The Efficacy and Adverse Events of Testosterone Replacement Therapy in Hypogonadal Men: A Systematic Review and Meta-Analysis of Randomized, Placebo-Controlled Trials

Oscar J. Ponce,[1,2,3]* Gabriela Spencer-Bonilla,[3,4]* Neri Alvarez-Villalobos,[3,5] Valentina Serrano,[3,6] Naykky Singh-Ospina,[7] Rene Rodriguez-Gutierrez,[3,5] Alejandro Salcido-Montenegro,[5] Raed Benkhadra,[1] Larry J. Prokop,[1,8] Shalender Bhasin,[9] and Juan P. Brito[3]

[1]Evidence-Based Practice Center, Mayo Clinic, Rochester, Minnesota 55905; [2]CONEVID (Unidad de Conocimiento y Evidencia), Universidad Peruana Cayetano Heredia, Lima 15102, Peru; [3]Knowledge and Evaluation Research Unit, Division of Endocrinology, Diabetes, Metabolism and Nutrition, Department of Medicine, Mayo Clinic, Rochester, Minnesota 55905; [4]University of Puerto Rico Medical Sciences Campus, San Juan 00921, Puerto Rico; [5]Universidad Autonoma de Nuevo Leon, Hospital Universitario "Dr. José E. Gonzalez," Plataforma INVEST-KER Mexico, Monterrey, Nuevo León 64460, México; [6]Department of Nutrition, Diabetes and Metabolism, Escuela de Medicina, Pontificia Universidad Catolica de Chile, Santiago 8330077, Chile; [7]Division of Endocrinology, Diabetes and Metabolism, University of Florida, Gainesville, Florida; [8]Mayo Clinic Libraries, Mayo Clinic, Rochester, Minnesota 55905; and [9]Section of Men's Health: Aging and Metabolism, Boston Claude D. Pepper Older Americans Independence Center for Function Promoting Therapies, Brigham and Women's Hospital-Harvard Medical School, Boston, Massachusetts 02115

Context: The efficacy and safety of testosterone replacement therapy (TRT) in hypogonadal men remain incompletely understood.

Objective: To conduct a systematic review and meta-analysis of randomized clinical trials (RCTs) to determine the effects of TRT on patient important outcomes and adverse events in hypogonadal men.

Data Sources: We searched Ovid Medline, Ovid Embase, Ovid Cochrane Database of Systematic Reviews, Ovid Cochrane Central Register of Controlled Trials, and Scopus, from inception to 2 March 2017.

Study Selection: Randomized clinical trials assessing the efficacy and adverse events of TRT of at least 12 weeks compared with placebo in adult men with hypogonadism, defined by morning total testosterone ≤300 ng/dL and at least one symptom or sign of hypogonadism.

Data Extraction: Reviewers working independently and in duplicate assessed the quality of RCTs and collected data on patient characteristics, interventions, and outcomes.

Results: We found four RCTs (including 1779 patients) at low risk of bias. Compared with placebo, TRT was associated with a small but significant increase in sexual desire or libido [standardized mean difference (SMD): 0.17; 95% confidence interval (CI), 0.01, 0.34; n = 1383], erectile function (SMD: 0.16; 95% CI, 0.06, 0.27; n = 1344), and sexual satisfaction (SMD: 0.16; 95% CI, 0.01, 0.31; n = 676) but had no effect on energy or mood. TRT was associated with an increased risk of developing erythrocytosis (relative risk: 8.14; 95% CI, 1.87, 35.40; n = 1579) compared with placebo but had no significant effect on lower urinary tract symptoms.

*These authors contributed equally to this study.

Abbreviations: AMS, Aging Male's Symptom; CI, confidence interval; EF, erectile function; I^2, percentage of variance in a meta-analysis that is attributable to study heterogeneity; IIEF, International Index of Erectile Function; LUTS, lower urinary tract symptoms; MD, mean difference; PDQ, Psychosexual Daily Questionnaire; PIO, patient important outcome; RR, relative risk; SMD, standardized mean difference; SR, systematic review; TRT, testosterone replacement therapy; TT, total testosterone.

Conclusion: In hypogonadal men, TRT improves sexual desire, erectile function and sexual satisfaction; however, it increases the risk of erythrocytosis. (*J Clin Endocrinol Metab* 103: 1745–1754, 2018)

The Endocrine Society recommends that testosterone replacement therapy (TRT) should be prescribed to symptomatic hypogonadal men with low levels of total testosterone (TT; ≤300 ng/dL) (1). The recent two- to fourfold rise in testosterone prescription in the last 20 years is disproportional to the rate of diagnosed testosterone deficiency (2). One study found that almost 20% of men who were prescribed testosterone did not meet the laboratory criteria (≤300 ng/dL) (3), suggesting that guidelines' recommendations may not being followed.

Despite the multiple TRT systematic reviews (SRs) available in the literature, none have summarized the evidence, including only patients for whom TRT is recommended. Most include hypogonadal men without a clear testosterone cutoff (4–9), and some include asymptomatic men with low levels of testosterone (*e.g.*, TT < 231 ng/dL) (10) and symptomatic men with normal to low levels of testosterone (*e.g.*, TT < 433 ng/dL) (11), which introduces a limitation when translating the body of evidence in patient care.

As a result of the alarming rise in testosterone prescription and uncertainty about the benefits and harms of TRT on patient important outcomes (PIOs), such as sexual function (4, 10, 12, 13) or erythrocytosis (8, 13), the Endocrine Society commissioned this SR to determine the efficacy and adverse events associated with TRT in symptomatic hypogonadal men.

Methods

This SR and meta-analysis were performed following a protocol established and approved *a priori* by an expert panel of the Endocrine Society. We followed the standards set in the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement (14).

Eligibility criteria

We included randomized, placebo-controlled trials that compared the efficacy and adverse events of TRT vs placebo in adult men with morning TT levels ≤300 ng/dL and one or more symptoms or signs of hypogonadism. The possibility of using free testosterone was discussed, but few studies recruited subjects based on free testosterone levels to reach a meaningful conclusion. The requirement for more than one TT level ≤300 ng/dL for defining hypogonadism was also considered but was deemed impractical, as very few studies required more than one TT measurement for participant eligibility. Trials in which patients received <3 months of TRT or placebo and randomized trials of testosterone formulations, such as selective androgen receptor modulators or an androgen other than testosterone, were excluded. Symptoms and signs of hypogonadism were defined by the endocrine practice guideline criteria (1) and input from the expert panel (Appendix 1). Trials

that included transgender individuals, patients with specific comorbidities (such as diabetes mellitus, human immunodeficiency virus-associated weight loss, or chronic obstructive lung disease), or men with drug-induced testosterone deficiency, such as that associated with the use of opioids or gonadotropin-releasing hormone agonists or antagonists, were also excluded.

