Published in final edited form as:
Menopause. 2014 September ; 21(9): 952–961. doi:10.1097/GME.0000000000000195.

Effects of Physiologic Testosterone Replacement on Quality of Life, Self-Esteem, and Mood in Women with Primary Ovarian Insufficiency

Gioia M. Guerrieri, D.O.[1], **Pedro E. Martinez, M.D.**[1], **Summer P. Klug, M.S.**[1], **Nazli A. Haq, M.A.**[1], **Vien H. Vanderhoof, R.N., C.R.N.P.**[2], **Deloris E. Koziol, Ph.D.**[3], **Vaishali B. Popat, M.D., M.P.H.**[2], **Sophia N. Kalantaridou, M.D.**[2], **Karim A. Calis, Pharm.D., M.P.H.**[2], **David R. Rubinow, M.D.**[4], **Peter J. Schmidt, M.D.**[1,*], and **Lawrence M. Nelson, M.D.**[2,*]

[1]Behavioral Endocrinology Branch, National Institute of Mental Health, National Institutes of Health, Department of Health & Human Services

[2]Integrative Reproductive Medicine Group, Intramural Research Program on Reproductive and Adult Endocrinology, National Institute of Child Health and Human Development, National Institutes of Health, Department of Health & Human Services

[3]Biostatistics and Clinical Epidemiology Service, Office of the Deputy Director of Clinical Care, Clinical Center, National Institutes of Health, Department of Health & Human Services

[4]Department of Psychiatry, University of North Carolina, Chapel Hill NC

Abstract

Objective—Low androgen levels occur in women with primary ovarian insufficiency(POI) and could contribute to mood and behavioral symptoms in this condition. We examined the effects of physiologic testosterone (T) replacement therapy added to standard estrogen/progestin replacement therapy (EPT) on quality of life, self-esteem, and mood in women with POI.

Methods—128 women with 46XX spontaneous POI participated in a 12 month randomized, placebo-controlled, parallel-design investigation of the efficacy of T augmentation of EPT. Quality of life, self-esteem, and mood symptoms were evaluated with standardized rating scales and a structured clinical interview. Differences in outcome measures between the T and placebo treatments were analyzed by Wilcoxon rank-sum tests.

Results—No differences were found in baseline characteristics including serum hormone levels(P >0.05). Baseline mean(SD) CES-D scores were 10.7(8.6)(T) and 9.2(7.8)(placebo) (P = 0.35). After 12 months of treatment, measures of quality of life, self-esteem, and the presence of mood symptoms did not differ between treatment groups. Serum T levels achieved physiologic

Corresponding author: Gioia M. Guerrieri, 10 Center Drive, Room 25330, MSC 1274, Bethesda MD 20892-1277, gioia.guerrieri@nih.gov, Tel: 301.451.4045, Fax: 301.480.4597.
*Both authors contributed equally to this work.

Conflicts of Interest/Financial Disclosure: None of the other Authors have any disclosures or conflicts of interest to report.

Preliminary results of this study were presented at the 2013 Society of Biological Psychiatry annual meeting and have not been presented or published elsewhere.

levels in the T group and were significantly higher compared to placebo ($P < 0.001$). Baseline T levels were not associated with either adverse or beneficial clinical effects.

Conclusions—The 150 microgram T patch achieves physiologic hormone levels in women with POI. Our findings suggest that augmentation of standard EPT with physiologic testosterone replacement in young women with POI neither aggravates nor improves baseline reports of quality of life, or self-esteem and had minimal effect on mood. Other mechanisms might play a role in the altered mood that accompanies this disorder.

Keywords

testosterone; POI; mood; QOL; self-esteem

Introduction

Primary ovarian insufficiency (POI) affects approximately 1% of women before age 40, and is accompanied by infertility, decreased quality of life, depression, increased risk for osteoporosis and cardiovascular disease (1–5). In addition to decreased ovarian estradiol and progesterone production, women with POI have lower circulating levels of androgens compared with women who have normal ovarian function (6–9). Total production of testosterone (T) in women is estimated to be approximately 300 mcg per day with ovarian production of T accounting for approximately 50% of this total daily secretion and, in women with POI, there is considerable evidence suggesting that ovarian T production is insufficient. Current hormone replacement strategies were developed for women who experience natural menopause at an average age of 50. However, despite the large number of studies on hormone therapy in normally menopausal women, less is known about the effects of menopausal hormone therapy on ovarian failure in young women. Women with POI will take estrogen replacement for decades and the potential long-term sequelae of prolonged reduced serum T levels in these women are unknown. Thus, our overall goal was to examine the effects (both beneficial and adverse) on quality of life, mood, and self-esteem, after the restoration of normal circulating levels of T in young women with POI. Specifically, this study was designed to neither examine the effects of pharmacologic doses of T on these behavioral measures, nor to specifically test the efficacy of physiologic T replacement on depressive symptoms in these women. Young women with androgen deficiency may experience a variety of physical symptoms secondary to their androgen depletion as well as psychological changes that affect their quality of life (1). Androgen deficiency in women could contribute to depressed mood and reduced quality of life (10, 11) and T replacement has been recommended in some women with POI who present with decreased well-being, persistent fatigue and low libido despite adequate estrogen replacement (12). Studies also demonstrate the beneficial effects of androgen therapy on mood and libido in hypogonadal women with a surgically-induced menopause, a concurrent medical or psychiatric illness (e.g., HIV and anorexia nervosa, respectively), or reproductive aging (13–18). A recent study also reported the beneficial effects of T replacement under placebo-controlled conditions on quality of life in women with Turner syndrome who, like women with POI, experience ovarian insufficiency (19). Finally, physiologic T replacement has been reported to improve mood in women with anorexia nervosa (18). Nonetheless,

some studies suggest that T replacement could negatively impact on mood and behavior (20–22), and, therefore, the evaluation of its effects in these women is warranted. However, there are less data available on the effects of T on quality of life and mood in young hypogonadal with POI. As part of a 3-year placebo controlled trial examining the effects of the addition of T to combined estrogen and progesterone therapy (EPT) on measures of bone density in POI, we examined the effect of adding the ovarian production rate of T on measures of quality of life, self-esteem, and depressive symptom severity in 128 women with spontaneous 46, XX POI.

