DATE: October 2017

OREGON HEALTH AND SCIENCE UNIVERSITY
OFFICE OF CLINICAL INTEGRATION AND EVIDENCE-BASED PRACTICE

Evidence-Based Practice Summary

Benefits and Harms of Flibanserin and Testosterone for the Treatment of Low Libido in Menopausal Women

Prepared for: Karen E. Adams, M.D. and Nicole Harrington Cirino, M.D.
Authors: Tovah Kohl, MA

BACKGROUND

Hypoactive sexual desire disorder (HSDD) is a commonly diagnosed form of female sexual dysfunction. The essential feature of female HSDD is a deficiency or absence of sexual fantasies and desire, for sexual activity (low libido) that causes marked distress or interpersonal difficulty. HSDD is a prevalent condition in women of all ages. There are discrepancies between studies, but it is estimated that HSDD affects between 12-7% of women (Vallejos 2017). Testosterone and flibanserin are two pharmacologic treatments for HSDD in menopausal women.

In both surgically and naturally menopausal women, testosterone therapy, alone or combined with hormonal replacement therapy, is suggested to be associated with improvement in sexual function, energy, and quality of life. There are, however, concerns regarding the use of testosterone in women need to be considered. Some of the most common side effects that are the androgenic side effects, especially hirsutism and acne (Elraiyah 2014). Many of the international society guidelines cautiously recommend using testosterone or using it if other treatments have failed (Endocrine Society 2014).

The Food and Drug Administration (FDA) approved flibanserin in August 2015 under the brand name Addyi. The once-daily, nonhormonal pill is indicated for the treatment of HSDD in premenopausal women. Flibanserin is the first prescription medicine for the treatment of this sexual dysfunction. The FDA approval is not without warnings; flibanserin is subject to the Risk Evaluation and Mitigation Strategy (REMS) to ensure safe use by patients, and has a boxed warning directed to prescribing providers. The goal is to inform patients and providers of the increased risk of hypotension and syncope (Vallejos 2017).

DATE: October 2017

ASK THE QUESTION

Question 1: In treatment of low libido in menopausal women, what are the benefits (and harms, if any) of flibanserin and testosterone?

SEARCH FOR EVIDENCE

Databases included Ovid MEDLINE, Cochrane Database of Systematic Reviews, National Guideline Clearinghouse

Search Strategy see Appendix B

Filters/limits included articles in the English language published from 2007– August 2017

CRITICALLY ANALYZE THE EVIDENCE

Primary Literature

The literature search resulted in more than 700 articles that analyzed the benefits and harms of androgens to treat low libido. Only ten of these considered harms and benefits among menopausal women. The ten studies included systematic reviews, randomized controlled trials, and observational studies.

1. Benefits of Flibanserin compared to placebo: Two systematic reviews examined the efficacy of flibanserin. The first systematic review (Jaspers 2016) found that there was an increase in sexually satisfying events (SSEs) by 0.49 events [95% CI, 0.32-0.67]), eDiary desire score (1.63 [95% CI, 0.45-2.82]), and Female Sexual Function Index (FSFI) scores (0.27 [95% CI, 0.17-0.38]) in patients taking flibanserin. The second review (Gao 2015) found an increase of SSEs by 0.59 events [95% CI = 0.37–0.80, p < 0.00001]); sexual desire score (1.91 [95% CI = 0.21 to 3.60, p = 0.03]) and FSFI desire score (0.32 [95% CI = 0.19–0.46, p < 0.00001]) in patients taking flibanserin.

Overall Level of Evidence: Moderate, downgraded due to design limitations

2. Harms of Flibanserin compared to placebo:
   Two systematic reviews examined the harms of flibanserin. Jaspers (2016) found that the risk for any adverse events (AEs), which also included non– drug-related AEs such as common cold, was 1.29 (95% CI, 1.15-1.45) times higher for Flibanserin than for placebo. The risk for dizziness was 4.00 (95% CI, 2.56-6.27) times higher with flibanserin than with placebo; for

somnolence,3.97 (95% CI, 3.01-5.24) times higher with flibanserin; for nausea, 2.35 (95%CI, 1.85-2.98) times higher with flibanserin; and for fatigue, 1.64 (95% CI, 1.27-2.13) times higher with flibanserin. The overall risk ratio for the four most common AEs was 2.86 (95%CI, 2.32-3.52). The absolute number of serious AEs was small, and the risk ratio did not differ between Flibanserin and placebo users (1.48 [95%CI, 0.91-2.41]). In the second review (Gao 2015), there was a higher proportion of women who experienced an AE while taking flibanserin (OR = 1.54 [95% CI = 1.34–1.76, p < 0.00001]). There was also a higher proportion of nervous system disorders (OR = 2.58 [95% CI = 2.10 to 3.18, p < 0.00001]) and fatigue (OR = 1.71 [95% CI = 1.20–2.43, p = 0.003]) in the flibanserin groups.

Overall Level of Evidence: Moderate, downgraded due to design limitations

3. Benefits of Testosterone Compared to Placebo: Four studies, one systematic reviews and three randomized controlled trials (RCTs), evaluated the benefits of treating low libido in women with testosterone. The systematic review (Achilli 2017), summarized the efficacy and safety of transdermal (TT) in postmenopausal women for the treatment of HSDD. The pooled results showed that the group receiving testosterone had significantly more SSEs (MD 0.92 [95% CI, 0.65, 1.19; p<0.00001]) and experienced significantly more desire (MD 6.09 [95% CI, 4.51, 7.68; p<0.00001]) compared with the placebo group. The first RCT (Fooladi 2014) investigated the efficacy of TT as a treatment for SSRI/SNRI-emergent loss of libido. The results of the study found no difference at 12 weeks between the TT group and the placebo group measured by the Sabbatsberg Sexual Self-Rating Scale from baseline. However, there was an increase of 2.3 SSEs with TT vs. 0.1 with placebo (p = 0.02). Labrie (2014) investigated the influence of moderate/severe pain at sexual activity at baseline on female sexual dysfunction (FSD) in post-menopausal women following prasterone administration. The benefits over placebo in prasterone-treated women for desire was improved at week 12 by 22% (p = 0.016), 51% (P = 0.0047), 31% (p = 0.2845) and 48% (p = 0.0072) in the placebo, 0.25%, 0.5% and 1.0% prasterone groups, respectively. The third RCT (Tungmunsakulchai 2015) evaluated the effectiveness of testosterone undecanoate on sexual function in postmenopausal women utilizing the FSFI score. After eight weeks of treatment, the FSFI scores significantly improved in both groups when compared to the baseline, but the FSFI scores from the testosterone group were significantly higher than in the placebo group post-treatment (28.6 ± 3.6, 25.3 ± 6.7, respectively, p = 0.04).

Overall Level of Evidence: Moderate, downgraded due to design limitations

4. Harms of Testosterone Compared to Placebo: Four studies, one systematic review and three RCTs, evaluated the harms of treating low libido in women with testosterone. The systematic review (Achilli 2017) summarized the efficacy and safety of transdermal T in postmenopausal women for the treatment of HSDD. The pooled results from all seven studies showed that there was no significant difference in total adverse events (RR, 1.01 [95% CI, 0.97-1.05; p=0.77]) or severe adverse events (RR, 1.02 [95% CI, 0.62-1.68; p=0.94]). However, when separated by types of events, the T group had more statistically significant androgenic events (RR 1.37 [95% CI, 1.12- 1.69; p=0.002]), acne (RR, 1.41 [95% CI, 1.05-1.88; p=0.02]), and hair growth (RR, 1.56 [95% CI, 1.17- 2.09; p=0.003]). The RCT conducted by Fooladi (2014) found no androgenic adverse events and no clinically

relevant changes for any vital sign measurements. Labrie (2014) did not report on adverse events and Tungmunsakulchai (2015) found that no statically significant differences in acne and hirsutism between the Testosterone group and the placebo group.