The outcomes of interest were the following: sexual function, physical function, mobility, frailty, fatigue, mood, cognition, and anemia. Trials that used only surrogate measures or endpoints that were not deemed patient or clinically important (*e.g.*, lean body mass, inflammation markers) were excluded.

We also evaluated two adverse events: lower urinary tract symptoms (LUTS) and erythrocytosis. Major adverse cardiovascular events and prostate cancer were not included, as no trial was long enough or large enough to provide a meaningful number of events, and only one trial used a prespecified standardized ascertainment of these adverse events.

Search strategy

A comprehensive search of several databases for manuscripts published in any language was conducted from their inception to 2 March 2017 (Appendix 2). The databases included Ovid Medline In-Process & Other Non-Indexed Citations, Ovid Medline, Ovid Embase, Ovid Cochrane Central Register of Controlled Trials, Ovid Cochrane Database of Systematic Reviews, and Scopus. The search strategy was designed and conducted by an experienced librarian with input from the study investigators. Additionally, controlled vocabulary, supplemented with keywords, was used to search for SRs and meta-analyses of testosterone therapy. We consulted content experts and manually screened the references from identified SRs to identify studies missed by our search strategy.

Selection of studies

Search results were uploaded into a SR software program (DistillerSR; Evidence Partners, Ottawa, ON, Canada). Reviewers, working independently and in duplicate, screened abstracts and titles for eligibility using standardized instructions. As part of calibration, eligibility criteria were iterated for clarity and consistency. Articles included by at least one reviewer were retrieved. Following abstract screening, eligibility of reports was assessed through full-text screening. To ensure consistency among reviewers, the performance of the reviewers on 10 full-text exemplary reports was assessed. We assessed the eligibility of each study using criteria that were described earlier. At the level of full-text screening, any disagreements were resolved by consensus between the two reviewers (O.J.P. and G.S.-B.).

Data collection and management

The reviewers performed data extraction independently and in duplicate using a standardized form. Reviewers used a web-based data collection form (DistillerSR) to extract the following: (1) inclusion criteria for each trial (age and TT level measurement time and cutoff), (2) intervention characteristics (number of arms being compared, testosterone dose, route of administration, frequency, and therapy duration), (3) population

baseline characteristics (age, sex, racial demographics, and mean baseline TT levels), (4) any scale, score, or questionnaire related to efficacy outcomes at longest follow-up (sexual function, physical function, mobility, frailty, anemia, fatigue, mood, and cognition), (5) measures of effect for adverse events (erythrocytosis, LUTS), and (6) risk of bias indicators.

Outcome classification

Development and classification of constructs

Before beginning data extraction, two reviewers classified the scales, domains, and questions, reported in the included studies, into prespecified constructs of interest likely to be considered important by patients (PIO) (15). To this end, we favored sexual function constructs, such as sexual desire, erectile function, and sexual satisfaction, which are recognized by experts and can be assessed using validated instruments [e.g., International Index of Erectile Function (IIEF) (16)]. Thus, the use of frequency (e.g., frequency of ejaculations) alone, rather than any of these constructs, might not give a complete assessment of the impact of testosterone on a patient's quality of life as it relates to sexual function. Sexual activity was also included, as a result of the importance it has been given by influential entities, such as the U.S. Food and Drug Administration. The constructs used for classification were developed based on qualitative studies (15, 17) and a prior SR (4) that evaluated these PIOs in patients with similar characteristics to our population of interest (symptomatic patients with TT levels ≤300 ng/dL). As a result of the process, we identify the following constructs for evaluation: (1) erectile function (EF), (2) sexual desire or libido, (3) sexual satisfaction, (4) energy, (5) mood, (6) physical function, and (7) cognition.

When trials reported multiple scales measuring the same domain (e.g., EF), we included that which most reflected the underlying domain, as judged by the extractors. Additionally, overall scores of scales were prioritized over single domains or individual questions from a larger scale. Furthermore, scales measuring "fatigue" were considered to be measuring the same construct as the "energy" scales; e.g., we considered "less fatigue" to be equivalent to "more energy."

Author contact

Authors were contacted when it was unclear whether full-text manuscripts were eligible for inclusion in this SR and if data necessary to perform a meta-analysis for the outcomes of interest were missing. Authors were contacted by an E-mail to the corresponding author. If no response was received after 1 week, then the author was contacted by phone. We contacted authors twice by phone, at 1-week intervals. If no response was received, then the first or last author was contacted by using E-mail and phone with the same interval period.

Risk of bias and confidence in the body of evidence

The risk of bias was assessed using the Cochrane Collaboration's tool for randomized clinical trials (18). This tool takes into consideration six domains: (1) random sequence generation; (2) allocation concealment; (3) blinding of participants; (4) incomplete outcome data; (5) selective outcome reporting; and (6) other sources of bias. Disagreements were resolved by consensus between the two reviewers (O.J.P. and G.S.-B.). The overall confidence or overall quality of evidence for each outcome was appraised by discussion between the two extractors using the Grading of Recommendations Assessment,

Development and Evaluation approach. This approach takes into account the risk of bias of the individual studies, inconsistency in the results, indirectness, imprecision, and other considerations to provide a global assessment of the confidence merited by the body of evidence (19).

Summary measures and synthesis of results

We calculated standardized mean difference (SMD) or relative risk (RR) for each outcome of interest and pooled results using random-effect models. All statistical analyses were performed using Stata v15.0 (StataCorp, College Station, TX). In trials with more than one eligible active treatment arm, we chose the treatment regimen that most resembled the majority of active treatment arms of the other included studies for pooling in the meta-analysis. The other eligible active treatment arms were used for sensitivity analyses.

For continuous outcomes reported in different trials with different scales but reflecting the same construct (e.g., mood, sexual desire), we calculated their SMD. We also pooled scales, domains, or results that were reported in more than one trial (e.g., orgasmic function domain from the IIEF scale). In general, SMDs of 0.2, 0.5, and 0.8 are considered small, medium, and large treatment effects, respectively (20). Inconsistency for each outcome, not attributable to chance, was assessed visually using forest plots and estimated using the percentage of variance in a meta-analysis that is attributable to study heterogeneity (I^2) statistic. $I^2 < 25\%$ reflects low inconsistency; $I^2 > 75\%$ reflects high inconsistency (21). Subgroup analyses were planned based on testosterone formulation, route of administration, risk of bias, elderly population (≥65 years old), and TT cutoff of ≤280 ng/dL.