Material and Methods

Participant Selection

Women between the ages of 18 and 42 years participated in this National Institute of Child Health and Human Development (NICHD) study at the NIH Clinical Research Center from 2000 to 2006. This study was part of a larger three-year randomized, controlled trial, investigating the effects of testosterone augmentation of a combined EPT regimen on measures of bone density (6). Participants were recruited through notices on the NIH home page or physician referral, and each woman provided written informed consent for participation in this IRB-approved protocol. The diagnosis of POI for this study was based on a history of at least four months of non-iatrogenic oligo-amenorrhea occurring in women < 40 years of age and at least two determinations of menopausal serum FSH levels (> 40 mIU/ml) with a normal, 46, XX peripheral karyotype. Participants with baseline serum free testosterone levels above 3.1 pg/mL or sex hormone binding globulin (SHBG) levels less than 36 nmol/L were excluded from the study. Additional exclusions included current smokers (> 2 cigarettes per day), those consuming more than two alcoholic beverages daily, and those with a BMI > 30 or < 19, or a hirsutism score > 8 (23). Medical histories, physical and gynecologic exams and psychiatric evaluations were then performed, after which women were offered participation in the trial of T replacement therapy in addition to standard EPT. After randomization, each woman was evaluated at 3 and 12 months after which their participation in this component of the trial was complete. Participants with current major depression or those on antidepressant therapy were not excluded from participation. Women who met criteria for depression on admission were offered standard antidepressant therapy by either their primary treating physician or an NIH physician re-evaluation prior to entering this trial. The dose or type of antidepressant was not altered in any woman during the 1 year trial (unless they wished to stop the medicine altogether).

Study Design and Procedures

This was a randomized, double blind, placebo-controlled, parallel design trial in which we evaluated measures of quality of life, self-esteem, and depressive symptoms for one year to examine the efficacy of physiologic T replacement in EPT (see CONSORT checklist, Supplemental Digital Content 1). This trial was part of a 3-year study that focused on bone density as the primary outcome in women with POI. After screening, all participants initiated three months of EPT to ensure that all women began the study on a standardized EPT regimen While continuing EPT, all women were randomized to one of two groups; a group consisting of EPT with placebo patch and a group consisting of EPT with testosterone

patch EPT consisted of 100 micrograms daily transdermal estradiol (Watson Laboratories, Salt Lake City, UT) with oral medroxyprogesterone acetate (10 milligrams daily for 12 days of each 28 day cycle). The dosage of T was 150 micrograms per day administered transdermally (Watson Laboratories, Salt Lake City). The NIH pharmacy generated sequentially numbered patches containing either testosterone or placebo and provided that randomization number with the study medication on the day of admission. Study participants and all raters were blinded to the intervention assignment and did not have access to lab results or rating scale scores from of prior clinic visits.

Outcome Measures

Symptom ratings and blood hormone levels were obtained at three time points as follows: at baseline prior to randomization, and at three and 12 months after initiation of T or placebo. Outcome measures for this trial were both self- and rater-administered, and included the 14-item Quality of Life Enjoyment and Satisfaction Questionnaire (Q-Les-Q) (24) and the 10-item Rosenberg Self-Esteem Scale (25) in which higher scores reflect better self-esteem and quality of life, respectively. We employed the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders (fourth edition) [DSM-IV] (SCID) (26) to establish the presence or absence of Axis I psychiatric illness, as well as the self-administered Center for Epidemiologic Studies Depression Scale (CES-D) (27) and the rater-administered Hamilton Rating Scale for Depression (HAM-D) (28) which assess depressive symptom severity. Blood samples for hormone assays were obtained from fasting women at 0800 h, separated within one hour, and frozen at −70 degrees C until the time of assay.

Hormonal Assays

Serum measurements of total testosterone (reference range 14–54.3 ng/dl) were made using radioimmunoassay (RIA) after extraction chromatography (Esoterix Endocrinology, Calabassa Hills, CA). Serum free testosterone (reference range 1.3–6.8 pg/ml) was determined with equilibrium dialysis (8) and was calculated from the ratio of radioactivity outside versus inside the cell multiplied by the total serum testosterone. Serum estradiol, FSH, and LH levels were performed at the NIH Clinical Center laboratory. Estradiol was measured by a competitive chemiluminescence immunoassay (Immulite 2000) and microparticle enzyme immunoassay (AxSYM System, Abbott Diagnostics, Abbott Park, IL) was used to measure FSH and LH. Serum SHBG was measured with an immunoradiometric assay (IRMA) developed at Esoterix Endocrinology (Calabasas Hills, CA). The serum hormone intra-assay and inter-assay coefficients of variation (CV) are as follows: total testosterone (intra-assay CV = 2.9% – 8.0%; inter-assay CV = 8.0%–15.0%), free testosterone (intra-assay CV = 6.9%; inter-assay CV = 8.9–9.4%), estradiol (both intra-assay and inter-assay CV were < 11.0%), FSH (intra-assay CV was 4.9%; inter-assay CV was 6.5%) and LH (intra-assay CV = 5.8%; and inter-assay CV = 6.4%), and SHBG (intra-assay CV = 2.4%–3.9%; inter-assay CV = 7.8% – 10.6%).

Statistical Analysis

All data were analyzed using SAS Version 9.2 statistical software (SAS Institute, Inc, Cary, NC). Baseline characteristics in women receiving T and those receiving placebo were

compared with Fisher's exact test for categorical variables and Wilcoxon rank-sum tests for continuous variables. Wilcoxon rank-sum tests were employed to compare the symptom rating scale scores and serum hormone levels of women randomized to receive T and those randomized to receive placebo. The values at each clinic visit (i.e., baseline, month 3, and month 12), as well as the differences in values between baseline and three months, and between baseline and 12 months were included in the analyses.

Additionally, we compared the numbers of women on T compared with placebo who met SCID criteria for major depression as well as the numbers who met criteria for either clinically significant depressive symptoms (defined as CES-D scores ≥ 16 (29–31)) or subsyndromal depression (defined as CES-D scores ≥ 8 and ≤16 (32, 33)), at each time point using Fisher's exact test. We also used the Fisher's exact test to compare the numbers of women in each treatment group who experienced a 50% or greater improvement (i.e., reduction) in their CES-D scale scores at 12 months compared with baseline. Results were presented as mean (SD) for interval data and counts and percentages for nominal and ordinal data. Because of the presence of missing data, we also performed a separate repeated measures analysis for each of the four symptom rating scores using SAS's PROC MIXED, comparing the T and placebo treatment groups at 3-months and 12-months. The predictor variable of interest (treatment) and month were modeled as fixed effects. The interaction term of month*treatment was not significant and was therefore dropped from the final models. We included the baseline symptom rating as a covariate and used a covariance structure of compound symmetry, with the Kenward and Roger method for computing the degrees of freedom. The power analysis for this study was based on the larger, 3-year study on bone density outcomes; whereby, using a power level of 0.90 and an alpha level of 0.05, a sample size of 50 women in each group was estimated to be required to detect a difference in bone mineral density related to the addition of testosterone. All P values were two tailed with a $P < 0.05$ considered significant.