Overall Level of Evidence: Moderate, downgraded due to design limitations

5. Benefits of Testosterone Compared to Other Treatments: There was one systematic review and four RCTs that compared the benefits of testosterone with other pharmacologic treatments. The systematic review (Elraiyah 2014), evaluated benefits and harms of systemic testosterone in postmenopausal women with normal adrenal function. Their analysis found that compared to the testosterone free regimen, the patients taking testosterone containing regimens had statistically significant improvements in number of satisfying sexual episodes (WMD, 1.20; 95%CI, 0.88 to 1.51), and interest in sex (SMD, 0.35; 95% CI, 0.19 to 0.52). The quality of encounters was also found to be improved by the testosterone containing regimens including; improvements in orgasm (SMD, 0.20; 95% CI, 0.09 to 0.31), arousal (SMD, 0.25; 95% CI, 0.09 to 0.40), and enjoyment of sex (SMD, 0.31; 95% CI, 0.12 to 0.51). The first RCT (Melisko 2017) evaluated the safety of intravaginal testosterone cream (IVT) or an estradiol-releasing vaginal ring (7.5 mug/d) in patients with early-stage breast cancer (BC) receiving Aromatase inhibitors (AI). Sexual interest improved for both vaginal ring (p=0.021) and IVT (p=0.02) patients. Sexual dysfunction also improved for both (p <0 .001). Single-item Cancer Rehabilitation Evaluation System sexual satisfaction score showed more improvement in vaginal ring patients (p= 0.004) than in IVT patients (p= 0.14). The second RCT (Poels 2013) assessed the hypothesis that treatment with on-demand use of T+PDE5i improves sexual functioning, particularly in women who suffer from HSDD as the result of a relative insensitivity for sexual cues. The participants either received a placebo, testosterone or testosterone plus PDE51i. For low sensitive women, sexual satisfaction during SSEs increased by a statistically significant amount (p=0.019) with T+PDE5i. There was no statistically significant effect for high sensitive women. Another publication from the same study as Poels (2013), Van Rooji (2013) investigates if treatment with a single dosage of T+5-HT(1A)ra will produce improvement in sexual functioning in women with HSDD as the result of dysfunctional high sexual inhibition and compared it with a placebo and T+PDE5i. In the high inhibition group, the T+5-HT1Ara condition “sexual satisfaction” was statistically significantly higher compared with the placebo condition (p = 0.005). In the low inhibition group, sexual satisfaction” was statistically significantly higher with T+PDE5i compared with T+5HT1Ara (p = 0.007). The third RCT (van Rooji 2015) investigated the possible effects of T+5-HT(1A)ra, and of sublingual testosterone combined PDE5-i on sexual functioning in women with SSRI-induced sexual dysfunction. Only women with relatively long cysteine, adenine, and guanine (CAG) repeats and using relatively low SSRI doses (n=8) reported statistically significantly more sexual satisfaction with T+PDE5i compared to placebo (p=0.002). The interaction between drug (placebo versus T+5-HT1Ara) and the two groups (SSRI dose and CAG repeat length) was statistically significant (p=0.002).

Overall Level of Evidence: Low, downgraded due to design limitations and inconsistency across included studies

6. Harms of Testosterone Compared to Other Treatments: There was one systematic review and four RCTs that compared the benefits of testosterone with other pharmacologic treatment. The systematic review (Elraiyah 2014), evaluated benefits and

harms of systemic testosterone in postmenopausal women with normal adrenal function. The analysis found that patients in the testosterone containing regimen had an increased risk of developing acne of 7.0% vs 4.7% for the testosterone free group (RR, 1.62; 95% CI, 1.28 to 2.06, p<0.001) and hirsutism of 10.7% vs 6.6% (RR, 1.45; 95% CI, p=0.011). The first RCT Melisko (2017) reported no major adverse events. Minor adverse events occurring in more than 2% of participants included vaginal discharge (4 vaginal ring, 2 IVT [8% overall]), facial hair growth (1 vaginal ring, 5 IVT [8% overall]), vaginal or vulvar itching and/or irritation (4 vaginal ring [5% overall]), vaginal odor (3 IVT [4%overall]), and urinary tract or yeast infection (1 vaginal ring, 3 IVT [5%overall]). Treatment related adverse events for Poels (2013) included flushing (23% for T+PDE5i and 3.7% for placebo), headache (15.9% for T+PDE5i and 2.4% for placebo), lightheadedness (0.9% for T+PDE5i and 0.6% for placebo), and dizziness (1.1% for T+PDE5i and 0.2% for placebo). van Rooji (2013) found that the T+5-HT(1A)ra group experienced more lightheadedness and dizziness than the placebo group (10.3% vs 0.6% and 11.3% vs 0.2%, respectively). The 2015 van Rooji study did not report on adverse events.

Overall Level of Evidence: Very Low, downgraded due to imprecision and inconsistency across the included studies.

PICO Question: In treatment of low libido in women, what are the benefits (and harms, if any) of Flibanserin?						Lower Quality Rating if:
Outcome: Benefits (monthly sexual desire intensity, increased number of SSEs per month, increased desire on the FSFI)						☐ Studies inconsistent (*wide variation of treatment effect across studies, populations, interventions, or outcomes varied)*
Author/Date	*Purpose of Study*	*Study Design & Methods*	*Sample*	*Outcomes*	*Design Limitations*	☐ Studies are indirect *(PICO question is quite different from the available evidence in regard to population, intervention, comparison, or outcome)*
Total # of Studies: 2 # of Systematic Reviews: 2						☐ Studies are imprecise (*When studies include few patients and few*
Jaspers, L., et al. (2016). *JAMA Internal Medicine*	To conduct a systematic review and meta-analysis of randomized clinical trials assessing efficacy and safety of flibanserin for the treatment of HSDD in women	Systematic review with meta-analysis	8 studies (incl 4 unpublished studies), 5914 women	**Pooled mean differences for SSE change from baseline were 0.49** (95% CI, 0.32-0.67) between 100-mg flibanserin and placebo, **1.63 (95% CI, 0.45-2.82) for eDiary desire, and 0.27 (95% CI, 0.17-0.38) for FSFI desire.**	Study Limitations = ☐ None ***Systematic Review*** ☐ Review did not address focused clinical question ☐ Search was not detailed or exhaustive ☒ Quality of the studies was not appraised or studies were of low quality ☐ Methods and/or results were inconsistent across studies	
Gao, Z., et al. (2015). *Journal of*	To assess the efficacy and safety of	Systematic review	Four publications involving a total of	The comparison of flibanserin with placebo, primary efficacy endpoints: **satisfying sexual events (the**	Study Limitations = ☒ None ***Systematic Review*** ☐ Review did not address	

Sexual Medicine	flibanserin in women with HSDD.		3,414 patients	**standardized mean difference [SMD] = 0.59,** 95% confidence interval [CI] = 0.37–0.80, P < 0.00001); s**exual desire score (the SMD = 1.91,** 95% CI = 0.21 to 3.60, P = 0.03) and **Female Sexual Function Index (FSFI) desire domain score (the SMD = 0.32**, 95% CI = 0.19–0.46, P < 0.00001)	focused clinical question ☐ Search was not detailed or exhaustive ☐ Quality of the studies was not appraised or studies were of low quality ☐ Methods and/or results were inconsistent across studies	*events and thus have wide confidence intervals and the results are uncertain)* ☐ Publication Bias *(e.g. pharmaceutical company sponsors study on effectiveness of drug, only small, positive studies found)* <u>Increase Quality Rating if:</u> ☐ Large Effect ☐ Dose-response gradient ☐ Plausible confounders or other biases increase certainty of effect Quality (certainty) of evidence for studies as a whole: ☐ High ☒ Moderate ☐ Low ☐ Very Low

The GRADE criteria were used to evaluate the quality of evidence presented in research articles reviewed during the development of this guideline. For more detailed information, see Appendix A.