Results

Study selection

A total of 2807 studies were identified through our search strategy. Full-text screening with moderate agreement between the reviewers (κ = 0.54) identified 11 (22–32) publications, reporting on four trials and 1779 patients (Fig. 1) that met the prespecified inclusion and exclusion criteria.

Trial characteristics

All of the trials included adult men with at least one symptom of hypogonadism and with at least one morning TT level ≤300 ng/dL. However, only three trials (22–30) reported that they performed two morning TT measurements; this was confirmed by contacting the authors. Furthermore, one trial included only older men (>65 years old) and required an average of two morning TT levels ≤275 ng/dL (22, 24, 27–29). The included trials used either a transdermal 1% testosterone gel or a 2% testosterone solution. The intervention duration in the included trials ranged from 12 to 52 weeks. The mean baseline TT levels ranged from 201.2 to 239 ng/dL (Table 1). Two trials reported that participants who were assigned to TRT achieved mean on-treatment testosterone levels in the normal male range (>300 ng/dL) (26, 31, 32). One trial reported that 73% of patients who received TRT achieved normal range levels (300 to 1050 ng/dL) at week 12 (23, 25, 30). The remaining study (22, 24,
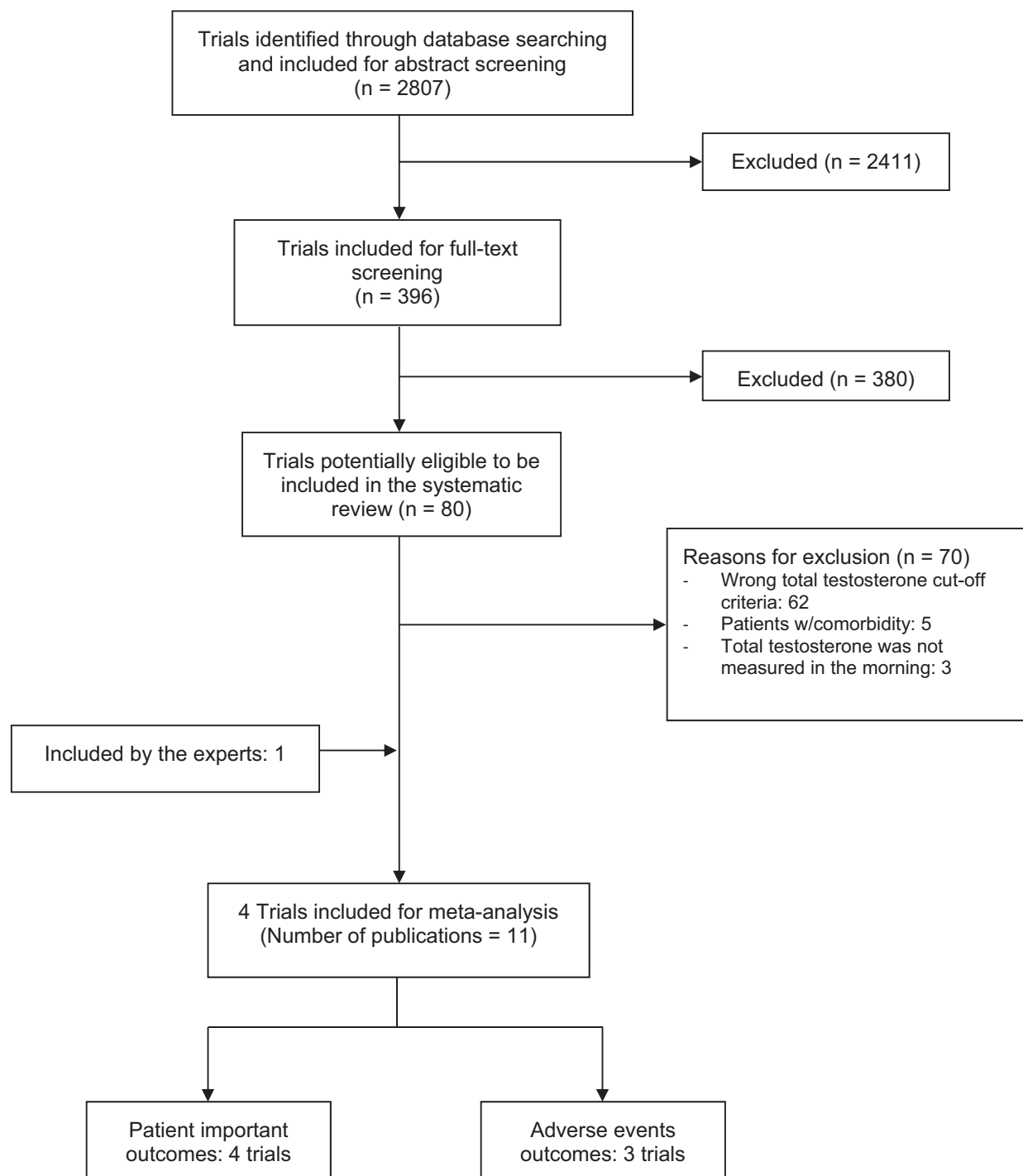
```
┌─────────────────────────────────────┐
│  Trials identified through database  │
│  searching and included for abstract │
│           screening                  │
│            (n = 2807)                │
└─────────────────────────────────────┘
                  │
                  │──────────────────────►  ┌──────────────────────────┐
                  │                          │   Excluded (n = 2411)    │
                  ▼                          └──────────────────────────┘
┌─────────────────────────────────────┐
│   Trials included for full-text      │
│            screening                 │
│             (n = 396)                │
└─────────────────────────────────────┘
                  │
                  │──────────────────────►  ┌──────────────────────────┐
                  │                          │   Excluded (n = 380)     │
                  ▼                          └──────────────────────────┘
┌─────────────────────────────────────┐
│  Trials potentially eligible to be   │
│   included in the systematic         │
│         review (n = 80)              │
└─────────────────────────────────────┘
```

Reasons for exclusion (n = 70)
- Wrong total testosterone cut-off criteria: 62
- Patients w/comorbidity: 5
- Total testosterone was not measured in the morning: 3

Included by the experts: 1

4 Trials included for meta-analysis
(Number of publications = 11)

Patient important outcomes: 4 trials

Adverse events outcomes: 3 trials

Figure 1. Flowchart.