Results

Baseline Characteristics

Of the 395 women with POI who were assessed for eligibility, 128 women (age range 19–41 years, mean age (31.6) were randomized into either T-treatment (n = 67) or placebo (n = 61) and 123 women completed this study (Figure 1). There was no group difference between the ages of women randomized to receive T (mean (SD) = 31.1 (5.8) years) and those randomized to receive placebo (mean 32.1 (5.5) years) nor were there group differences in the BMI of the women randomized to receive T (mean (SD) = 23.4 (3.3)) or placebo (mean (SD) = 24.2 (3.2)) or in the number of women with POI who were treated for a thyroid condition (n = 5, and 6 in T and placebo groups, respectively) (P = NS for all comparisons). Additionally, no differences were observed between groups in baseline serum free testosterone (P = 0.7) or total testosterone measures (P = 0.2). (Table 1).

Similarly, there were no significant differences between the T treatment and placebo groups in the numbers of women meeting criteria for either major depression or subsyndromal depression at baseline. Eight women in each arm met SCID criteria for major depression (11.9% in T treatment and 13.1% in placebo groups [P = NS]), while 4.5% (n = 3) and

11.5% (n = 7 [P = NS]) women in the T and placebo groups, respectively, met criteria for current minor depression. Baseline CES-D scores did not differ between those women randomized to T and those randomized to placebo (P = NS). No group differences were observed in the number of women who met baseline CES-D criteria for clinically significant depressive symptoms (i.e., a CES-D score ≥ 16), with 15 (22.4%) in the T arm and 7 (11.7%) in the placebo arm meeting this threshold value (P = 0.16); nor were group differences observed in the number of women who met criteria for subsyndromal depression (26 (38.8%) in the T and 23 (37.7%) in the placebo groups). A history of past major depression was diagnosed by SCID interview in thirty-three women in both the T treatment arm (49%) and placebo arm (54.1%). Eighty-five women had no lifetime DSM-IV axis-I mood disorder diagnosis (n = 43 in T, n = 42 in placebo groups respectively) (P = NS for all comparisons). Finally, a similar number of women currently receiving antidepressant treatment were randomized to T treatment (18% (n = 12) and placebo (19% n = 11) (P = NS). Of the 128 enrolled participants, 123 completed the study (Figure 1).

Treatment Trial

Symptom Rating Scores (Table 2)—We found no differences in either the Q-Les-Q or Rosenberg self-esteem scale scores in women randomized to T or placebo at baseline (P = 0.4 and 0.7, respectively), at three months (P = 0.3 and 0.8, respectively) or after 12 months of treatment (P = 0.2 and 0.9, respectively).

Mood rating scale scores and syndromal measures of depression did not differ across treatment groups after 12 months. No significant differences were observed in either the CES-D scores or the change from baseline in CES-D scores between women randomized to T treatment or placebo after three months or 12 months of treatment (P = NS for all comparisons). Similarly, no differences were observed in the numbers of women with CES-D scores ≥ 16 at either the three or 12 month visits (P = 1.0, and 0.2, respectively) or the numbers of women with CES-D scores ≥ 8 and <16 at either the three or 12 month visits (P = 1.0 and 0.66, respectively). Finally, the number of women who experienced a 50% or greater reduction in CES-D scores after 12 months was no different in the T-treatment group compared with placebo-treated women (n = 28 (50.9%) and n = 19 (36.5%) in T and placebo, respectively, (P = 0.2)).

A similar pattern was observed in the effects of T and placebo treatments on HAM-D scores. No group differences were observed in HAM-D rating scale scores between T treatment and placebo at baseline (P = 0.4), nor were treatment-related differences observed in either the scores at three months (P = 0.7) and 12 months (P = 0.1) or the change in scores at three months (P = 0.4) and 12 months (P = 0.8). Similarly, no differences were observed in the number of women who met SCID criteria for major depression across treatment groups at three months (P = 1.0) or 12 months (P = 0.5). Additionally, in the sample of women who were not depressed at baseline (i.e., women who met SCID criteria for depression were excluded), no differences were observed in CES-D scores between the numbers of women who were randomized to T (N =59) and placebo (N = 52) treatment at 3 months or 12 months (P =0.8 and P =0.6, respectively). Finally, we found no differences in measures of either Q-Les-Q or Rosenberg self-esteem scores at 3 months or 12 months (P=NS for all

comparisons) among women randomized to T and placebo who were not depressed at baseline (either by SCID diagnosis or by CES-D scores for syndromal symptoms of depression).

Adverse Effects (Table 3)—No negative effects on mood were uniformly observed between groups. Indeed, no women in either treatment group dropped out of the study because of negative mood symptoms. Five women from the T treatment arm dropped out of the study after three months, two of whom reported skin irritation or rash while the other three were lost to follow-up. Of those five women, four had nearly identical CES-D and HAM-D scores at baseline and at three-months with no score > 9 on either measure. One woman reported a reduction in depressive symptoms in both CES-D and HAM-D scores from baseline (19 and 27, respectively) to three months (3 and 4, respectively) and discontinued the trial due to the development of a rash. Androgenic side effects were also reported by several women in the trial but these side effects were infrequent and did not significantly differ between treatment groups.

Serum Hormone Levels (Table 4)—Significant differences were found in serum total testosterone levels among women with POI receiving T treatment, compared with placebo after three months and at 12 months ($P < 0.001$ for both groups at each time interval respectively). Similarly, significant differences were observed in free testosterone levels among women receiving T treatment, compared with placebo after both three and 12 months ($P < 0.001$ for both groups respectively). In contrast, no group differences were observed in serum measures of estradiol, FSH, LH or SHBG at baseline, three months or at 12 months (P = NS for all comparisons).

Discussion

This is the first large scale study to evaluate the effects of physiologic T replacement with EPT on measures of quality of life and depressive symptoms in a sample of women with well characterized POI. T replacement had neither detrimental nor beneficial effects, compared to placebo, on measures of quality of life, self-esteem, and several measures of mood. Despite significant elevations of serum T levels during treatment, we did not observe significant improvements compared to placebo in any of the behavioral outcomes we measured. A considerable number of women in both treatment groups met criteria for either subsyndromal or syndromal depression (i.e., 61.2% and 50.0 % in T and placebo, respectively); nonetheless, only a small number of women in this study met criteria for major depression, most of whom received concurrent use of antidepressant medications. Thus the results of this study cannot be interpreted to define the efficacy of T replacement in women with POI who have major depression. Albeit indirect, evidence of the failure of T to significantly improve mood may suggest that depressive symptoms in women with POI are androgen independent, and perhaps related to other environmental or reproductive factors experienced by these women.

