

Published in final edited form as:

J Sex Med. 2011 January ; 8(1): 191–201. doi:10.1111/j.1743-6109.2010.02007.x.

Behavior and Symptom Change Among Women Treated with Placebo for Sexual Dysfunction

Andrea Bradford, PhD^{*,†} and Cindy M. Meston, PhD[‡]

^{*}Baylor College of Medicine—Department of Family and Community Medicine, Houston, TX, USA

[†]Houston Center for Quality of Care and Utilization Studies, Houston, TX, USA

[‡]University of Texas at Austin—Department of Psychology, Austin, TX, USA

Abstract

Introduction—In clinical trials of drug treatments for women’s sexual dysfunction, placebo responses have often been substantial. However, little is known about the clinical significance, specificity, predictors, and potential mechanisms of placebo response in sexual dysfunction.

Aim—We aimed to determine the nature and predictors of sexual function outcomes in women treated with placebo for female sexual arousal disorder (FSAD).

Methods—We conducted a secondary analysis of data from the placebo arm of a 12-week, multisite, randomized controlled pharmaceutical trial for FSAD (N = 50). We analyzed the magnitude, domain specificity, and clinical significance of sexual function scores at baseline, 4, 8, and 12 weeks (post-treatment). We examined longitudinal change in sexual function outcomes as a function of several baseline variables (e.g., age, symptom-related distress) and in relation to changes in sexual behavior frequency during the trial.

Main Outcome Measure—Female Sexual Function Index total score.

Results—The magnitude of change at post-treatment was clinically significant in approximately one-third of placebo recipients. Effect sizes were similar across multiple aspects of sexual function. Symptom improvement was strongly related to the frequency of satisfying sexual encounters during treatment. However, the relationship between sexual encounter frequency and outcome varied significantly between participants.

Conclusions—A substantial number of women experienced clinically significant improvement in sexual function during treatment with placebo. Changes in sexual behavior during the trial, more so than participant age or symptom severity at baseline, appeared to be an important determinant of outcome. Contextual and procedural aspects of the clinical trial may have influenced outcomes in the absence of an active drug treatment.

Keywords

Placebo Response; Sexual Dysfunction; Sexual Arousal Disorders; Women

© 2010 International Society for Sexual Medicine

Corresponding Author: Andrea Bradford, PhD, Family and Community Medicine, Baylor College of Medicine, 3701 Kirby Drive Suite 600, Houston, Texas 77098, USA. Tel: 713-487-5073; Fax: 713-748-7359; Andrea.Bradford2@gmail.com.

Conflict of Interest: None.

Introduction

Many clinical trials for women's sexual dysfunctions have yielded substantial clinical responses following administration of placebo [1]. Although efforts have been made to isolate "true" placebo responses from the natural course of illness and other artifacts [2], in practice placebo response is difficult to separate from the experiences of treatment-seeking, clinical observation, and adherence to clinical trial procedures. The "placebo effect" is perhaps best described as the outcome of a richly contextualized clinical encounter [3] in which elements other than the presumed active treatment are beneficial. Hence, response to placebo should be understood within the context of the "healing situation," [4] which comprises factors both internal and external to the individual.

Placebo response appears to vary across medical conditions, cultures, and settings [5], and little is known about predictors of placebo response in sexual disorders. A small pilot study suggested that older age predicts a greater placebo response and that changes in relationship adjustment may covary with symptom reduction [6]. However, the magnitude, time course, and other predictors of placebo response in sexual dysfunction have not been examined. Understanding factors that promote symptom relief in the absence of an active treatment may inform the development of future interventions. For this purpose, the term placebo and its somewhat pejorative connotation may be misleading; rather, examining the context in which changes occur "spontaneously" may reveal underappreciated mechanisms of change.

Changes in patient behavior are believed to be a potential mediating factor in placebo response [7,8] and have great relevance to the treatment of sexual dysfunction. For instance, behavioral exercises are an important component of conventional sex therapy, and compliance with behavioral homework is a favorable prognostic indicator in the treatment of sexual dysfunctions [9]. From a cognitive-behavioral perspective, activating behavior provides opportunities for reinforcement, forces confrontation of underlying problems that are perpetuated through avoidance, and may yield evidence that change is possible. Thus, it is possible that placebo response is enhanced by changes in behavior that are congruent with the patient's wishes or expectations for improvement.