27–29) showed a statistically significant increase in serum testosterone in the TRT group compared with placebo. The overall risk of bias was judged to be low for all outcomes (Table 2).

Meta-analysis of PIOs by construct

Symptomatic hypogonadal men treated with TRT had a small but statistically significantly improvement in sexual desire [SMD: 0.16; 95% confidence interval (CI),

0.01, 0.31], EF (SMD: 0.16; 95% CI, 0.06, 0.27), and sexual satisfaction (SMD: 0.17; 95% CI, 0.01, 0.34) compared with patients receiving placebo. There was a nonstatistically significant improvement in energy and mood in the testosterone group compared with the placebo group (Fig. 2).

Two sensitivity analyses were performed. The first one included a third arm of the Steidle *et al.* (32) trial—intervention consisted of a higher daily dose (100 mg)

Table 1.   Baseline Characteristics

| Author, Year | Inclusion Criteria | | Intervention | | | Age, Years (SD) | n | Baseline Morning TT, ng/dL (SD) |
	Age	TT Cutoff, ng/dL	Drug	Route of Administration	Weeks			
Brock, 2016 (23, 25, 30)	≥18	<300	Testosterone solution 2%, 60 mg daily	Transdermal	12	54.7 (10.6)	358	202.2 (66.3)
			Placebo			55.9 (11.4)	357	201.2 (67.3)
Snyder, 2016 (22, 24, 27–29)	≥65	<275	Testosterone gel 1%, 50 mg daily	Transdermal	52	72.1 (5.7)	395	232 (63)
			Placebo			72.3 (5.8)	395	236 (67)
Paduch, 2015 (26)	≥26	<300	Testosterone solution 2%, 60 mg daily	Transdermal	16	48.4 (9.8)	36	214 (56)
			Placebo			52.7 (9.3)	40	223 (53)
Steidle, 2003 (31, 32)	20–80	≤300	Testosterone gel 1%, 50 mg daily	Transdermal	12.9	58.1 (9.7)	99	234 (58)
			Placebo			56.8 (10.8)	99	228 (81)
			Testosterone gel 1%, 100 mg daily[a]			56.8 (10.6)	106	234 (63)
			Testosterone patch, 5 mg daily[b]			60.5 (9.7)	102	239 (69)

Abbreviation: SD, standard deviation.

[a]Sensitivity analysis was performed with this arm.

[b]Not included in meta-analysis.

of 1% transdermal testosterone gel (31, 32). Findings in EF (SMD: 0.23; 95% CI, 0.12, 0.33) and sexual desire (SMD: 0.25; 95% CI, 0.06, 0.44) remained consistent. For the second analyses, we excluded a non-PIO scale from the EF construct. This result was derived from excluding the orgasmic function domain of the IIEF scale used by Paduch *et al.* (26), which is solely based on two frequency questions about ejaculation and orgasm. Effectiveness of testosterone in EF remained unchanged (SMD: 0.17; 95% CI, 0.06, 0.29).

Meta-analysis of PIOs by scales

Testosterone therapy had a small but statistically significant improvement in the EF domain of IIEF (SMD: 0.20; CI 95%, 0.08, 0.32) compared with placebo.

Two additional outcomes were available. The first one was a non-PIO, called orgasmic function, which was measured with the IIEF and is derived from two questions about ejaculation and orgasm frequency (16). The second, sexual activity, a non-PIO but important for policymakers, such as the US Food and Drug Administration, was measured with the Psychosexual Daily Questionnaire (PDQ), which includes these two questions, as well as items inquiring about sexual daydreams, anticipation of sex, flirting, sexual interactions with partner, erection, masturbation, and intercourse (33). The analysis on both outcomes found that testosterone treatment had a small, statistically significant improvement in overall sexual activity (SMD: 0.23; CI 95%, 0.13, 0.33; Fig. 2) but had no significant effect on orgasmic function (SMD: 0.11; CI 95%, −0.04, 0.26).

Narrative synthesis

One trial (22, 24, 27–29) reported that 1% transdermal testosterone gel significantly improved self-reported physical function compared with placebo [mean difference (MD): 3.06; 95% CI, 1.18, 4.94],

Table 2.   Risk of Bias of Included Studies

| Study | Random Sequence Generation | Allocation Concealment | Blinding of Participants and Personnel | Blinding of Outcome Assessment | | Incomplete Outcome Data | | Selective Reporting | Loss to Follow-Up, % |
				PRO	AEO	PRO	AEO		
Brock *et al.*, 2016	Low risk	Low risk	Low risk	Unclear	Low risk	High risk	Low risk	Low risk	16.1
Snyder *et al.*, 2016	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	4.8
Paduch *et al.*, 2015	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	13
Patient Reported Outcomes									
Steidle *et al.*, 2003	Unclear	Unclear	Low risk	Low risk		Unclear		High risk	28.6