Our finding that T augmentation of EPT was no more effective than EPT alone on measures of quality of life and mood is consistent with some but not all studies. Indeed, overall the results of studies examining the effects of T replacement in addition to estrogen therapy on

mood and well-being are mixed. In ostensibly the same study, Sherwin and Gelfand (34, 35) compared the effects of intramuscular preparations of estrogens with and without androgen augmentation on symptoms of depression and well-being in women who had undergone a surgical menopause. Improvements in symptoms were observed relative to baseline with both estrogen alone and estrogen/androgen preparations; however, depression measures did not differ between women receiving combined estrogen/androgen and those receiving estrogens alone, whereas feelings of well-being were significantly higher in those women treated with the addition of T replacement. Mood enhancement also was reported using oral preparations of estrogen and testosterone in postmenopausal (36) and oophorectomized women (37), respectively. As in our study, most of the women in these negative trials were recruited because they were hypogonadal but not necessarily symptomatic, and therefore, baseline levels of mood and quality of life could have been too high or too low, respectively to show a beneficial effect of T treatment. Several studies have reported benefits of T replacement in hypogonadal women (after oophorectomy), whose selection for entry into the trials was based on their complaints of clinically significant loss of sexual function (38–41). In the study by Shifren *et al.*(40), significant differences between T and placebo in measures of mood were only observed in women taking a higher dose of transdermal T (i.e., 300 micrograms per day), but not in those receiving a dose equivalent to that employed in this trial. Thus it is possible that had we selected more symptomatic women with POI, or employed higher doses of transdermal T, a more robust difference in mood measures or quality of life would have been observed in our study. Nonetheless, physiologic doses of T replacement were observed to improve mood in a group of women with anorexia nervosa in whom depression rating scale scores were consistent with clinically significant depression (i.e., Beck Depression Inventory ≥ 10) in approximately 50% of the sample (18). Women with POI differ from women who have undergone a surgical menopause in several clinical characteristics including the early age of onset of ovarian insufficiency, the impact on potential fertility, and frequency of vasomotor symptoms. Thus, a better comparison group is women with Turner Syndrome (TS), who also experience ovarian insufficiency from an early age. Indeed, in a prior study we observed similar levels of anxiety, shyness, and low self-esteem in women with TS and POI, both groups being more symptomatic than a healthy comparison group (4). There has been one placebo-controlled trial of T replacement in addition to estrogen therapy in women with TS (19). In this study, quality of life measures improved on 1.5 mg a day of methyltestosterone compared with placebo, as did several metabolic, cognitive, and sexual function measures. However, as with some studies in women who had undergone a surgical menopause, the women with TS were taking oral estrogens and had lower baseline free T levels than the women in this study (i.e., mean free T levels at baseline were 1.1 pg/ml vs 2.2 pg/ml respectively). The dose of T employed in the women with TS was 1.5 mg of methyltestosterone, and, therefore, also could have been supraphysiologic compared to the doses we employed. Overall, our findings of a lack of efficacy of physiologic low dose testosterone addition to estrogen therapy are consistent with several previous studies in hypogonadal women and do not support the routine use of T to improve mood symptoms in women with POI. However, we cannot rule out the possibility that the selection of more symptomatic women or those with lower baseline T levels, or the use of higher doses of T, would result in a more robust pattern of effects of T on mood symptoms, quality of life, and self-esteem measures.

A recent meta-analysis supported several reports that POI was accompanied by low testosterone levels compared with healthy young women (9). Evidence of testosterone's beneficial effects on measures of quality of life, depressive symptoms, and well-being have provided indirect evidence that hypoandrogenism could underlie these symptoms. The results of this trial then would suggest that current mood symptoms, self-esteem, and quality of life are not reflective of the androgen deficiency state present in women with POI. In a previous study, we observed a higher than expected prevalence of a past history of major depression in women with POI (42) suggesting that POI conveys a vulnerability to develop depressive illness. The lack of therapeutic effects of T on depressive symptoms in these women with POI does not rule out the possibility that testosterone deficiency contributes to an underlying vulnerability to depressive illness. Thus it is possible that as with depression during the natural menopause (43) abnormalities of T secretion could contribute to an increased risk for depression. Moreover, a preventive effect of T replacement on the risk for depression in these women remains a possible intervention to examine. Similarly, since estrogen has reported antidepressant effects in women with perimenopausal depression (44; 45), it is also possible that the pre-study treatment with EPT reduced depressive symptoms sufficiently to preclude any observable additional antidepressant effects of testosterone in these women.

The findings of this study initially could appear to conflict with studies that have examined the effects of T and EPT replacement (14; 37–41) or T alone (46; 47) on measures of sexual dysfunction. There is considerable overlap in the items measured in most QOL scales and the symptoms of both depression and decreased sexual function. Studies showing an improvement in QOL or well-being after T treatment in hypogonadal women with sexual dysfunction do not necessarily reflect a concurrent improvement in mood symptoms, nor can the observed impact of T on mood and well-being be generalized to hypogonadal women who do not have complaints of sexual dysfunction. Thus comparisons are less meaningful across studies in which differing presenting symptoms are the targets of treatment. In this study, in which women were recruited for a study of bone density, QOL measures could reflect a range of phenomena unrelated to depressive symptoms, and depressive symptoms would best be measured with specific depression rating scales and structured diagnostic interviews. Since neither measures of depression nor QOL improved after EPT+T treatment compared with EPT, our data suggests that in women with POI neither measure should be an indication for T treatment in keeping with current Clinical Practice Guidelines (48).

The limitations in this study should be discussed. First, the presence of a small proportion of women who met criteria for major depression prevents an adequate evaluation of the antidepressant effects of T compared with placebo. This was a clinic-based sample not selected for the presence or absence of depression and our sample reflected the profile of symptoms encountered in a clinic-based practice more so than a sample selected for psychiatric morbidity. While we did include women with depression, only eight women in our study (approximately 12%) in the testosterone treatment arm met DSM criteria for major depression (compared with a 12 month prevalence of 8.5% for major depression in the general population of women (49)). The number of depressed women was not sufficiently large to permit the demonstration of a beneficial effect of T on mood, particularly in the context of concurrent EPT, which has demonstrated antidepressant effects (40; 41).