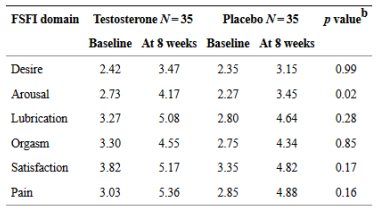
PICO Question: In treatment of low libido in women, what are the benefits (and harms, if any) of Flibanserin						<u>Lower Quality Rating if:</u> ☐ Studies inconsistent (*wide variation of treatment effect across studies, populations, interventions, or*)
Outcome: Harms (Adverse Events i.e. dizziness, somnolence, nausea, and fatigue)						
Author/Date	*Purpose of Study*	*Study Design & Methods*	*Sample*	*Outcomes*	*Design Limitations*	
Total # of Studies: 2 # of Systematic Reviews: 2						
Jaspers, L., et al. (2016). *JAMA*	To conduct a systematic	Systematic review with meta-	8 studies (incl 4 unpublished studies),	**The risk for any AEs, which also included non– drug-related AEs**	Study Limitations = ☐ None	

Internal Medicine	review and meta-analysis of randomized clinical trials assessing efficacy and safety of flibanserin for the treatment of HSDD in women	analysis	5914 women	**such as common cold, was 1.29** (95% CI, 1.15-1.45) times higher for flibanserin than for placebo. **The risk for dizziness was 4.00 (95% CI, 2.56-6.27) times higher with flibanserin** than with placebo; **for somnolence,3.97 (95% CI, 3.01-5.24) times higher with flibanserin; for nausea, 2.35 (95%CI, 1.85-2.98) times higher with flibanserin**; and **for fatigue, 1.64 (95% CI, 1.27-2.13) times higher with flibanserin**. **The overall risk ratio for the 4 most common AEs was 2.86 (95%CI, 2.32-3.52). The absolute number of serious AEs was small, and the risk ratio did not differ between flibanserin and placebo users (1.48[95%CI,0.91-2.41)**	***Systematic Review*** ☐ Review did not address focused clinical question ☐ Search was not detailed or exhaustive ☒ Quality of the studies was not appraised or studies were of low quality ☐ Methods and/or results were inconsistent across studies	*outcomes varied)* ☐ Studies are indirect *(PICO question is quite different from the available evidence in regard to population, intervention, comparison, or outcome)* ☐ Studies are imprecise (*When studies include few patients and few events and thus have wide confidence intervals and the results are uncertain)* ☐ Publication Bias *(e.g. pharmaceutical company sponsors study on effectiveness of drug, only small, positive studies found)* <u>Increase Quality Rating if:</u> ☐ Large Effect ☐ Dose-response gradient ☐ Plausible confounders or other biases increase certainty of effect Quality (certainty) of evidence for studies as a whole:
Gao, Z., et al. (2015). *Journal of Sexual Medicine*	To assess the efficacy and safety of flibanserin in women with HSDD.	Systematic review	Four publications involving a total of 3,414 patients	**The proportion of women who experienced an AE odds ratio [OR] = 1.54** (95% CI = 1.34–1.76, P < 0.00001). ***Nervous system disorders* OR = 2.58** (95% CI = 2.10 to 3.18, P < 0.00001) ***Fatigue* OR = 1.71** (95% CI = 1.20–2.43, P = 0.003)	Study Limitations = ☒ None ***Systematic Review*** ☐ Review did not address focused clinical question ☐ Search was not detailed or exhaustive ☐ Quality of the studies was not appraised or studies were of low quality ☐ Methods and/or results were inconsistent across studies	

						☐ High ☒ Moderate ☐ Low ☐ Very Low

The GRADE criteria were used to evaluate the quality of evidence presented in research articles reviewed during the development of this guideline. For more detailed information, see Appendix A.

PICO Question: In treatment of low libido in women, what are the benefits of Testosterone compared to placebo? **Outcome:** Benefits (monthly sexual desire intensity, increased number of SSEs per month, increased desire on the FSFI)						Lower Quality Rating if:
Author/Date	*Purpose of Study*	*Study Design & Methods*	*Sample*	*Outcomes*	*Design Limitations*	
Total # of Studies: 5 # of Systematic Reviews: 2 # of RCTs: 3						
Achilli, C., et al. (2017). *Fertility & Sterility*	To systematically review and summarize the existing evidence related to the efficacy and safety of transdermal T in postmenopausal women for the treatment of hypoactive sexual desire disorder (HSDD).	Systematic reviews and meta-analysis.	Seven randomized controlled trials enrolled 3,035 participants; 1,350 women were randomized to treatment with T patch, and 1,379 women were randomized to placebo.	**Primary Outcome Measure SSE:** Five studies reported on the MD change in SSE. **Pooling the results of these studies showed that the T-group had significantly more SSE compared with the placebo group MD 0.92** (95% CI, 0.65, 1.19; P<.00001). **Secondary Outcome Measure PFSF domains desire:** Six studies reported mean change in sexual desire experienced by women. **Pooling the results of these studies showed that the T group had experienced significantly more desire compared with the placebo group MD 6.09** (95% CI, 4.51, 7.68; P<.00001)	Study Limitations = ☒ None ***Systematic Review*** ☐ Review did not address focused clinical question ☐ Search was not detailed or exhaustive ☐ Quality of the studies was not appraised or studies were of low quality ☐ Methods and/or results were inconsistent across studies	☐ Studies inconsistent (*wide variation of treatment effect across studies, populations, interventions, or outcomes varied)* ☐ Studies are indirect *(PICO question is quite different from the available evidence in regard to population, intervention, comparison, or outcome)* ☐ Studies are imprecise (*When studies include few patients and few events and thus have wide confidence intervals and the results are uncertain)* ☐ Publication Bias *(e.g. pharmaceutical company sponsors study on effectiveness of drug, only small,*
Fooladi, E., et al. (2014). *Journal of Sexual Medicine*	To investigate the efficacy of transdermal testosterone (TT) as a treatment for SSRI/SNRI-emergent loss	Double-blind, randomized, placebo-controlled study. Women, on a stable dose of SSRI or SNRI with treatment-emergent loss of libido were randomly allocated to treatment with a TT patch delivering 300 mcg of testosterone/day or an	Forty-four women, aged 35-55 years	**The primary outcome measure for the trial was the change in sexual function measured by the Sabbatsberg Sexual Self-rating Scale (SSS), completed at baseline and week 12:** The change from baseline in the total SSS score **did not differ between the two treatment**	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome)	