Abbreviations: AEO, adverse events outcomes; PRO, patient reported outcomes.
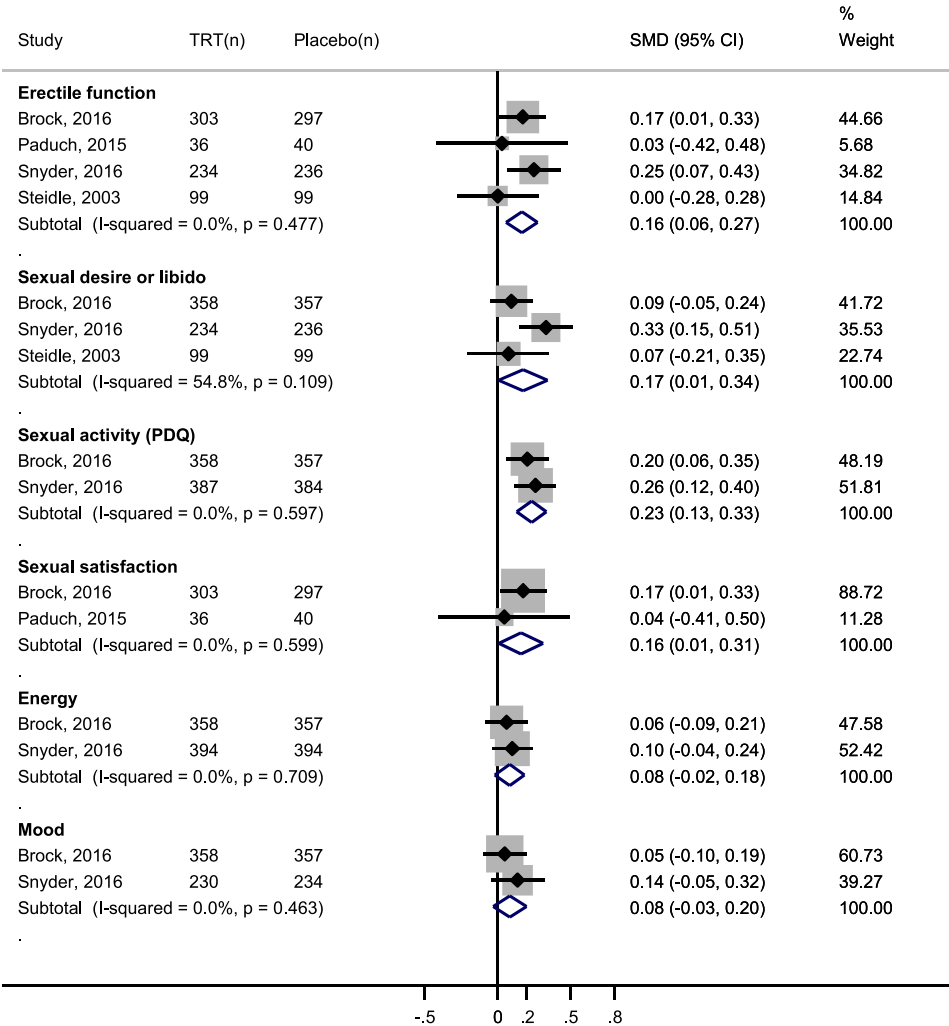
Figure 2. Meta-analysis of sexual function by constructs, sexual activity, energy, and mood. The testosterone gel 1% 50 mg daily arm in Steidle, 2003, was compared with placebo for this analysis. (n), number of patients; PDQ, Psychosexual Daily Questionnaire.

measured by the physical function domain from the quality of life questionnaire, Medical Outcomes Study Short Form 36. However, it did not significantly improve memory, as measured by the Memory Complaint Questionnaire score compared with placebo (MD: −0.24; 95% CI, −0.72, 0.23). Memory Complaint Questionnaire is a self-reported instrument that asks patients to assess their current memory relative to the past or their memory performance over time (34).

Meta-analysis of adverse events

The participants assigned to testosterone arms of the included trials had a statistically, significantly higher frequency of erythrocytosis (hematocrit >54% or hemoglobin >17.5 g/dL) compared with placebo (RR: 8.14; CI 95%, 1.87, 35.40; Fig. 3). There was no statistically significant difference in LUTS scores between the testosterone and placebo arms of two trials (MD: 0.38; 95% CI, −0.67, 1.43; Fig. 4).

Grading of Recommendations Assessment, Development and Evaluation approach to assess the quality of evidence

Quality of evidence was high for EF, sexual satisfaction, sexual desire, physical function, and erythrocytosis, as a result of low risk of bias, low heterogeneity, evidence included that answers the review question (indirectness may not be an issue), and precise results (a considerable number of events and consistent message about benefit or harm) (Table 3). Although results for sexual desire were somewhat heterogeneous (I^2 = 54.8%), there was not a substantial variation in the effect estimates across studies, and there was a clear overlap of CIs among trials. Energy, mood, subjective memory complaint, LUTS, and orgasmic function were moderate as a result of nonsignificant treatment effects. Estimates of effects on sexual activity were also judged to be moderate quality; although it is an important outcome for many policymakers and experts, it is not a PIO (indirectness). Furthermore,

Table 3. Summary of Findings and Confidence in the Body of Evidence

	Effect Size (95% CI)	No. of Participants (Total Studies)	Quality of Evidence (Domains of Concern)
Efficacy outcomes by constructs			
EF	SMD: 0.16 (0.06, 0.27)	1344 (4)	High
Sexual satisfaction	SMD: 0.16 (0.01, 0.31)	676 (2)	High
Sexual desire	SMD: 0.17 (0.01, 0.34)	1383 (3)	High
Energy	SMD: 0.08 (−0.02, 0.18)	1503 (2)	Moderate (imprecision)
Mood	SMD: 0.08 (−0.03, 0.20)	1179 (2)	Moderate (imprecision)
Physical function	MD: 3.06 (1.18, 4.94)	790 (1)	High
Subjective memory complaint	MD: −0.24 (−0.72, 0.23)	790 (1)	Moderate (imprecision)
Efficacy outcomes by scales			
IIEF EF	SMD: 0.20 (0.08, 0.32)	1179 (2)	High
IIEF orgasmic function	SMD: 0.11 (−0.04, 0.26)	676 (2)	Low (indirectness, not a PIO and imprecision)
PDQ sexual activity	SMD: 0.23 (0.13, 0.33)	1486 (2)	Moderate (indirectness, not a PIO)
Adverse events outcomes			
LUTS	MD: 0.38 (−0.67, 1.43)	866 (2)	Moderate (imprecision)
Erythrocytosis	RR: 8.14 (1.87, 35.40)	1579 (3)	High

orgasmic function was downgraded, as a result of nonsignificant treatment effect and not being a PIO, resulting in low quality of evidence.

Discussion

High-quality evidence shows that compared with placebo, testosterone therapy was associated with a small but substantial improvement in sexual desire (libido), EF, sexual activity, and sexual satisfaction in hypogonadal men. However, TRT was also associated with a substantially increased risk of developing erythrocytosis. No statistically significant effect was found on energy, mood, or LUTS.

Comparison with previous findings

To generate an accurate comparison, an additional systematic search was performed by using databases mentioned before. The search was tailored to find SRs and meta-analysis in English with similar inclusion and exclusion criteria to our study. We did not find any SR with these characteristics, making this SR one that truly included hypogonadal men for whom TRT is recommended.