However, to achieve 80% power in our analysis, we would need a sample size of 556 women with POI in each treatment arm for a total 1112 participants in the study in order to demonstrate antidepressant effects - a sample size of depressed women prohibitively large to ethically conduct an appropriately powered study of the antidepressant efficacy of T. Moreover, as stated earlier this study was not designed to specifically evaluate the efficacy of T as a treatment for depression. Nonetheless, a sample containing a larger proportion of women with depressive symptoms could allow us to make a more definitive evaluation of T's mood elevating effects. Additionally, the concurrent use of antidepressant medication could have obscured T's potential antidepressant effects in those depressed women on standard antidepressant medications. Finally, had we employed supraphysiologic doses of T we may have identified a greater positive effect of T on quality of life measures or mood symptoms.

Conclusion

Our data suggest that neither low self-esteem nor depressive symptoms are caused by low androgen levels in women with POI (or improve with physiologic repletion of T). Moreover, reports of decreased quality of life and self-esteem or the presence of depressive symptoms could reflect a range of emotional stresses related to both the physiology of POI and the impact of the diagnosis of POI on an individual woman independent of an underlying gonadal steroid deficiency (both T and E2). There are important theoretical and practical implications to our findings. The findings beg the question, "Does physiological ovarian sex-hormone therapy in young women with primary ovarian insufficiency improve quality of life and parameters of self-esteem and mood?" Despite the fact that normal ovarian function provides menstruating women with a daily production of approximately 150 microgram of T, T replacement in these young women is not current standard of care (1). We have demonstrated that the 150 microgram T patch does indeed achieve physiologic hormone levels in women with POI with no significant increase in side effects or worsening of mood compared to placebo. However, our findings suggest that T replacement has a non-significant trend toward mood improvement in women with POI, suggesting that other mechanisms might play a more significant role in the altered mood that may accompany this disorder.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Dr. Rubinow is a funded member of the Editorial Board for a Servier sponsored journal (Dialogues in Clinical Neuroscience).

Sources of Funding: This work was supported by the Intramural Research Program of the National Institute of Child Health and Human Development, National Institutes of Health, and Procter & Gamble under Cooperative Research and Development Agreement (CRADA 00983). This work was written as part of Peter Schmidt's and Lawrence Nelson's official duties as Government employees. The views expressed in this article do not necessarily represent the views of the NIMH, NIH, HHS, or the United States Government.

The authors would like to thank James Troendle, for his statistical analyses on an earlier version of the manuscript.

References

1. Nelson LM. Primary ovarian insufficiency. N Engl J Med. 2009; 360:606–614. [PubMed: 19196677]
2. Nelson LM, Covington SN, Rebar RW. An update: spontaneous premature ovarian failure is not an early menopause. Fertil Steril. 2005; 83:1327–1332. [PubMed: 15866564]
3. Hartmann BW, Kirchengast S, Albrecht A, Laml T, Soregi G, Huber JC. Androgen serum levels in women with premature ovarian failure compared to fertile and menopausal controls. Gynecol Obstet Invest. 1997; 44:27–131.
4. Schmidt PJ, Cardoso GMP, Ross JL, Haq N, Rubinow DR, Bondy CA. Shyness, social anxiety, and impaired self-esteem in Turner syndrome and premature ovarian failure. J A M A. 2006; 295:1374–1376.
5. Pal L, Bevilacqua K, Zeitlian G, Shu J, Santoro N. Implications of diminished ovarian reserve (DOR) extend well beyond reproductive concerns. Menopause. 2008; 15:1086–1094. [PubMed: 18791485]
6. Kalantaridou SN, Calis KA, Mazer NA, Godoy H, Nelson LM. A pilot study of an investigational testosterone transdermal patch system in young women with spontaneous premature ovarian failure. J Clin Endocrinol Metab. 2005; 90:6549–6552. [PubMed: 16174721]
7. Papalia MA, Davis SR. What is the rationale for androgen therapy for women? Treat Endocrinol. 2003; 2:77–84. [PubMed: 15871544]
8. Kalantaridou SN, Calis KA, Vanderhoof VH, et al. Testosterone deficiency in young women with 46,XX spontaneous premature ovarian failure. Fertil Steril. 2006; 86:1475–1482. [PubMed: 17070197]
9. Janse F, Tanahatoe SJ, Eijkemans MJC, Fauser BCJM. Testosterone concentrations, using different assays, in different types of ovarian insufficiency: a systematic review and meta-analysis. Hum Reprod Update. 2012; 18:405–419. [PubMed: 22525963]
10. Bachmann G, Bancroft J, Braunstein G, et al. Female androgen insufficiency: the Princeton consensus statement on definition, classification, and assessment. Fertil Steril. 2002; 77:660–665. [PubMed: 11937111]
11. Santoro N, Torrens J, Crawford S, et al. Correlates of circulating androgens in mid-life women: the study of women's health across the nation. J Clin Endocrinol Metab. 2005; 90:4836–4845. [PubMed: 15840738]
12. Kalantaridou SN, Davis SR, Nelson LM. Premature ovarian failure. Endocrinol Metab Clin North Am. 1998; 27:989–1006. [PubMed: 9922918]
13. Burd ID, Bachmann GA. Androgen replacement in menopause. Curr Womens Health Rep. 2001; 1:202–205. [PubMed: 12112971]
14. Davis SR, Goldstat R, Papalia MA, et al. Effects of aromatase inhibition on sexual function and well-being in postmenopausal women treated with testosterone: a randomized, placebo-controlled trial. Menopause. 2006; 13:37–45. [PubMed: 16607097]
15. Miller K, Corcoran C, Armstrong C, et al. Transdermal testosterone administration in women with acquired immunodeficiency syndrome wasting: a pilot study. J Clin Endocrinol Metab. 1998; 83:2717–2725. [PubMed: 9709937]
16. Mazer NA. Testosterone deficiency in women: etiologies, diagnosis, and emerging treatments. Int J Fertil Womens Med. 2002; 47:77–86. [PubMed: 11991434]
17. Khastgir G, Studd J. Hysterectomy, ovarian failure, and depression. Menopause. 1998; 5:13–122.
18. Miller KK, Grieco KA, Klibanski A. Testosterone Administration in Women with Anorexia Nervosa. The Journal of Clinical Endocrinology and Metabolism. 2005; 90:1428–1433. [PubMed: 15613421]
19. Zuckerman-Levin N, Frolova-Bishara T, Militianu D, Levin M, Aharon-Peretz J, Hochberg Z. Androgen replacement therapy in Turner syndrome: a pilot study. J Clin Endocrinol Metab. 2009; 94:4820–4627. [PubMed: 19846743]
20. Strauss RH, Liggett MT, Lanese RR. Anabolic Steroid Use and Perceived Effects in Ten Weight-Trained Women Athletes. J A M A. 1985; 253:2871–2873.