	of libido.	identical placebo patch (Pl) for 12 weeks.		**groups** at 12 weeks (P=.10) **SSEs:** significantly greater for the testosterone group than for the placebo group at 12 weeks (**an increase of 2.3 events vs. 0.1,** P = 0.02)	☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	*positive studies found)* Increase Quality Rating if: ☐ Large Effect ☐ Dose-response gradient ☐ Plausible confounders or other biases increase certainty of effect Quality (certainty) of evidence for studies as a whole: ☐ High ☒ Moderate ☐ Low ☐ Very Low
Labrie, F., et al. (2014). *Journal of Sexual Medicine*	Investigate the influence of moderate/severe pain at sexual activity (dyspareunia) (MSD) at baseline on female sexual dysfunction (FSD) following prasterone administration.	RCT. The effect of daily administration of prasterone (0, 3.25mg, 6.5mg or 13mg) for 12 weeks on FSD in women with or without MSD at baseline was evaluated.	215 postmenopausal women	Comparable benefits were observed in women not having MSD (n = 56) vs. those having MSD (n = 159). **The benefits over placebo in prasterone-treated women for desire is improved at week 12 by 22% (P = 0.016), 51% (P = 0.0047), 31% (P = 0.2845) and 48% (P = 0.0072) in the placebo, 0.25%, 0.5% and 1.0% prasterone groups, respectively.**	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	
Tungmunsakulchai, R., et al. (2015). *BMC Women's Health*	To evaluate the effectiveness of testosterone undecanoate on sexual function in postmenopausal women utilizing the standardized questionnaire FSFI score	RCT Participants were randomly assigned to 8-week treatment with either oral testosterone undecanoate 40 mg or placebo twice weekly with daily oral estrogen. The FSFI scores before and after treatment were compared to assess any improvement of sexual function.	70 post menopausal women	The baseline characteristics and baseline FSFI scores were comparable between both groups. After 8 weeks of treatment, the FSFI scores significantly improved in both groups when compared to the baseline but the **FSFI scores from the testosterone group were significantly higher than in the placebo group post-treatment (28.6 ± 3.6, 25.3 ± 6.7, respectively, p = 0.04).** Effects of testosterone and placebo on various FSFI domains 	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	

FSFI domain	Testosterone $N=35$		Placebo $N=35$		p value[b]
	Baseline	At 8 weeks	Baseline	At 8 weeks	
Desire	2.42	3.47	2.35	3.15	0.99
Arousal	2.73	4.17	2.27	3.45	0.02
Lubrication	3.27	5.08	2.80	4.64	0.28
Orgasm	3.30	4.55	2.75	4.34	0.85
Satisfaction	3.82	5.17	3.35	4.82	0.17
Pain	3.03	5.36	2.85	4.88	0.16

The GRADE criteria were used to evaluate the quality of evidence presented in research articles reviewed during the development of this guideline. For more detailed information, see Appendix A.

PICO Question: In treatment of low libido in women, what are the harms of Testosterone compared to placebo?						Lower Quality Rating if: ☒ Studies inconsistent (*wide variation of treatment effect across studies, populations, interventions, or outcomes varied)* ☐ Studies are indirect *(PICO question is quite different from the available evidence in regard to population, intervention, comparison, or outcome)* ☐ Studies are imprecise (*When studies include few patients and few events and thus have wide confidence intervals and the results are uncertain)* ☐ Publication Bias *(e.g. pharmaceutical company sponsors study on effectiveness of drug, only small, positive studies found)* Increase Quality Rating if: ☐ Large Effect
Outcome: Harms (Adverse Events [AEs], acne, hair growth, facial hair, alopecia)						
Author/Date	*Purpose of Study*	*Study Design & Methods*	*Sample*	*Outcomes*	*Design Limitations*	
Total # of Studies: 5 # of Systematic Reviews: 2 # of RCTs: 3						
Achilli, C., et al. (2017). *Fertility & Sterility*	To systematically review and summarize the existing evidence related to the efficacy and safety of transdermal T in postmenopausal women for the treatment of hypoactive sexual desire disorder (HSDD).	Systematic reviews and meta-analysis.	Seven randomized controlled trials enrolled 3,035 participants; 1,350 women were randomized to treatment with T patch, and 1,379 women were randomized to placebo.	**Pooling the results from all seven studies showed that there was no significant difference in total adverse events or** (RR, 1.01; 95% CI, 0.97, 1.05; P=.77;); **severe adverse event**s (RR, 1.02; 95% CI, 0.62, 1.68; P=.94) Androgenic Adverse Events: **T group had significantly more total androgenic adverse events compared with the placebo group RR, 1.37 (95% CI, 1.12- 1.69; P=.002).** Acne: **T group had significantly more acne compared with the placebo group RR, 1.41(95% CI, 1.05-1.88; P=.02).** Hair Growth**: T group had significantly more hair growth compared with the placebo group RR, 1.56 (95% CI, 1.17- 2.09; P=.003).** Facial Hair, alopecia, voice deepening: **No statistical difference**	Study Limitations = ☒ None ***Systematic Review*** ☐ Review did not address focused clinical question ☐ Search was not detailed or exhaustive ☐ Quality of the studies was not appraised or studies were of low quality ☐ Methods and/or results were inconsistent across studies	
Fooladi, E., et al. (2014). *Journal of Sexual Medicine*	To investigate the efficacy of transdermal testosterone (TT) as a treatment for SSRI/SNRI-emergent loss	Double-blind, randomized, placebo-controlled study. Women, on a stable dose of SSRI or SNRI with treatment-emergent loss of libido were randomly allocated to treatment with a TT patch delivering 300 mcg of	Forty-four women, aged 35-55 years	**There were no androgenic adverse events. There were no clinically relevant changes for any vital sign measurements.**	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of	

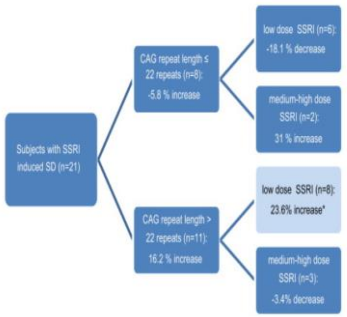
	of libido.	testosterone/day or an identical placebo patch (Pl) for 12 weeks.			measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	☐ Dose-response gradient ☐ Plausible confounders or other biases increase certainty of effect Quality (certainty) of evidence for studies as a whole: ☐ High ☒ Moderate ☐ Low ☐ Very Low
Labrie, F., et al. (2014). *Journal of Sexual Medicine*	Investigate the influence of moderate/severe pain at sexual activity (dyspareunia) (MSD) at baseline on female sexual dysfunction (FSD) following prasterone administration.	RCT. The effect of daily administration of prasterone (0, 3.25mg, 6.5mg or 13mg) for 12 weeks on FSD in women with or without MSD at baseline was evaluated.	215 postmenopausal women	**Did not report on adverse events**	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	
Tungmunsakulchai, R., et al. (2015). *BMC Women's Health*	To evaluate the effectiveness of testosterone undecanoate on sexual function in postmenopausal women utilizing the standardized questionnaire FSFI score	RCT Participants were randomly assigned to 8-week treatment with either oral testosterone undecanoate 40 mg or placebo twice weekly with daily oral estrogen. The FSFI scores before and after treatment were compared to assess any improvement of sexual function.	70 post-menopausal women	**Table 4** Adverse effects Symptoms | Testosterone N = 35 | Placebo N = 35 | *p* value Acne N (%) | 6 (17.6) | 5 (14.2) | 0.10 Hirsutism N (%) | 3 (8.8) | 3 (8.5) | 0.70	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	

The GRADE criteria were used to evaluate the quality of evidence presented in research articles reviewed during the development of this guideline. For more detailed information, see Appendix A.