Aims

In order to better understand the nature of placebo response in sexual dysfunction, we conducted a descriptive study of symptom severity during placebo treatment in a sample of women with female sexual arousal disorder. The primary aim of the study was to test the hypothesis that symptom severity at a given time point is a function of recent frequency of sexual behavior, and specifically satisfactory sexual encounters. Secondary aims were to describe the magnitude and clinical significance of placebo response and predictors of variation in placebo response between individuals.

Methods

Data Source

This study was a secondary analysis of data from a large, multisite, Phase III randomized controlled trial of tadalafil for female sexual arousal disorder. The data set was from a 12-week parallel-group trial in which 200 participants at 13 sites were assigned to receive either placebo or one of three doses of tadalafil. The trial sponsor (Eli Lilly/ICOS) agreed to release data from all trial enrollees who were randomized to treatment with placebo. Data were converted from their original format and analyzed using SPSS version 15 (SPSS, Inc., Chicago, IL, USA) unless otherwise specified.

Participants

Fifty participants in the parent trial were randomized to the placebo arm and constituted the sample for the present study. Participants were required to meet diagnostic criteria for female sexual arousal disorder as specified in the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV [10]). They were also required to be premenopausal, between the ages of 35 and 55, in a stable relationship with a sexual partner, and using a medically approved form of contraception. Participants agreed to attempt sexual activity (either masturbation or sexual activity with a partner) at least three times during a 4-week pre-baseline run-in period and at least three times every 4 weeks during treatment. Participants were excluded for primary (but not secondary) sexual pain conditions, lifelong (but not acquired) hypoactive sexual desire disorder, recent or current pregnancy, active breastfeeding, and contraindicated medical conditions or treatments (e.g., cardiac disease, chemotherapy for cancer).

Measures

Female Sexual Function Index (FSFI)—Our primary outcome measure was the FSFI [11]. The FSFI is a 19-item questionnaire divided into 6 content domains: desire, arousal, lubrication, orgasm, satisfaction, and pain. The measure reliably discriminates women with DSM-IV diagnosed sexual dysfunctions from control patients [11,12]. Although many clinical trials for female sexual dysfunction use sexual activity frequency as an end point, Rellini and Meston [13] found that the FSFI was more sensitive than sexual behavior frequency in detecting clinician-rated improvement after treatment. In the present study, the FSFI was administered at baseline, 4, 8, and 12 weeks (post-treatment). We examined the FSFI Total score (sum of all 6 content domain scores) as the primary outcome, although we also separately analyzed outcomes for each domain score. Using data collected at baseline (first administration), Cronbach's alpha for the full scale (19 items) was 0.90 in this sample.

Sexual Activity Record (SAR)—The SAR was developed to measure the frequency of “successful and satisfactory sexual events” as a primary end point in clinical trials of sexual dysfunction [14]. However, in the present study the SAR was used as a *predictor* of treatment outcome. The SAR is a brief form that is completed after the respondent engages in sexual activity. Its seven items assess the respondent's experience of the most recent sexual encounter in the areas of sexual arousal, orgasm, and overall satisfaction with sexual arousal. Respondents indicate whether sexual events include self-stimulation, partnered sexual activity, or a combination of both. We focused primarily on predicting outcomes as a function of *satisfactory* sexual events (i.e., those events rated by the respondent as either moderately or very satisfying), regardless of whether they resulted in orgasm. However, in exploratory analyses we also predicted outcomes from the frequency of sexual events resulting in orgasm and the number of sexual events that included a partner (i.e., excluding self-stimulation only).

