function in men. *Neuropsychologia*. 2001;**39**(13):1385–1394.
49. Kenny AM, Bellantonio S, Gruman CA, Acosta RD, Prestwood KM. Effects of transdermal testosterone on cognitive function and health perception in older men with low bioavailable testosterone levels. *J Gerontol A Biol Sci Med Sci*. 2002;**57**(5):M321–M325.
50. Kenny AM, Fabregas G, Song C, Biskup B, Bellantonio S. Effects of testosterone on behavior, depression, and cognitive function in older men with mild cognitive loss. *J Gerontol A Biol Sci Med Sci*. 2004;**59**(1):75–78.
51. Cherrier MM, Matsumoto AM, Amory JK, Ahmed S, Bremner W, Peskind ER, Raskind MA, Johnson M, Craft S. The role of aromatization in testosterone supplementation: effects on cognition in older men. *Neurology*. 2005;**64**(2):290–296.
52. Cherrier MM, Matsumoto AM, Amory JK, Asthana S, Bremner W, Peskind ER, Raskind MA, Craft S. Testosterone improves spatial memory in men with Alzheimer disease and mild cognitive impairment. *Neurology*. 2005;**64**(12):2063–2068.
53. Haren MT, Wittert GA, Chapman IM, Coates P, Morley JE. Effect of oral testosterone undecanoate on visuospatial cognition, mood and quality of life in elderly men with low-normal gonadal status. *Maturitas*. 2005;**50**(2):124–133.
54. Cherrier MM, Matsumoto AM, Amory JK, Johnson M, Craft S, Peskind ER, Raskind MA. Characterization of verbal and spatial memory changes from moderate to supraphysiological increases in serum testosterone in healthy older men. *Psychoneuroendocrinology*. 2007;**32**(1): 72–79.
55. Maki PM, Ernst M, London ED, Mordecai KL, Perschler P, Durso SC, Brandt J, Dobs A, Resnick SM. Intramuscular testosterone treatment in elderly men: evidence of memory decline and altered brain function. *J Clin Endocrinol Metab*. 2007;**92**(11):4107–4114.
56. Vaughan C, Goldstein FC, Tenover JL. Exogenous testosterone alone or with finasteride does not improve measurements of cognition in healthy older men with low serum testosterone. *J Androl*. 2007; **28**(6):875–882.
57. Emmelot-Vonk MH, Verhaar HJJ, Nakhai Pour HR, Aleman A, Lock TM, Bosch JL, Grobbee DE, van der Schouw YT. Effect of testosterone supplementation on functional mobility, cognition, and other parameters in older men: a randomized controlled trial. *JAMA*. 2008;**299**(1):39–52.
58. Borst SE, Yarrow JF, Fernandez C, Conover CF, Ye F, Meuleman JR, Morrow M, Zou B, Shuster JJ. Cognitive effects of testosterone and finasteride administration in older hypogonadal men. *Clin Interv Aging*. 2014;**9**:1327–1333.

59. Cherrier MM, Anderson K, Shofer J, Millard S, Matsumoto AM. Testosterone treatment of men with mild cognitive impairment and low testosterone levels. *Am J Alzheimers Dis Other Demen*. 2015;**30**(4): 421–430.
60. Huang G, Wharton W, Bhasin S, Harman SM, Pencina KM, Tsitouras P, Li Z, Hally KA, Asthana S, Storer TW, Basaria S. Effects of long-term testosterone administration on cognition in older men with low or low-to-normal testosterone concentrations: a prespecified secondary analysis of data from the randomised, double-blind, placebo-controlled TEAAM trial. *Lancet Diabetes Endocrinol*. 2016;**4**(8): 657–665.
61. Melehan KL, Hoyos CM, Yee BJ, Wong KK, Buchanan PR, Grunstein RR, Liu PY. Increased sexual desire with exogenous testosterone administration in men with obstructive sleep apnea: a randomized placebo-controlled study. *Andrology*. 2016;**4**(1):55–61.
62. Wahjoepramono EJ, Asih PR, Aniwiyanti V, Taddei K, Dhaliwal SS, Fuller SJ, Foster J, Carruthers M, Verdile G, Sohrabi HR, Martins RN. The effects of testosterone supplementation on cognitive functioning in older men. *CNS Neurol Disord Drug Targets*. 2016;**15**(3):337–343.
63. Resnick SM, Matsumoto AM, Stephens-Shields AJ, Ellenberg SS, Gill TM, Shumaker SA, Pleasants DD, Barrett-Connor E, Bhasin S, Cauley JA, Cella D, Crandall JP, Cunningham GR, Ensrud KE, Farrar JT, Lewis CE, Molitch ME, Pahor M, Swerdloff RS, Cifelli D, Anton S, Basaria S, Diem SJ, Wang C, Hou X, Snyder PJ. Testosterone treatment and cognitive function in older men with low testosterone and age-associated memory impairment. *JAMA*. 2017;**317**(7):717–727.
64. Bhasin S, Brito JP, Cunningham GR, Hayes FJ, Hodis HN, Matsumoto AM, Snyder PJ, Swerdloff RS, Wu FC, Yialamas MA. Testosterone therapy in men with hypogonadism: an Endocrine Society clinical practice guideline [published comment appears in *J Clin Endocrinol Metab*. 2019;201(1):13–14]. *J Clin Endocrinol Metab*. 2018;**103**(5):1715–1744.
65. Zakaib GD. Case closed: testosterone does not boost cognition. Cambridge, MA: Alzforum. 24 February 2017. Available at: www.alzforum.org/news/research-news/case-closed-testosterone-does-not-boost-cognition. Accessed 13 December 2018.
66. Rhoden EL, Morgentaler A. Risks of testosterone-replacement therapy and recommendations for monitoring. *N Engl J Med*. 2004;**350**(5):482–492.
67. Smith MA, Bowen RL, Nguyen RQ, Perry G, Atwood CS, Rimm AA. Putative gonadotropin-releasing hormone agonist therapy and dementia: an application of Medicare hospitalization claims data. *J Alzheimers Dis*. 2018;**63**(4):1269–1277.
68. Chao J, Rubinow KB, Kratz M, Amory JK, Matsumoto AM, Page ST. Short-term estrogen withdrawal increases adiposity in healthy men. *J Clin Endocrinol Metab*. 2016;**101**(10):3724–3731.
69. Finkelstein JS, Lee H, Burnett-Bowie SA, Pallais JC, Yu EW, Borges LF, Jones BF, Barry CV, Wulczyn KE, Thomas BJ, Leder BZ. Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med*. 2013;**369**(11):1011–1022.
70. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, Mierzwinski-Urban M, Clifford T, Hutton B, Rabb D. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care*. 2012;**28**(2): 138–144.
71. Hutchinson AD, Hosking JR, Kichenadasse G, Mattiske JK, Wilson C. Objective and subjective cognitive impairment following chemotherapy for cancer: a systematic review. *Cancer Treat Rev*. 2012; **38**(7):926–934.
72. Srisurapanont M, Suttajit S, Eurviriyanukul K, Varnado P. Discrepancy between objective and subjective cognition in adults with major depressive disorder. *Sci Rep*. 2017;**7**(1):3901.
73. Burmester B, Leathem J, Merrick P. Subjective cognitive complaints and objective cognitive function in aging: a systematic review and meta-analysis of recent cross-sectional findings. *Neuropsychol Rev*. 2016;**26**(4):376–393.
74. Mayor S. Testosterone increases coronary artery plaque in older men but helps anaemia and bone density. *BMJ*. 2017;**356**:j885.