PICO Question: In treatment of low libido in women, what are the benefits of Testosterone compared to another treatment?						Lower Quality Rating if: ☒ Studies inconsistent (*wide*
Outcome: Benefits (monthly sexual desire intensity, increased number of SSEs per month, increased desire on the FSFI)						
Author/Date	*Purpose of Study*	*Study Design & Methods*	*Sample*	*Outcomes*	*Design Limitations*	

Total # of Studies: 5 # of Systematic Reviews 1 # of RCTs: 4						*variation of treatment effect across studies, populations, interventions, or outcomes varied)* ☐ Studies are indirect *(PICO question is quite different from the available evidence in regard to population, intervention, comparison, or outcome)* ☐ Studies are imprecise (*When studies include few patients and few events and thus have wide confidence intervals and the results are uncertain)* ☐ Publication Bias *(e.g. pharmaceutical company sponsors study on effectiveness of drug, only small, positive studies found)* Increase Quality Rating if: ☐ Large Effect ☐ Dose-response gradient ☐ Plausible confounders or other biases increase certainty of effect
Elraiyah et al (2014). *The Journal of Clinical Endocrinology & Metabolism*	To summarize the best available evidence regarding the benefits and harms of systemic testosterone in postmenopausal women with normal adrenal function.	Systematic Review with meta-analysis	17 studies, 3288 patients	Compared T-free regimen (TFR), **the T-containing regimen (TCR) was associated with statistically significant improvement in number of satisfying sexual episodes** (WMD, 1.20; 95%CI, 0.88 to 1.51), **and interest in sex** (SMD, 0.35; 95% CI, 0.19 to 0.52). **The quality of encounters was also found to be improved by TCR, including improvements in orgasm** (SMD, 0.20; 95% CI, 0.09 to 0.31), **arousal** (SMD, 0.25; 95% CI, 0.09 to 0.40), and **enjoyment of sex** (SMD, 0.31; 95% CI, 0.12 to 0.51).	Study Limitations = ☒ None ***Systematic Review*** ☐ Review did not address focused clinical question ☐ Search was not detailed or exhaustive ☐ Quality of the studies was not appraised or studies were of low quality ☐ Methods and/or results were inconsistent across studies	
Melisko, M. E., et al. (2017). *JAMA Oncology*	To evaluate safety of intravaginal testosterone cream (IVT) or an estradiol-releasing vaginal ring (7.5 mug/d) in patients with early-stage breast cancer (BC) receiving Aromatase inhibitors (AI).	RCT Postmenopausal (PM) women with hormone receptor (HR)-positive stage I to III BC taking AIs with self-reported vaginal dryness, dyspareunia, or decreased libido were randomized to 12 weeks of IVT or an estradiol vaginal ring. Intervention was considered unsafe if more than 25% of patients had persistent elevation in estradiol (E2), defined as E2 greater than 10 pg/mL (to convert to pmol/L, multiply by 3.671) and at least 10 pg/mL above baseline after treatment initiation on 2 consecutive tests at least 2 weeks apart.	76 women	**For patients using the vaginal ring, mean (SD) sexual interest improved from 1.2 (0.9) at baseline to 0.9 (0.7) at week 12 (P = .021) and for IVT patients, from 1.4 (0.8) to 1.0 (0.6) (P = .02). Mean (SD) sexual dysfunction improved from 2.9 (1.1) at baseline to 2.0 (1.1) at week 12 in vaginal ring patients and from 2.9 (0.8) to 1.9 (1.1) in IVT patients (both P < .001). Mean (SD) single-item Cancer Rehabilitation Evaluation System sexual satisfaction score showed more improvement in vaginal ring patients, increasing from 2.5 (1.6) at baseline to 4.0 (1.5) at week 12 (P = .004) than in IVT patients, changing from 3.2 (1.6) to 4.0 (1.5) (P = .14**	Study Limitations = ☐ None ***RCTS*** ☒ Lack of blinding ☒ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	
Poels, S., et al. (2013). *Journal of Sexual Medicine*	To assess the hypothesis that treatment with on-demand use of T+PDE5i improves	RCT. women with HSDD underwent three medication treatment regimes (placebo, T+PDE5i, and T with a serotonin receptor agonist, which lasted 4 weeks each. In a participant-controlled ambulatory psychophysiological experiment at home (the first week of each	56 women with HSDD	*Low Sensitive Women:* **Sexual Satisfaction During Sexual Events. Treatment with T+PDE5i produced a statistically significant** [F(1,23) = 6.34, P = 0.019] increase in sexual satisfaction (M = 3.36, SE = 0.16) as compared with placebo (M = 2.96, SE = 0.20)	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of	

	sexual functioning, particularly in women who suffer from Hypoactive Sexual Desire Disorder (HSDD) as the result of a relative insensitivity for sexual cues.	drug treatment), physiological and subjective indices of sexual functioning were measured. In a bedroom experiment (the subsequent 3 weeks), sexual functioning was evaluated following each sexual event after the self-administration of study medication. Subjective evaluation of sexual functioning was also measured by weekly and monthly reports.		*High Sensitive Women:* **The results of the high sensitive subgroup showed no statistically significant drug effects** in the bedroom experiment.	measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	Quality (certainty) of evidence for studies as a whole: ☐ High ☐ Moderate ☒ Low ☐ Very Low
van Rooij, K., et al. (2013). *Journal of Sexual Medicine*	To investigate if treatment with a single dosage of T+5-HT(1A)ra will produce improvement in sexual functioning in women with Hypoactive Sexual Desire Disorder (HSDD) as the result of dysfunctional high sexual inhibition.	RCT Participants underwent three different medication treatments: (i) placebo: placebo for testosterone (cyclodextrin solution without testosterone) and placebo for the PDE5i (PDE5i = sildenafil) and 5-HT1Ara (5-HT1Ara = buspirone) (powder-filled gelatin capsule without sildenafil/buspirone); (ii) T+PDE5i: the combination of testosterone (0.5 mg) sublingually with cyclodextrin as carrier and sildenafil (50 mg, hidden in a powder-filled gelatin capsule); (iii) T+5-HT1Ara; the combination of testosterone (0.5 mg) sublingually with cyclodextrin as carrier and buspirone (10 mg, hidden in a powder-filled gelatin capsule).	56 women with HSDD	*High Inhibition group:* **In the T+5-HT1A ra condition "sexual satisfaction" was statistically significant higher (M = 2.98, SE = 0.13) compared with the placebo condition (M = 2.51, SE = 0.14) [F(1,25) = 9.51, P = 0.005].** *Low Inhibition Group:* **Low inhibition group (N = 26) revealed that treatment, "sexual satisfaction" was statistically significant higher during T+PDE5i (M = 3.82, SE = 0.11) compared with the T+5HT1A ra condition (M = 3.37, SE = 0.18) [F(1,22) = 8.84, P = 0.007].**	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	
van Rooij, K., et al. (2015). *European Journal of Pharmacology*	To investigate the possible effects of sublingual testosterone combined with a serotonin (5-HT)1A receptor agonist, and	RCT Women underwent three different medication regimes: (i) placebo: placebo for testosterone (cyclodextrin solution without testosterone) and placebo for the PDE5 inhibitor (PDE5i=sildenafil)/5-HT1A receptor agonist (5-HT1Ara=buspirone) (powder-filled gelatine capsule without sildenafil/buspirone); (ii)	21 women	**Only women with relatively long cysteine, adenine, and guanine (CAG) repeats and using relatively low SSRI doses (n=8) reported statistically significant more sexual satisfaction in the T+PDE5i (M=3.53, SE=0.26) condition compared to placebo (M=2.92, SE=0.21 [F(1,7)=−4.67, P=0.002]).** **The interaction between drug (placebo versus T+5-HT1Ara) and**	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U	

	of sublingual testosterone combined with a phosphodiesterase type 5 inhibitor (PDE5-i) on sexual functioning in women with SSRI-induced sexual dysfunction.	T+PDE5i: the combination of testosterone (0.5 mg) sublingually with cyclodextrin as carrier and sildenafil (50 mg, hidden in a powder-filled gelatine capsule); (iii) T+5-HT1Ara; the combination of testosterone (0.5 mg) sublingually with cyclodextrin as carrier and buspirone (10 mg, hidden in a powder-filled gelatin capsule). Each medication regime lasted four weeks and the order of the three medication regimes was randomized		**the two groups (SSRI dose and CAG repeat length) was statistically significant F(1,15)=13.37, P=0.002.** 	☐ Difference in important prognostic factors at baseline	

The GRADE criteria were used to evaluate the quality of evidence presented in research articles reviewed during the development of this guideline. For more detailed information, see Appendix A.