Female Sexual Distress Scale (FSDS)—The FSDS [15] is a self-report questionnaire developed to measure sexually-related personal distress in women. This measure was examined as a predictor of treatment outcome. The FSDS lists 12 feelings or problems and asks the respondent to indicate how often each problem has caused distress in the past 30 days. Response choices are “never,” “rarely,” “occasionally,” “frequently,” and “always.” The questionnaire is scored by summing the item responses (scaled such that “never” equals 0 and “always” equals 4). Using data collected at intake (first administration), Cronbach's alpha for the FSDS was 0.93 in this sample.

Procedure

After completing an initial telephone screening interview, eligible persons attended a clinic visit that included informed consent, a diagnostic interview for DSM-IV sexual dysfunctions, and a medical evaluation to rule out health-related exclusion criteria. Participants also completed the FSDS at this session. Upon enrollment, participants entered a 4-week baseline run-in period and began recording their sexual encounters using the SAR. Participants then returned to the clinic for a second visit, at which point they completed the FSDS and the FSFI and received instructions pertaining to their treatment. Visit 2 (hereafter referred to as “baseline”) marked the beginning of the 12-week double-blind treatment period. Treatment was self-administered on demand; participants were instructed to ingest one tablet prior to each instance of sexual activity. Throughout the 12-week treatment period, participants continued to record sexual activities using the SAR.

Data Analysis

Magnitude and Significance of Outcome—We conducted a repeated measures analysis of variance ($_{ANOVA}$) to determine the statistical significance of the overall mean change in outcome and computed the within-group effect size (Cohen’s d) to quantify the magnitude of the change in FSFI scores pre- to post-treatment. We also estimated and compared effect sizes for each of the six FSFI domain scores according to the method described by Cumming and Finch (2001). All effect size calculations were performed using Exploratory Software for Confidence Intervals [16].

To determine whether participants exhibited change scores that were unlikely to be caused by measurement error, we computed a Reliable Change Index according to the method described by Jacobson and Truax [17]. To compute a threshold for reliable change at the 0.05 alpha level, we determined the difference score necessary to exceed 1.96 times the standard error of difference (s_{diff}) between pre- and post-treatment scores [17]. We also defined a clinically significant outcome as an FSFI post-treatment score of ≥ 26 [18]. However, regardless of score, if a participant did not meet the reliable change criterion (e.g., a participant who entered the study with a score of 25 and improved by 2 points), she was not considered to have experienced clinically significant change. All tests were performed separately for completer and intent-to-treat (last observation carried forward) samples.

Predictors of Outcome—We computed the 12-week response to placebo by subtracting the FSFI Total score at baseline from the post-treatment FSFI Total score. We then correlated this change score with several baseline predictors of interest: age, baseline FSFI Total score, FSFI Satisfaction domain score (as a proxy for the perceived quality of the sexual relationship), and FSDS score. We also determined the univariate association (Pearson correlation) of satisfying sexual event (SSE) frequency and FSFI scores within 4-week measurement intervals. Predictors that were significantly associated with the pre-post change score were retained for inclusion in a multivariable, longitudinal model predicting change in FSFI Total score across the four assessment intervals.

For the purpose of this study, an SSE was defined as any recorded sexual encounter for which the participant rated herself “moderately satisfied” or “very satisfied” in response to the summary question, “How satisfied were you with your sexual arousal during this sexual activity?” on the SAR. We computed the correlation between SSE count and FSFI Total score for each measurement period and tested for differences in the strength of association across time.

Multilevel Linear Modeling of FSFI Outcome—Whereas baseline predictors are single values that are fixed in time or not expected to vary substantially over time, sexual

encounters are variable over time and may covary with symptom reduction. To accommodate both types of variables in a multiple regression model, we used a multilevel linear modeling strategy. Multilevel models, also known as linear mixed models or hierarchical linear models, are appropriate for simultaneous analysis of within-subjects variables that vary over time (i.e., “level-1” variables [19,20]) and time-invariant, between-subjects variables (i.e., “level-2” variables). Multilevel models are particularly well suited to longitudinal analysis as they allow for modeling of individual change parameters with autocorrelated errors as well as higher-level factors that are systematically related to individual change parameters [21]. Furthermore, multilevel models do not require all subjects to be measured at the same number of time points, and in most cases parameter estimates are robust to missing data.