Although no SR retrieved by our search met our inclusion criteria, three important SRs bore some resemblance. The first one included hypogonadal or healthy men with sexual dysfunction and performed a subgroup analysis on patients with low mean levels of TT at baseline
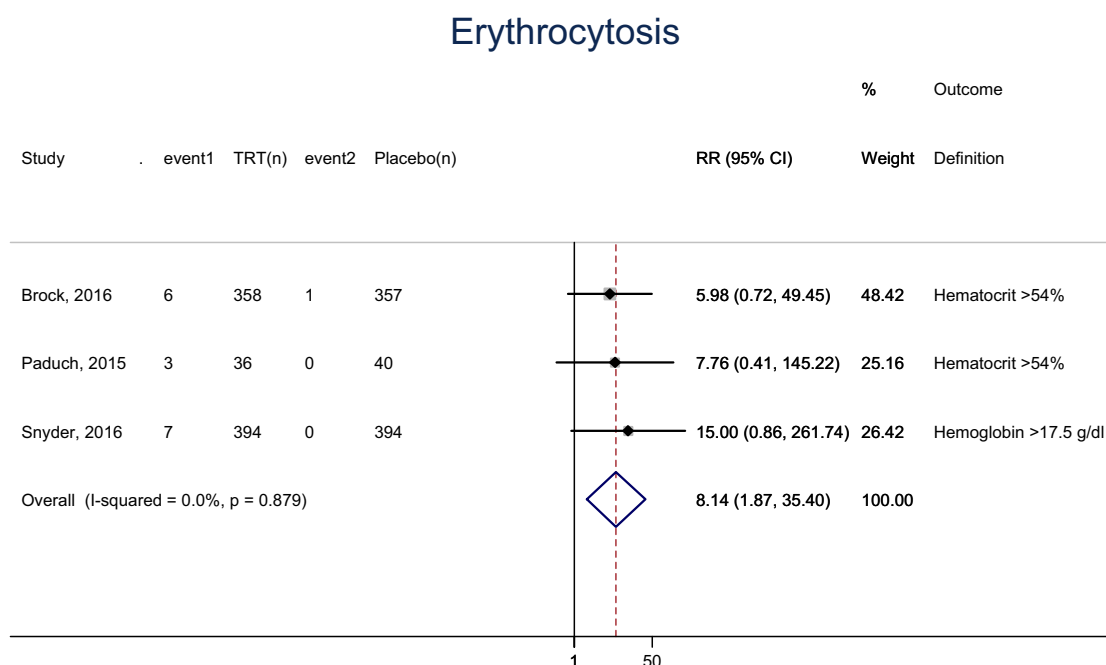

Figure 3. Meta-analysis of erythrocytosis. event1 and event2, the number of events in each arm; (n), number of patients.
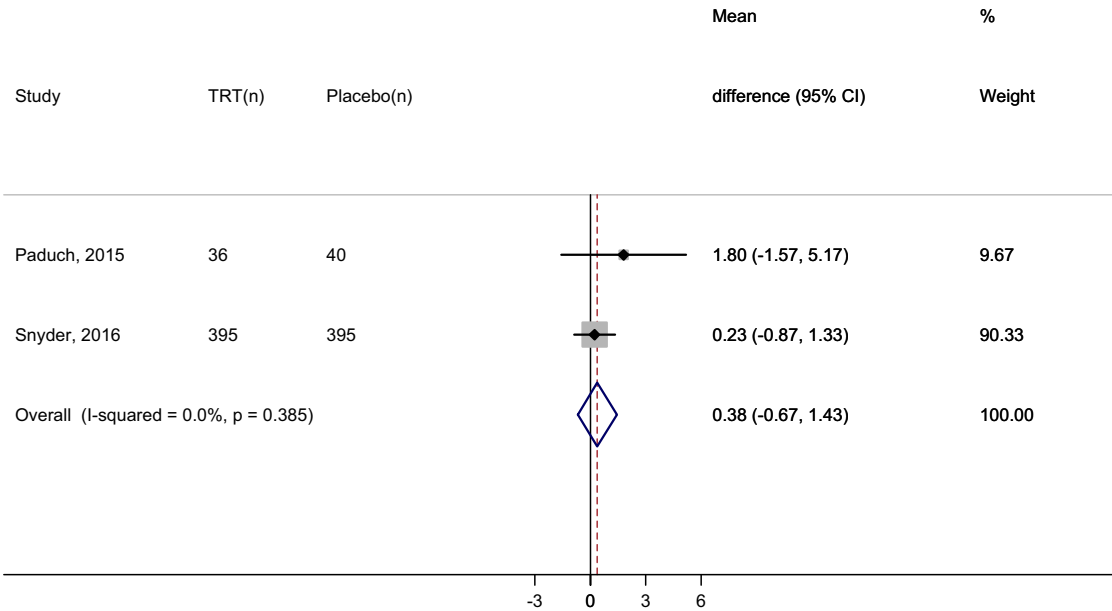
International Prostate Symptom Score

Figure 4. Meta-analysis of the International Prostate Symptom Score. (n), number of patients.

and found that similar to our results, testosterone treatment improved sexual desire (SMD: 1.24; 95% CI, 0.12, 2.36; n=519); nonetheless, it did not improve EF (SMD: 0.80; 95% CI, −0.10, 1.60; n=511) or sexual satisfaction (SMD: 1.20; 95% CI, −0.50, 2.90; n=110) (4); however, there was inconsistency in the estimates of treatment effect. Hence, a comparison with our study may be misleading because of their subgroup analysis definition and high inconsistency in effects of included trials.

Two additional SRs included male populations with late-onset hypogonadism (9, 11), which is defined by the presence of hypogonadal symptoms and decreased levels of testosterone (35). One of them, however, had a TT cutoff for inclusion of <433 ng/dL; thus, the included trials likely recruited eugonadal, as well as hypogonadal, men, as defined by current Endocrine Society recommendations (11). The outcomes of interest were quality of life and the Aging Male's Symptom (AMS) scale. They found that AMS psychological (SMD: −0.89; 95% CI, −1.41, −0.37; n=1212) and sexual (SMD: −1.29; 95% CI, −1.75, −0.83; n= 1212) subscales improved with TRT compared with placebo. These trials did not use modern psychometrically robust instruments for the ascertainment of sexual function, and the AMS psychological subscale also differs from the mood and energy scales used in the trials included in our SR. The last SR included participants without regard to a specific eligibility threshold for testosterone level, therefore, included participants who were not hypogonadal. Similar to our findings, this SR also did not find a change in LUTS after

testosterone treatment compared with placebo ($P > 0.05$, n=2029) (27).

Thirteen additional SRs and meta-analysis with some characteristics similar to our study were found. The risk of the development of erythrocytosis was also increased in one SR of men ≥45 years old with low or low-normal testosterone levels (odds ratio: 3.67; 95% CI, 1.82, 7.51; n = 1084) (8). Moreover, EF, measured by the IIEF questionnaire was improved across all studies that included male participants with low levels of testosterone (10, 12). These SRs used variable definitions of hypogonadism and often included men who were not hypogonadal; some included patients with other comorbidities, and some compared TRT with other treatments.