PICO Question: In treatment of low libido in women, what are the harms of Testosterone compared to another treatment?						Lower Quality Rating if: ☒ Studies inconsistent (*wide variation of treatment effect across studies, populations, interventions, or outcomes varied)* ☐ Studies are indirect *(PICO question is quite different from the available evidence in regard to population, intervention, comparison, or outcome)* ☒ Studies are
Outcome: Harms (Adverse Events [AEs], acne, hair growth, facial hair, alopecia)						
Author/Date	*Purpose of Study*	*Study Design & Methods*	*Sample*	*Outcomes*	*Design Limitations*	
Total # of Studies: 5# of Systematic Reviews: 1 # of RCTs: 4						
Elraiyah et al (2014). *The Journal of Clinical Endocrinology & Metabolism*	To summarize the best available evidence regarding the benefits and harms of systemic testosterone in postmenopausal women with normal	Systematic Review with meta-analysis	17 studies, 3288 patients	Compared T-free regimen (TFR), **the T-containing regimen (TCR) was associated with statistically significant improvement in number of satisfying sexual episodes** (WMD, 1.20; 95%CI, 0.88 to 1.51), **and interest in sex** (SMD, 0.35; 95% CI, 0.19 to 0.52). **The quality of encounters was also found to be improved by TCR, including improvements in orgasm** (SMD, 0.20; 95% CI, 0.09 to 0.31), **arousal** (SMD, 0.25; 95% CI, 0.09 to 0.40), and **enjoyment of sex** (SMD, 0.31; 95% CI,	Study Limitations = ☒ None ***Systematic Review*** ☐ Review did not address focused clinical question ☐ Search was not detailed or exhaustive ☐ Quality of the studies was not appraised or studies were of low quality ☐ Methods and/or results were inconsistent across studies	

	adrenal function.			0.12 to 0.51).		imprecise (*When studies include few patients and few events and thus have wide confidence intervals and the results are uncertain)* ☐ Publication Bias *(e.g. pharmaceutical company sponsors study on effectiveness of drug, only small, positive studies found)* Increase Quality Rating if: ☐ Large Effect ☐ Dose-response gradient ☐ Plausible confounders or other biases increase certainty of effect Quality (certainty) of evidence for studies as a whole: ☐ High ☐ Moderate ☐ Low ☒ Very Low
Melisko, M. E., et al. (2017). *JAMA Oncology*	To evaluate safety of intravaginal testosterone cream (IVT) or an estradiol-releasing vaginal ring (7.5 mug/d) in patients with early-stage breast cancer (BC) receiving Aromatase inhibitors (AI).	RCT Postmenopausal (PM) women with hormone receptor (HR)-positive stage I to III BC taking AIs with self-reported vaginal dryness, dyspareunia, or decreased libido were randomized to 12 weeks of IVT or an estradiol vaginal ring. Intervention was considered unsafe if more than 25% of patients had persistent elevation in estradiol (E2), defined as E2 greater than 10 pg/mL (to convert to pmol/L, multiply by 3.671) and at least 10 pg/mL above baseline after treatment initiation on 2 consecutive tests at least 2 weeks apart.	76 women	All treatment-related adverse events were grade 1 or 2. **Events occurring in more than 2% of participants included vaginal discharge (4 vaginal ring, 2 IVT [8% overall]), facial hair growth (1 vaginal ring, 5 IVT [8% overall]), vaginal or vulvar itching and/or irritation (4 vaginal ring [5% overall]), vaginal odor (3 IVT [4%overall]), and urinary tract or yeast infection (1 vaginal ring, 3IVT[5%overall]).Two patients discontinued therapy for toxic effects (vaginal odor or irritation in vaginal ring and vulvar irritation in IVT).**	Study Limitations = ☐ None ***RCTS*** ☒ Lack of blinding ☒ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	
Poels, S., et al. (2013). *Journal of Sexual Medicine*	To assess the hypothesis that treatment with on-demand use of T+PDE5i improves sexual functioning, particularly in women who suffer from Hypoactive Sexual Desire Disorder (HSDD) as the result of a relative insensitivity for sexual cues.	In a randomized, double-blind, placebo-controlled, crossover design. women with HSDD underwent three medication treatment regimes (placebo, T+PDE5i, and T with a serotonin receptor agonist, which lasted 4 weeks each. In a participant-controlled ambulatory psychophysiological experiment at home (the first week of each drug treatment), physiological and subjective indices of sexual functioning were measured. In a bedroom experiment (the subsequent 3 weeks), sexual functioning was evaluated following each sexual event after the self-administration of study medication. Subjective evaluation of sexual functioning was also measured by weekly and monthly reports.	56 women with HSDD	**Table 2** Treatment related adverse events |	T+PDE5i (%)*	Placebo (%)†	Total (%)‡
---	---	---	---
Flushing	23.0	3.7	9.6
Headache	15.9	2.4	7.4
Lightheadedness	0.9	0.6	3.9
Dizziness	1.1	0.2	4.2	 *Percentage T+PDE5i medication = AE T+PDE5i/552 units †Percentage placebo medication = AE placebo/542 units ‡Percentage total medication = AE total/1,636 units T = testosterone; PDE5i = phosphodiesterase type 5 inhibitor; AE = adverse events	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	

van Rooij, K., et al. (2013). *Journal of Sexual Medicine*	To investigate if treatment with a single dosage of T+5-HT(1A)ra will produce improvement in sexual functioning in women with Hypoactive Sexual Desire Disorder (HSDD) as the result of dysfunctional high sexual inhibition.	RCT Participants underwent three different medication treatments: (i) placebo: placebo for testosterone (cyclodextrin solution without testosterone) and placebo for the PDE5i (PDE5i = sildenafil) and 5-HT1Ara (5-HT1Ara = buspirone) (powder-filled gelatin capsule without sildenafil/buspirone); (ii) T+PDE5i: the combination of testosterone (0.5 mg) sublingually with cyclodextrin as carrier and sildenafil (50 mg, hidden in a powder-filled gelatin capsule); (iii) T+5-HT1Ara; the combination of testosterone (0.5 mg) sublingually with cyclodextrin as carrier and buspirone (10 mg, hidden in a powder-filled gelatin capsule).	56 women	**Table 2** Treatment-related adverse events T+PDE5i (%)*; T+5-HT$_{1A}$ra (%)†; Placebo (%)‡; Total (%)§ Flushing 23.0 1.9 3.7 9.6 Headache 15.9 3.7 2.4 7.4 Lightheadedness 0.9 10.3 0.6 3.9 Dizziness 1.1 11.3 0.2 4.2 *Percentage T+PDE5i medication = AE T+PDE5i/552 units †Percentage T+5-HT$_{1A}$ra medication = AE 5-HT$_{1A}$ra/542 units ‡Percentage placebo medication = AE placebo/542 units §Percentage total medication = AE total/1,636 units T = testosterone; PDE5i = phosphodiesterase type 5 inhibitor; 5-HT$_{1A}$ra = 5-HT$_{1A}$ receptor agonist; AE = adverse events	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	
van Rooij, K., et al. (2015). *European Journal of Pharmacology*	To investigate the possible effects of sublingual testosterone combined with a serotonin (5-HT)1A receptor agonist, and of sublingual testosterone combined with a phosphodiesterase type 5 inhibitor (PDE5-i) on sexual functioning in women	RCT Women underwent three different medication regimes: (i) placebo: placebo for testosterone (cyclodextrin solution without testosterone) and placebo for the PDE5 inhibitor (PDE5i=sildenafil)/5-HT1A receptor agonist (5-HT1Ara=buspirone) (powder-filled gelatine capsule without sildenafil/buspirone); (ii) T+PDE5i: the combination of testosterone (0.5 mg) sublingually with cyclodextrin as carrier and sildenafil (50 mg, hidden in a powder-filled gelatine capsule); (iii) T+5-HT1Ara; the combination of testosterone (0.5 mg) sublingually with cyclodextrin as carrier and buspirone (10 mg, hidden in a powder-filled gelatin capsule). Each medication regime lasted four weeks and the order of the three medication regimes was randomized	21 women	**Did not report on adverse events**	Study Limitations = ☒ None ***RCTS*** ☐ Lack of blinding ☐ Lack of allocation concealment ☐ Stopped early for benefit ☐ Incorrect analysis of ITT ☐ Selective reporting of measures (e.g., no effect outcome) ☐ Large losses to F/U ☐ Difference in important prognostic factors at baseline	

	with SSRI-induced sexual dysfunction.					