21. Hohlagschwandtner M, Husslein P, Klier C, Ulm B. Correlation between serum testosterone levels and peripartal mood states. Acta Obstet Gynecol Scand. 2001; 80:326–330. [PubMed: 11264607]
22. Sherwin BB, Gelfand MM. The role of androgen in the maintenance of sexual functioning in oophorectomized women. Psychosom Med. 1987; 49:397–409. [PubMed: 3615768]
23. Ferriman D, Gallwey JD. Clinical assessment of body hair growth in women. J Clin Endocrinol Metab. 1961; 21:1440–1447. [PubMed: 13892577]
24. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. Psychopharmacol Bull. 1993; 29:321–326. [PubMed: 8290681]
25. Rosenberg, M. Society and the adolescent self-image. Princeton, NJ: Princeton University Press; 1965.
26. First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured clinical interview for DSMIV axis I disorders - patient edition. New York, NY: Biometrics Research Department, New York State Psychiatric Institute; 1996.
27. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas. 1977; 1:385–401.
28. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960; 23:56–62. [PubMed: 14399272]
29. Comstock GW, Helsing KJ. Symptoms of depression in two communities. Psychol Med. 1976; 6:551–563. [PubMed: 1005571]
30. Freeman EW, Sammel MD, Liu L, Gracia CR, Nelson DB, Hollander L. Hormones and menopausal status as predictors of depression in women in transition to menopause. Arch Gen Psychiatry. 2004; 61:62–70. [PubMed: 14706945]
31. Roberts RE, Vernon SW. The center for epidemiologic studies depression scale: its use in a community sample. Am J Psychiatry. 1983; 140:41–46. [PubMed: 6847983]
32. Cohen CI, Goh KH, Gustave M. A prospective study of outcome and predictors of subclinical and clinical depression in an older biracial sample of psychiatric outpatients. J Affective Disord. 2010; 121:204–211.
33. Meeks TW, Vahia IV, Lavretsky H, Kulkarni G, Jeste DV. A tune in "a minor" can "b major": A review of epidemiology, illness course, and public health implications of subthreshold depression in older adults. J Affective Disord. 2011; 129:126–142.
34. Sherwin BB, Gelfand MM. Sex steroids and affect in the surgical menopause: a double-blind, cross-over study. Psychoneuroendocrinology. 1985; 10:325–335. [PubMed: 3903822]
35. Sherwin BB, Gelfand MM. Differential symptom response to parenteral estrogen and/or androgen administration in the surgical menopause. Am J Obstet Gynecol. 1985; 151:153–160. [PubMed: 3881960]
36. Myers LS, Dixen J, Morrissette D, Carmichael M, Davidson JM. Effects of estrogen, androgen, and progestin on sexual psychophysiology and behavior in postmenopausal women. J Clin Endocrinol Metab. 1990; 70:1124–1131. [PubMed: 1690746]
37. Floter A, Nathorst-Boos J, Carlstrom K, Von Schoultz B. Addition of testosterone to estrogen replacement therapy in oophorectomized women: effects on sexuality and well-being. Climacteric. 2002; 5:357–365. [PubMed: 12626215]
38. Simon J, Braunstein G, Nachtigall L, et al. Testosterone patch increases sexual activity and desire in surgically menopausal women with hypoactive sexual desire disorder. J Clin Endocrinol Metab. 2005; 90:5226–5233. [PubMed: 16014407]
39. Braunstein GD, Sundwall DA, Katz M, et al. Safety and efficacy of a testosterone patch for the treatment of hypoactive sexual desire disorder in surgically menopausal women: a randomized, placebo-controlled trial. Arch Intern Med. 2005; 165:1582–1589. [PubMed: 16043675]
40. Shifren JL, Braunstein GD, Simon JA, et al. Transdermal testosterone treatment in women with impaired sexual function after oophorectomy. N Engl J Med. 2000; 343:682–688. [PubMed: 10974131]
41. Nathorst-Boos J, Floter A, Jarkander-Rolff M, Carlstrom K, Von Schoultz B. Treatment with percutanous testosterone gel in postmenopausal women with decreased libido - effects on sexuality and psychological general well-being. Maturitas. 2006; 53:11–18. [PubMed: 16183220]

42. Schmidt PJ, Luff JA, Haq NA, et al. Depression in women with spontaneous 46, XX primary ovarian insufficiency. J Clin Endocrinol Metab. 2011; 96:E278–E287. [PubMed: 21047929]
43. Bromberger JT, Schott LL, Kravitz HM, et al. Longitudinal change in reproductive hormones and depressive symptoms across the menopausal transition: results from the study of women's health across the nation (SWAN). Arch Gen Psychiatry. 2010; 67:598–607. [PubMed: 20530009]
44. Schmidt PJ, Nieman L, Danaceau MA, et al. Estrogen replacement in perimenopause-related depression: a preliminary report. Am J Obstet Gynecol. 2000; 183:414–420. [PubMed: 10942479]
45. Soares CN, Almeida OP, Joffe H, Cohen LS. Efficacy of estradiol for the treatment of depressive disorders in perimenopausal women: a double-blind, randomized, placebo-controlled trial. Arch Gen Psychiatry. 2001; 58:529–534. [PubMed: 11386980]
46. Goldstat R, Briganti E, Tran J, Wolfe R, Davis SR. Transdermal testosterone therapy improves well-being, mood, and sexual function in premenopausal women. Menopause. 2003; 10:390–398. [PubMed: 14501599]
47. Davis SR, Moreau M, Kroll R, et al. Testosterone for low libido in postmenopausal women not taking estrogen. N Engl J Med. 2008; 359:2005–2017. [PubMed: 18987368]
48. Wierman ME, Basson R, Davis SR, et al. Androgen therapy in women: an endocrine society clinical practice. J Clin Endocrinol Metab. 2006; 91:3697–3710. [PubMed: 17018650]
49. Kessler RC. Appendix table 1: lifetime prevalence of DSM-IV/WMH-CIDI disorders by sex and cohort. National Comorbidity Survey. 2005 Available at: URL: http://www.hcp.med.harvard.edu/ncs/ftpdir/table_ncsr_by_gender_and_age.pdf.
50. Rush AJ, Kraemer HC, Sackeim HA, et al. Report by the ACNP task force on response and remission in major depressive disorder. Neuropsychopharmacology. 2006; 31:1841–1853. [PubMed: 16794566]
51. Ehrmann DA, Rosenfield RL. Clinicial review 10: an endocrinologic approach to the patient with hirsutism. J Clin Endocrinol Metab. 1990; 71:1–4. [PubMed: 2196276]
52. Cremoncini C, Vignati E, Libroia A. Treatment of hirsutism and acne in women with two combinations of cyproterone acetate and ethinylestradiol. Acta Eur Fertil. 1976; 7:299–314. [PubMed: 140576]