Implications for practice

These SR and meta-analysis fill an evidence gap and summarize the effects of TRT on patients who are considered to be hypogonadal based on current clinical recommendations. Patients, who place high value on improving sexual desire and EF and consider TRT to have a low value on burden of treatment in terms of cost, need for follow up, and risk adverse effects, such as erythrocytosis, may be good candidates for therapy. Patients may wish to know that TRT has not been shown to improve energy, mood, or cognition in symptomatic hypogonadal men and be offered other, evidence-based therapies for these symptoms. Clinicians and patients might elect not to initiate therapy if harm of therapy for that particular patient outweighs the benefits.

Individualization of care should be arrived at through shared decision-making. A more personalized treatment approach, based on patient values and preferences, will likely bring more patient-centered testosterone prescription patterns with the potential to decrease the overprescribing of hormone replacement therapy to millions of Americans (36, 37). To inform better these shared decision-making conversations, future testosterone studies should be monitored over longer periods of time to evaluate other potential benefits and harm of testosterone therapy. For instance, it is still unclear what the effect of TRT is on patient-important cardiovascular outcomes, fractures, and possible sequelae of long-term erythrocytosis (38). To this end, larger clinical trials or more rigorous observational studies are needed.

Strengths and limitations

These SR and meta-analyses are unique for their inclusion of trials that included participants who met the criteria for the diagnosis of hypogonadism, based on the Endocrine Society's clinical guideline (4), *i.e.*, trials that recruited hypogonadal men based both on low TT levels, defined as TT level ≤300 ng/dL, and the presence of one or more symptoms or signs of hypogonadism. The reported outcomes included in this meta-analysis were those that the Endocrine Society's expert panel and patients (15) deemed clinically relevant and patient important and that were ascertained using validated instruments. Almost all of the outcomes have moderate to high quality of evidence.

Limitations of this report include the heterogeneity of instruments used to ascertain outcomes across trials. The patient population recruited in these trials was also heterogeneous and likely included a mix of men with organic hypogonadism as a result of known diseases of the testes, pituitary, and the hypothalamus; age-related decline; and possibly low testosterone levels as a result of other conditions. We did not obtain data on how the symptoms improved based on the on-treatment testosterone levels, which would have required an individual patient data meta-analysis. There have been efforts to develop and standardize patient-reported outcomes to ascertain the efficacy of treatment in hypogonadal men (5); however, their uptake in research and clinical settings is still low. None of the trials was long enough or large enough to have sufficient statistical power to determine the effects of TRT on the incidence of prostate cancer, major adverse cardiovascular events, or bone fractures. The data on performance-based measures of physical function were available only in one trial, precluding a meta-analysis.

Conclusion

TRT improves sexual desire, EF, and sexual satisfaction but not mood, energy, or cognition in hypogonadal men. Testosterone treatment increases the risk of erythrocytosis in hypogonadal men but does not affect LUTS. Practicing clinicians may wish to incorporate these findings when discussing the benefits and harms of TRT with their patients.

References

1. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, Montori VM; Task Force, Endocrine Society. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2010;**95**(6):2536–2559.
2. Bandari J, Ayyash OM, Emery SL, Wessel CB, Davies BJ. Marketing and testosterone treatment in the USA: a systematic review. *Eur Urol Focus.* 2017;**3**(4–5):395–402.
3. Baillargeon J, Urban RJ, Kuo YF, Holmes HM, Raji MA, Morgentaler A, Howrey BT, Lin YL, Ottenbacher KJ. Screening and monitoring in men prescribed testosterone therapy in the U.S., 2001–2010. *Public Health Rep.* 2015;**130**(2):143–152.
4. Boloña ER, Uraga MV, Haddad RM, Tracz MJ, Sideras K, Kennedy CC, Caples SM, Erwin PJ, Montori VM. Testosterone use in men with sexual dysfunction: a systematic review and meta-analysis of randomized placebo-controlled trials. *Mayo Clin Proc.* 2007;**82**(1):20–28.
5. Guo C, Gu W, Liu M, Peng BO, Yao X, Yang B, Zheng J. Efficacy and safety of testosterone replacement therapy in men with hypogonadism: a meta-analysis study of placebo-controlled trials. *Exp Ther Med.* 2016;**11**(3):853–863.
6. Fernández-Balsells MM, Murad MH, Lane M, Lampropulos JF, Albuquerque F, Mullan RJ, Agrwal N, Elamin MB, Gallegos-Orozco JF, Wang AT, Erwin PJ, Bhasin S, Montori VM. Clinical review 1: adverse effects of testosterone therapy in adult men: a systematic review and meta-analysis. *J Clin Endocrinol Metab.* 2010;**95**(6):2560–2575.
7. 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.
8. Calof OM, Singh AB, Lee ML, Kenny AM, Urban RJ, Tenover JL, Bhasin S. Adverse events associated with testosterone replacement in middle-aged and older men: a meta-analysis of randomized, placebo-controlled trials. *J Gerontol A Biol Sci Med Sci.* 2005;**60**(11):1451–1457.

9. Kohn TP, Mata DA, Ramasamy R, Lipshultz LI. Effects of testosterone replacement therapy on lower urinary tract symptoms: a systematic review and meta-analysis. *Eur Urol.* 2016;**69**(6):1083–1090.

10. Corona G, Isidori AM, Buvat J, Aversa A, Rastrelli G, Hackett G, Rochira V, Sforza A, Lenzi A, Mannucci E, Maggi M. Testosterone supplementation and sexual function: a meta-analysis study. *J Sex Med.* 2014;**11**(6):1577–1592.

11. Nian Y, Ding M, Hu S, He H, Cheng S, Yi L, Li Y, Wang Y. Testosterone replacement therapy improves health-related quality of life for patients with late-onset hypogonadism: a meta-analysis of randomized controlled trials. *Andrologia.* 2017;**49**(4):e12630.

12. Corona G, Rastrelli G, Morgentaler A, Sforza A, Mannucci E, Maggi M. Meta-analysis of results of testosterone therapy on sexual function based on International Index of Erectile Function scores. *Eur Urol.* 2017;**72**(6):1000–1011.

13. 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.

14. 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.

15. Gelhorn HL, Vernon MK, Stewart KD, Miller MG, Brod M, Althof SE, DeRogatis LR, Dobs A, Seftel AD, Revicki DA. Content validity of the Hypogonadism Impact of Symptoms Questionnaire (HIS-Q): a patient-reported outcome measure to evaluate symptoms of hypogonadism. *Patient.* 2016;**9**(2):181–190.