The GRADE criteria were used to evaluate the quality of evidence presented in research articles reviewed during the development of this guideline. For more detailed information, see Appendix A.

External Guidelines

Two international societies and institutes offer recommendations related to testosterone use for the treatment of low libido in menopausal women. The recommendations, and the quality rating for each guideline are noted below.

Guideline Ratings

Guideline Issuer	Endocrine Society 2014	National Institute for Health and Care Excellence 2016
1. Transparency	A	A
2. Conflict of interest	A	A
3. Development group	B	A
4. Systematic Review	A	A
5. Supporting evidence	A	A
6. Recommendations	A	A
7. External Review	NR	A
8. Currency and updates	B	A

The two published clinical guidelines were evaluated for this review using the **University of Pennsylvania's Center for Evidence-Based Practice Trustworthy Guideline rating scale.** The scale is based on the Institute of Medicine's "Standards for Developing Trustworthy Clinical Practice Guidelines" (IOM), as well as a review of the AGREE Enterprise and Guidelines International Network domains.

See appendix C for full description of the Trustworthy Guideline grading system.

Guideline Recommendations:

In 2014, the **Endocrine Society** released an updated guideline for Androgen Therapy in Women:

Recommendation 1: We suggest a 3- to 6-month trial of a dose of T for postmenopausal women who request therapy for properly diagnosed HSDD and in whom therapy is not contraindicated resulting in a midnormal premenopausal value in a reference assay to avoid pharmacological T administration. *Weak Recommendation, Low Quality Evidence.*

Recommendation 2: If T therapy is prescribed, we suggest measuring T levels at baseline and after 3–6 weeks of initial treatment to assess patient overuse. *Weak Recommendation, Low Quality Evidence.*

Recommendation 3: In cases of ongoing T therapy, we suggest reviewing T levels every 6 months to monitor for excessive use and signs of androgen excess. *Weak Recommendation, Low Quality Evidence.*

Recommendation 4: We suggest cessation of T therapy for women who have not responded to treatment by 6 months. *Weak Recommendation, Low Quality Evidence.*

In 2015, the **National Institute for Health and Care Excellence** in the UK released an updated clinical guideline for Menopause:

Recommendation 1: Consider testosterone supplementation for menopausal women with low sexual desire if HRT alone is not effective. *Weak Recommendation, Very Low Quality Evidence.*

REFERENCES

1. Elraiyah et al. (2014) “The Benefits and Harms of Systemic Testosterone Therapy in Postmenopausal Women With Normal Adrenal Function: A Systematic Review and Meta-analysis.” The Journal of Clinical Endocrinology & Metabolism 99 (10):3543–3550.
2. Vallejos, X. and C. Wu (2017). "Flibanserin." Journal of Pharmacy Practice 30(2): 256-260.
3. Wierman ME, Arlt W, Basson R, Davis SR, Miller KK, Murad MH, Rosner W, Santoro N. Androgen therapy in women: a reappraisal: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2014 Oct;99(10):3489-510.
4. Jaspers, L., et al. (2016). "Efficacy and Safety of Flibanserin for the Treatment of Hypoactive Sexual Desire Disorder in Women: A Systematic Review and Meta-analysis." JAMA Internal Medicine 176(4): 453-462.
5. Gao, Z., et al. (2015). "Efficacy and Safety of Flibanserin in Women with Hypoactive Sexual Desire Disorder: A Systematic Review and Meta-Analysis." Journal of Sexual Medicine 12(11): 2095-2104.
6. Achilli, C., et al. (2017). "Efficacy and safety of transdermal testosterone in postmenopausal women with hypoactive sexual desire disorder: a systematic review and meta-analysis." Fertility & Sterility 107(2): 475-482.e415.
7. Fooladi, E., et al. (2014). "Testosterone improves antidepressant-emergent loss of libido in women: findings from a randomized, double-blind, placebo-controlled trial." Journal of Sexual Medicine 11(3): 831-839.
8. Labrie, F., et al. (2014). "Lack of influence of dyspareunia on the beneficial effect of intravaginal prasterone (dehydroepiandrosterone, DHEA) on sexual dysfunction in postmenopausal women." Journal of Sexual Medicine 11(7): 1766-1785
9. Tungmunsakulchai, R., et al. (2015). "Effectiveness of a low dose testosterone undecanoate to improve sexual function in postmenopausal women." BMC Women's Health 15: 113.
10. Melisko, M. E., et al. (2017). "Vaginal Testosterone Cream vs Estradiol Vaginal Ring for Vaginal Dryness or Decreased Libido in Women Receiving Aromatase Inhibitors for Early-Stage Breast Cancer: A Randomized Clinical Trial." JAMA Oncology 3(3): 313-319.
11. Poels, S., et al. (2013). "Toward personalized sexual medicine (part 2): testosterone combined with a PDE5 inhibitor increases sexual satisfaction in women with HSDD and FSAD, and a low sensitive system for sexual cues." Journal of Sexual Medicine 10(3): 810-823.
12. van Rooij, K., et al. (2013). "Toward personalized sexual medicine (part 3): testosterone combined with a Serotonin1A receptor agonist increases sexual satisfaction in women with HSDD and FSAD, and dysfunctional activation of sexual inhibitory mechanisms." Journal of Sexual Medicine 10(3): 824-837.
13. van Rooij, K., et al. (2015). "Efficacy of testosterone combined with a PDE5 inhibitor and testosterone combined with a serotonin (1A) receptor agonist in women with SSRI-induced sexual dysfunction. A preliminary study." European Journal of Pharmacology 753: 246-251.

14. National Collaborating Centre for Women's and Children's Health. Menopause: diagnosis and management. London (UK): National Institute for Health and Care Excellence (NICE); 2015 Nov 12. 29 p. (NICE guideline; no. 23)

Appendix A. GRADE criteria for rating a body of evidence on an intervention

Developed by the GRADE Working Group

Grades and interpretations:

High: Further research is very unlikely to change our confidence in the estimate of effect. Moderate: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate. Low: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. Very low: Any estimate of effect is very uncertain.

Type of evidence and starting level

Randomized trial–high Observational study–low Any other evidence–very low

Criteria for increasing or decreasing level

Reductions Study quality has serious (–1) or very serious (–2) problems Important inconsistency in evidence (–1) Directness is somewhat (–1) or seriously (–2) uncertain Sparse or imprecise data (–1) Reporting bias highly probable (–1) **Increases** Evidence of association† strong (+1) or very strong (+2) †Strong association defined as significant relative risk (factor of 2) based on consistent evidence from two or more studies with no plausible confounders Very strong association defined as significant relative risk (factor of 5) based on direct evidence with no threats to validity.