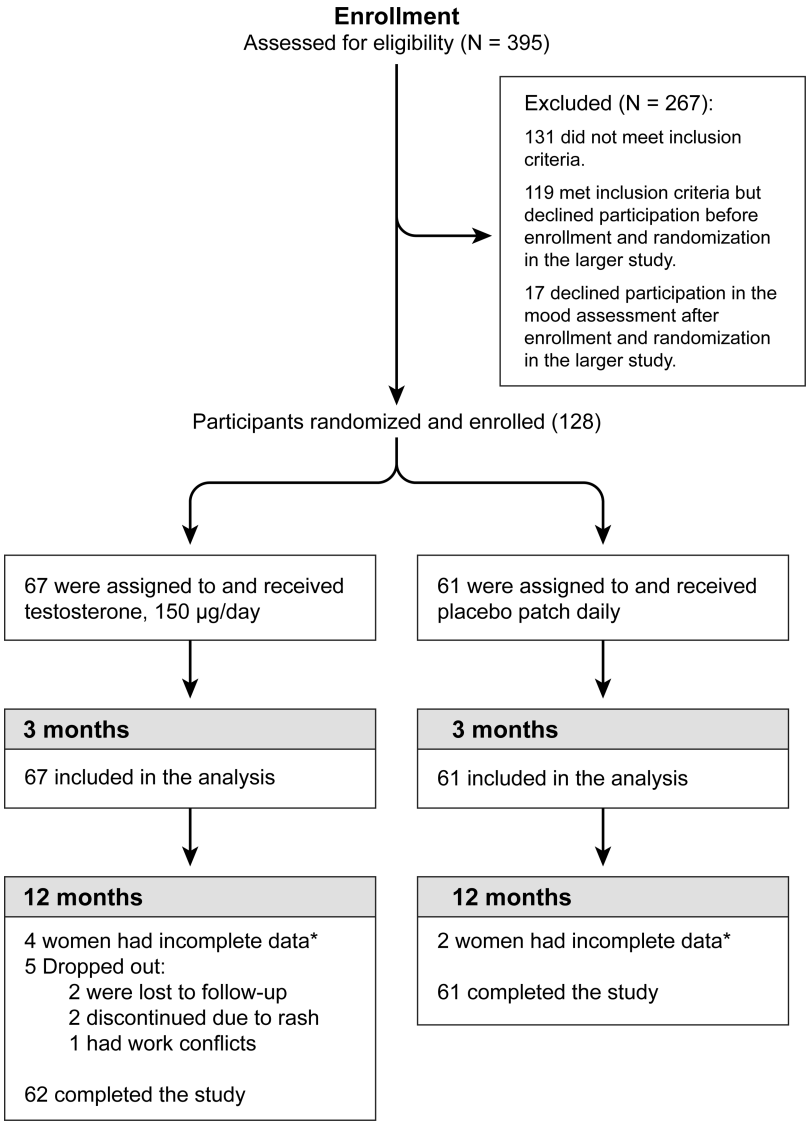

Figure 1. Study population enrollment and outcomes

An initial sample of 395 women with POI was admitted to the NIH Clinical Center and participated in the screening portion of the longer term study evaluating the effects on bone health between July 2000 and February 2005. Eligibility was reported as the number of women enrolled in the larger 3-year trial. Of these, 267 met exclusion criteria and 145 met inclusion criteria and were enrolled in the larger study, of the 145 women randomized to T or placebo treatment; 17 declined participation in the mood component of the study. Thus, 128 women with POI who were randomized to either T or placebo treatment were enrolled in our study with 123 women completing the study. *Of the six women who had incomplete ratings at the 12-month visit, two in the placebo arm and one in the T treatment arm had incomplete behavior ratings (i.e., CES-D or HAM-D or SCID); the other three women, present in the T treatment arm, completed the Q-Les-Q and one of those also completed the SCID). Therefore, the completed data obtained from these women was included in the analysis.

Table 1

Baseline characteristics of women with POI randomized to T (n = 67) or placebo (n = 61)

Variable	Testosterone	Placebo	P-value
Age (years)	31.1 (5.8)	32.1 (5.5)	NS
BMI	23.4 (3.3)	24.2 (3.2)	NS
Parity	N = 10 0.19 (0.5)	N = 13 0.26 (0.54)	NS
Number on Thyroid therapy	N = 5	N = 6	NS
CES-D ≥16	N= 15 (22.4%)	N = 7 (11.7 %)	NS
CES-D ≥8 and <16	N= 26 (38.8%)	N= 23 (37.7%)	NS
Current psychotropic medication use[a]	N= 12 (18.5%)	N= 11 (19.0%)	NS
Past history of Axis I DX[b]	N= 43 (64%)	N= 42 (69%)	NS
MDE	33 (49%)	33 (54.1%)	
Minor Depression	8 (11.9%)	8 (13.1%)	
Dysthymia	1 (0.01%)	1 (0.02%)	
PTSD	1 (0.01%)	None	
SCID (current MDE)	N= 8 (11.9%)	N= 8 (13.1%)	NS
SCID (current minor depression)	N= 3 (4.5%)	N= 7 (11.5%)	NS
Baseline Free testosterone (ng/dl)	2.2 (1.4)	2.1(1.2)	NS
Baseline Total testosterone (ng/dl)	21.7 (9.9)	18.7 (6.8)	NS

Reported values are expressed as mean (SD) unless otherwise indicated. Eligible women met the following inclusion criteria: 1) diagnosis of spontaneous 46, XX POI before the age of 40 yr (i.e. at least 4 months of oligomenorrhea or amenorrhea and FSH levels in the menopausal range, as defined by laboratory confirmation on two separate occasions, at least 1 month apart); 2) age between 18 and 42; and 3) non-iatrogenic cause of POI.

[a] Current psychotropic medication use included citalopram, paroxetine, sertraline, and buproprion).

[b] Past diagnoses and current depression were based on SCID scoring using DSM-IV diagnostic criteria for axis-I one mood disorders.

Table 2

Outcome measures of symptom ratings at baseline, three months and twelve months. Unless otherwise indicated, data are presented in means (SD).