16. Rosen RC, Riley A, Wagner G, Osterloh IH, Kirkpatrick J, Mishra A. The International Index of Erectile Function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology.* 1997;**49**(6):822–830.

17. Hayes RP, Ni X, Heiselman DE, Kinchen K. Psychometric testing of two new patient-reported outcome instruments for the evaluation of treatment for hypogonadism. *Int J Clin Pract.* 2016;**70**(7):587–595.

18. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Baltimore, MD: The Cochrane Collaboration. 2011. http://handbook.cochrane.org. Accessed 15 February 2018.

19. Ryan R, Hill S. How to GRADE the Quality of the Evidence. London, UK: Cochrane Consumers and Communication Group. Dcember 2016. http://cccrg.cochrane.org/author-resources. Accessed 15 February 2018.

20. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Erlbaum; 1988.

21. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;**327**(7414):557–560.

22. Roy CN, Snyder PJ, Stephens-Shields AJ, Artz AS, Bhasin S, Cohen HJ, Farrar JT, Gill TM, Zeldow B, Cella D, Barrett-Connor E, Cauley JA, Crandall JP, Cunningham GR, Ensrud KE, Lewis CE, Matsumoto AM, Molitch ME, Pahor M, Swerdloff RS, Cifelli D, Hou X, Resnick SM, Walston JD, Anton S, Basaria S, Diem SJ, Wang C, Schrier SL, Ellenberg SS. Association of testosterone levels with anemia in older men: a controlled clinical trial. *JAMA Intern Med.* 2017;**177**(4):480–490.

23. Brock G, Heiselman D, Maggi M, Kim SW, Rodríguez Vallejo JM, Behre HM, McGettigan J, Dowsett SA, Hayes RP, Knorr J, Ni X, Kinchen K. Effect of testosterone solution 2% on testosterone concentration, sex drive and energy in hypogonadal men: results of a placebo controlled study. *J Urol.* 2016;**195**(3):699–705.

24. Cunningham GR, Stephens-Shields AJ, Rosen RC, Wang C, Bhasin S, Matsumoto AM, Parsons JK, Gill TM, Molitch ME, Farrar JT, Cella D, Barrett-Connor E, Cauley JA, Cifelli D, Crandall JP, Ensrud KE, Gallagher L, Zeldow B, Lewis CE, Pahor M, Swerdloff RS, Hou X, Anton S, Basaria S, Diem SJ, Tabatabaie V, Ellenberg SS, Snyder PJ. Testosterone treatment and sexual function in older men with low testosterone levels. *J Clin Endocrinol Metab.* 2016;**101**(8):3096–3104.

25. Maggi M, Heiselman D, Knorr J, Iyengar S, Paduch DA, Donatucci CF. Impact of testosterone solution 2% on ejaculatory dysfunction in hypogonadal men. *J Sex Med.* 2016;**13**(8):1220–1226.

26. Paduch DA, Polzer PK, Ni X, Basaria S. Testosterone replacement in androgen-deficient men with ejaculatory dysfunction: a randomized controlled trial. *J Clin Endocrinol Metab.* 2015;**100**(8):2956–2962.

27. 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.

28. Snyder PJ, Bhasin S, Cunningham GR, Matsumoto AM, Stephens-Shields AJ, Cauley JA, Gill TM, Barrett-Connor E, Swerdloff RS, Wang C, Ensrud KE, Lewis CE, Farrar JT, Cella D, Rosen RC, Pahor M, Crandall JP, Molitch ME, Cifelli D, Dougar D, Fluharty L, Resnick SM, Storer TW, Anton S, Basaria S, Diem SJ, Hou X, Mohler ER III, Parsons JK, Wenger NK, Zeldow B, Landis JR, Ellenberg SS; Testosterone Trials Investigators. Effects of testosterone treatment in older men. *N Engl J Med.* 2016;**374**(7):611–624.

29. Snyder PJ, Kopperdahl DL, Stephens-Shields AJ, Ellenberg SS, Cauley JA, Ensrud KE, Lewis CE, Barrett-Connor E, Schwartz AV, Lee DC, Bhasin S, Cunningham GR, Gill TM, Matsumoto AM, Swerdloff RS, Basaria S, Diem SJ, Wang C, Hou X, Cifelli D, Dougar D, Zeldow B, Bauer DC, Keaveny TM. Effect of testosterone treatment on volumetric bone density and strength in older men with low testosterone: a controlled clinical trial. *JAMA Intern Med.* 2017;**177**(4):471–479.

30. Wu F, Zitzmann M, Heiselman D, Donatucci C, Knorr J, Patel AB, Kinchen K. Demographic and clinical correlates of patient-reported improvement in sex drive, erectile function, and energy with testosterone solution 2. *J Sex Med.* 2016;**13**(8):1212–1219.

31. Seftel AD, Mack RJ, Secrest AR, Smith TM. Restorative increases in serum testosterone levels are significantly correlated to improvements in sexual functioning. *J Androl.* 2004;**25**(6):963–972.

32. Steidle C, Schwartz S, Jacoby K, Sebree T, Smith T, Bachand R; North American AA2500 T Gel Study Group. AA2500 testosterone gel normalizes androgen levels in aging males with improvements in body composition and sexual function. *J Clin Endocrinol Metab.* 2003;**88**(6):2673–2681.

33. Lee KK, Berman N, Alexander GM, Hull L, Swerdloff RS, Wang C. A simple self-report diary for assessing psychosexual function in hypogonadal men. *J Androl.* 2003;**24**(5):688–698.

34. Crook TH III, Feher EP, Larrabee GJ. Assessment of memory complaint in age-associated memory impairment: the MAC-Q. *Int Psychogeriatr.* 1992;**4**(2):165–176.

35. Schubert M, Jockenhövel F. Late-onset hypogonadism in the aging male (LOH): definition, diagnostic and clinical aspects. *J Endocrinol Invest.* 2005;**28**(3, Suppl):23–27.

36. Jasuja GK, Bhasin S, Reisman JI, Berlowitz DR, Rose AJ. Ascertainment of testosterone prescribing practices in the VA. *Med Care.* 2015;**53**(9):746–752.

37. Morgan DJ, Dhruva SS, Wright SM, Korenstein D. 2016 update on medical overuse: a systematic review. *JAMA Intern Med.* 2016;**176**(11):1687–1692.

38. Morgentaler A, Miner MM, Caliber M, Guay AT, Khera M, Traish AM. Testosterone therapy and cardiovascular risk: advances and controversies. *Mayo Clin Proc.* 2015;**90**(2):224–251.