Appendix B. Search Strategy

Search strategy included:

Database: Ovid MEDLINE(R) <1946 to September Week 2 2017>

1 exp Libido/ (4758)

2 exp Sexual Dysfunctions, Psychological/ (24675)
3 exp Sexual Dysfunction, Physiological/ (27985)
4 1 or 2 or 3 (34837)
5 exp Androgens/ (92920)
6 (Flibanserin or addyi).mp. (133)
7 5 or 6 (93043)
8 4 and 7 (2551)
9 (libido* or ((sex* or coit* or intercours* or copulat*) adj3 (driv* or desir* or arous* or want* or need* or function* or dysfunction* or initia* or participa*))).mp. (45106)
10 8 and 9 (1757)
11 limit 10 to female (806)
12 ((reduc* or low* or decreas* or hypoactiv* or rais* or increas* or high* or elevat*) adj3 (driv* or desir* or arous* or function* or dysfunction* or want* or need* or function* or initia* or participa*)).mp. (364319)
13 8 and 12 (475)
14 limit 13 to female (235)
15 11 or 14 (810)
16 limit 15 to english language (705)
17 limit 15 to abstracts (560)
18 16 or 17 (745)
19 limit 18 to (meta analysis or systematic reviews) (43)
20 limit 18 to (controlled clinical trial or guideline or randomized controlled trial) (84)
21 limit 18 to (comparative study or evaluation studies) (44)
22 exp Epidemiologic Studies/ (2177321)
23 18 and 22 (82)
24 19 or 20 or 21 or 23 (217)
25 18 not 24 (528)

Filters/limits included articles published in English in the last 10 years.

Appendix C. Trustworthy Guideline rating scale

The University of Pennsylvania's Center for Evidence-Based Practice Trustworthy Guideline rating scale is based on the Institute of Medicine's "Standards for Developing Trustworthy Clinical Practice Guidelines" (IOM), as well as a review of the AGREE Enterprise and Guidelines International Network domains.

The purpose of this scale is to focus on the weaknesses of a guideline that may reduce the trust a clinical user can have in the guideline, and distinguish weaknesses in documentation (e.g. guide-line does not have a documented updating process) from weaknesses in the guidance itself (e.g. recommendations are outdated). Current quality scales like AGREE emphasize documentation. They are important checklists for developers of new guidelines, but are less useful for grading existing guidelines. These scales also are harder for clinicians and other persons who are not methodology experts to apply, and their length discourages their use outside formal technology assessment reports. This new scale is brief, balanced, and easy and consistent to apply.

We do not attempt to convert the results of this assessment into a numeric score. Instead we present a table listing the guidelines and how they are rated on each standard. This facilitates qualitative understanding by the reader, who can see for what areas the guideline base as a whole is weak or strong as well as which guidelines are weaker or stronger.

1. Transparency

A	Guideline development methods are fully disclosed.
B	Guideline development methods are partially disclosed.
C	Guideline development methods are not disclosed.

The grader must refer to any cited methods supplements or other supporting material when evaluating the guideline. Methods should include:
Who wrote the initial draft
How the committee voted on or otherwise approved recommendations
Evidence review, external review and methods used for updating are not addressed in this standard.

2. Conflict of interest

A	Funding of the guideline project is disclosed, disclosures are made for each individual panelist, and financial or other conflicts do not apply to key authors of the guideline or to more than 1 in 10 panel members).
B	Guideline states that there were no conflicts (or fewer than 1 in 10 panel members), but does not disclose funding source.
C	Lead author, senior author, or guideline panel members (at least 1 in 10) have conflict of interest, or guideline project was funded by industry sponsor with no assurance of independence.
NR	Guideline does not report on potential conflict of interests.

For purposes of this checklist, conflicts of interest include employment by, consulting for, or holding stock in companies doing business in fields affected by the guideline, as well as related financial conflicts. This definition should not be considered exclusive. As much as anything,

this is a surrogate marker for thorough reporting, since it may be assumed that guideline projects are funded by the sponsoring organization and many authors think it unnecessary to report a non-conflict.

3. Guideline development group

A	Guideline development group includes 1) methodological experts and clinicians and 2) representatives of multiple specialties.
B	Guideline development group includes one of the above, but not both.
C	Guideline developers all from one specialty or organization, and no methodologists.
NR	Affiliations of guideline developers not reported

The purpose of this standard is to ensure that supporters of competing procedures, or clinicians with no vested interest in utilization of one procedure or another, are involved in development of the guideline. Both AGREE II and IOM call for patient or public involvement: very few guideline panels have done so to date, so this is not necessary for guidelines to be rated A. Involvement of methodologists or HTA specialists in the systematic review is sufficient involvement in the guideline development group for our purposes. In the absence of any description of the guideline group, assume the named authors are the guideline group.

4. Systematic review

A	Guideline includes a systematic review of the evidence or links to a current review.
B	Guideline is based on a review which may or may not meet systematic review criteria.
C	Guideline is not based on a review of the evidence.

In order to qualify as a systematic review, the review must do all of the following:
Describe itself as systematic or report search strategies using multiple databases
Define the scope of the review (including key questions and the applicable population)
Either include quantitative or qualitative synthesis of the data or explain why it is not indicated

Note: this element does not address the quality of the systematic review: simply whether or not it exists. Concerns about quality or bias of the review will be discussed in text, where the analyst will explain whether the weaknesses of the review weaken the validity or reliability of the guideline.
Note: a guideline may be rated B on this domain even if the review on which it is based is not available to us. This potential weakness of the guideline should be discussed in text of the report.

5. Grading the supporting evidence

A	Specific supporting evidence (or lack thereof) for each recommendation is cited and graded
B	Specific supporting evidence (or lack thereof) for each recommendation is cited but the recommendation is not graded.
C	Recommendations are not supported by specific evidence.

To score a B on this domain there should be specific citations to evidence tables or individual references for each relevant recommendation in the guideline, or an indication that no evidence was available. Any standardized grading system is acceptable for purposes of this rating. If a guideline reports that there is no evidence available despite a thorough literature search, it may be scored B on this domain, or even A if evidence for other recommendations is cited and graded.

6. Recommendations

A	Considerations for each recommendation are documented (i.e. benefits and harms of a particular action, and/or strength of the evidence); and recommendations are presented in an actionable form.
B	Either one or the other of the above criteria is met.
C	Neither of the above criteria are met

In order to be actionable, the guideline should specify the specific population to which the guideline applies, the specific intervention in question, and the circumstances under which it should be carried out (or not carried out). The language used in the recommendations should also be consistent with the strength of the recommendation (e.g. directive and active language like “should” or “should not” for strong recommendations, and passive language like “consider” for weak recommendations). A figure or algorithm is considered actionable as long as it is complete enough to incorporate all the applicable patients and interventions. Please see the forthcoming NICE manual (24) for a good discussion of actionability in guidelines.

7. External review

A	Guideline was made available to external groups for review.
B	Guideline was reviewed by members of the sponsoring body only.
C	Guideline was not externally reviewed.
NR	No external review process is described.

8. Updating and currency of guideline

A	Guideline is current and an expiration date or update process is specified.
B	Guideline is current but no expiration date or update process is specified.
C	Guideline is outdated.

A guideline is considered current if it is within the developers’ stated validity period, or if no period or expiration data is stated, the guideline was published in the past three years (NOTE: the specific period may be changed at the analyst’s discretion, based on whether the technology is mature and whether there is a significant amount of recent evidence). A guideline must address new evidence when it is updated. A guideline which is simply re-endorsed by the panel without searching for new evidence must be considered outdated.