Rating Scale[a]	Testosterone [N=67]	Placebo [N=61]	*P*-value[b]
Q-Les-Q			
Baseline	3.9 (0.7) [N=67]	4.0 (0.7) [N=61]	0.4
3 months	4.3 (0.7) [N=67]	4.2 (0.7) [N=58]	0.3
12 months	4.3 (0.7) [N=58]	4.2 (0.7) [N=57]	0.2
Rosenberg's Self Esteem Scale			
Baseline	34.0 (5.3) [N=67]	34.7 (4.2) [N=61]	0.7
3 months	35.8 (4.2) [N=67]	35.9 (4.3) [N=59]	0.8
12 months	36.4 (3.7) [N=58]	36.3 (4.1) [N=58]	0.9
CES-D			
Baseline	10.7 (8.6) [N=67]	9.2 (7.8) [N=60]	0.3
3 months	6.5 (5.4) [N=67]	6.6 (5.8) [N=60]	1.0
12 months	5.9 (5.4) [N=58]	8.2 (8.5) [N=58]	0.4
HAM-D			
Baseline	9.5 (9.5) [N=67]	10.2 (8.8) [N=61]	0.4
3 months	6.1 (6.4) [N=67]	6.4 (6.5) [N=59]	0.7
12 months	5.1 (5.7) [N=57]	7.4 (8.1) [N=57]	0.1
SCID (reported as N (%) current depressed)			
Baseline	8 (12%) [N=67]	8 (13%) [N=61]	1.0
3 months	6 (9.1%) [N=66]	5 (8.3%) [N=60]	1.0
12 months	9 (15%) [N=59]	13 (22%) [N=59]	0.5

[a] Scale scoring for the Q-Les-Q is 0–100 where higher numbers reflect increasing quality of life. Rosenberg's self-esteem scale scoring is 10–40 with higher scores representing increased self-esteem. CES-D scores reflect depressive symptom severity using a scale range of 0–60 with higher scores reflecting greater symptom severity (scores of 16 or higher are consistent with major depression). HAM-D scale scoring is based on a 21 item questionnaire with a score of 14 or higher consistent with moderate depression (higher scores are indicative of symptom severity (50). SCID values represent the number of women who met DSM-IV criteria for a current MDE.

[b] *P*-values presented were determined by Wilcoxon rank-sum tests in all outcome measures except the SCID, which was analyzed using Fisher's exact test.

Table 3

Summary of side effects/adverse events between treatment groups at baseline, 3 months and 12 months, reported as numbers of women with event (unless otherwise noted).

Event Measure[a]	Baseline	3 months	12 months
Skin changes[b]			
Testosterone	None	2	1
Placebo	1	2	1
Hirsutism[c]			
Testosterone	None	None	1
Placebo	None	None	None
Acne[d]			
Testosterone	5	7	5
Placebo	2	4	2
Depilation Frequency[e]			
Testosterone	3	5	3
Placebo	2	3	3
Dropped out due to side effects/ adverse events[f]			
Testosterone	2 (1.34%)		
Placebo	None		

[a] *P*-values were > 0.05 for all comparisons.

[b] Skin changes were reported as the number of episodes women experienced irritation, rash, or pigment changes during the course of treatment. The reported skin changes did not require medical attention. One woman in each treatment group experienced localized skin darkening near the EPT patch site.

[c] Hirsutism was measured using the Ehrman scale scoring (51), where a code of 1 represents a score of 1–8 (normal) and a score > 8 is coded as a 2 (abnormal). Only abnormal hirsutism scores are reported (one woman in the T treatment group scored a 10 at 12 months).

[d] Acne measurement represents a modified Cremonici scale scoring (52) where a score of 0–1 represents less than 10 acne pustules (normal) and a score of 2–4 represents greater than 10 pustules (abnormal). Only the number of women who had abnormal acne scores is presented.

[e] Depilation frequency is reported as the number of times in the past month that hair has been removed from lip or chin (range = 0–3). A score of 3 was reported as abnormal.

[f] Reported adverse events at 3 months: two women withdrew from the study due to rash at the study patch site, neither of whom reported other (androgenic) side effects as the basis for their decision to withdraw from the study.

Table 4

Serum hormone levels between T treatment and Placebo in women with POI at baseline and after 3 and 12 months of treatment. Scores are presented in means (SD).

Variable[a]	Baseline	3 months	12 months
Free Testosterone[b] (pg/ml) reference range 1.3–6.8			
Testosterone	2.2 (1.4) [N=65]	4.7 (1.9) [N=63]	5.0 (3.4) [N=56]
Placebo	2.1 (1.2) [N=59]	1.9 (0.9) [N=57]	1.6 (0.8) [N=54]
Total Testosterone[b] (ng/dl) reference range 14–54.3			
Testosterone	21.7 (9.9) [N=66]	51.9 (20.1) [N=65]	49.6 (22.9) [N=56]
Placebo	18.7 (6.8) [N=59]	17.5 (6.0) [N=57]	16.4 (5.8) [N=54]
Estradiol (pg/ml) reference range 34–225			
Testosterone	90.7 (58.1) [N=67]	111.1 (89.2) [N=67]	90.0 (67.3) [N=59]
Placebo	100.5 (109.2) [N=61]	102.3 (85.2) [N=60]	91.7 (55.2) [N=59]
FSH[c] (IU/l) reference range 20–131			
Testosterone	24.9 (24.4) [N=67]	19.9 (18.6) [N=67]	20.0 (20.5) [N=59]
Placebo	22.7 (24.2) [N=61]	17.5 (13.4) [N=60]	19.7 (16.3) [N=59]
LH (IU/l) reference range 2–49			
Testosterone	15.2 (13.7) [N=67]	12.6 (12.7) [N=67]	13.8 (13.7) [N=59]
Placebo	14.6 (16.9) [N=61]	9.9 (8.8) [N=60]	11.9 (13.2) [N=59]
SHBG (nmol/l) reference range 40–120			
Testosterone	108.8 (51.8) [N=66]	110.4 (63.8) [N=67]	103.9 (45.5) [N=55]
Placebo	95.3 (37.7) [N=59]	93.3 (42.3) [N=58]	96.9 (45.5) [N=54]

[a] Reference ranges for total and free testosterone and SHBG correspond to Esoterix values, while those of estradiol, FSH, and LH correspond to the NIH clinical center laboratory database.

[b] *P*-values were < 0.001 in both free and total testosterone measures at 3and 12 months between treatment groups using Wilcoxon rank-sum tests. *P*-values for all other serum hormone comparisons were > 0.05.

[c] FSH levels represent study participation values when women were on EPT and do not reflect the diagnostic serum FSH levels (> 40 IU/l on two occasions) obtained during screening, three months prior to study initiation when EPT was started.