We followed Singer and Willett’s [20] general guidance for multilevel modeling of longitudinal data. All models predicted FSFI Total score as the outcome variable using full maximum likelihood estimation of fixed and random effects. Model selection proceeded in a stepwise fashion beginning with the simplest model and adding parameters at each step. The level-1 model was fully parametrized before the addition of level-2 predictors. Single-parameter hypothesis tests (z -tests) were used to guide selection of parameters at each step, with nonsignificant parameters omitted. HLM 6 (Scientific Software International, Inc., Lincolnwood, IL, USA) was used to estimate model parameters. We also tested the relative goodness-of-fit (model deviance) between entire models as a criterion for model refinement and comparison [19,20].

After tests of our main hypotheses, we tested several exploratory models by replacing SSEs with other categories of sexual events as predictors: (i) any sexual event; (ii) any sexual event including a partner; any (iii) sexual event resulting in orgasm. To compare the fit of these models we used Akaike’s Information Criterion, which consists of the model’s deviance statistic adjusted for the number of estimated parameters [20].

Results

Participant Characteristics

Fifty women entered the study and completed assessments at baseline. Seven women discontinued participation prior to post-treatment, leaving 43 treatment completers and a total of 184 outcome measurements across participants and measurement intervals. In addition to female sexual arousal disorder, 45 participants (90%) also met diagnostic criteria for hypoactive sexual desire disorder (acquired type), 37 participants (74%) met criteria for female orgasmic disorder, and 12 participants (24%) met criteria for dyspareunia. Table 1 displays demographic and health-related characteristics of the full sample at baseline.

Magnitude and Clinical Significance of Response to Placebo Treatment

Although the FSFI was not administered at the initial intake session, FSDS scores suggested no change in symptom distress during the run-in period (i.e., from intake to baseline), $t(49) = 1.45$, $P = 0.15$. Using all recorded observations, mean FSFI Total scores were 17.98 (standard deviation [SD] = 4.63) at baseline, 24.05 (SD = 5.83) at 4 weeks, 22.84 (SD = 6.44) at 8 weeks, and 23.80 (SD = 6.60) at post-treatment. A repeated measures ANOVA confirmed an overall significant change in scores over time ($F = 32.653$, d.f. = 3, $P < 0.001$). Change in FSDS scores from baseline to post-treatment was significantly and strongly associated with change in FSFI scores across the same period ($r = -0.60$, $P < 0.001$). Thus, distress about sexual problems did not appear to change merely in response to enrolling in the clinical trial, but was strongly correlated with symptom reduction during placebo treatment.

Among treatment completers, the average change of 6.87 points from baseline to post-treatment constituted an effect size (Cohen's d) of 1.067. Effect sizes for individual domain scores (Figure 1) did not differ significantly, with the exception of the Pain domain score. In an intent-to-treat analysis, dropouts and completers had similar magnitudes of change from baseline to post-treatment; $t = -1.178$, $P = 0.245$.

Out of all treatment completers ($n = 43$), 26 participants (60.5%) met the reliable change criterion described above. Fifteen completers (34.9%) met both the reliable change criterion and the most clinically significant change criterion. In an intent-to-treat analysis, assuming that no further change would have taken place among participants who discontinued treatment, 30 participants (60.0%) met the reliable change criterion, and 17 participants (34.0%) met both the reliable change criterion and the clinically significant change criterion.

Baseline Predictors of Treatment Outcome

Table 2 displays Pearson correlation coefficients between FSFI change score (post-treatment minus pretreatment FSFI) and several variables recorded at baseline. The only significant finding was that women with lower FSFI Satisfaction domain scores tended to report greater change on the FSFI Total score from baseline to post-treatment.

SAR Data and Trends

Participants recorded a total of 1,292 sexual events across the 12 weeks of the clinical trial. When analyzed by 4-week intervals, the total number of recorded sexual events increased during the first and second 4-week treatment periods but returned to baseline levels during the final 4 weeks of the trial. However, the proportion of sexual activities labeled as satisfying increased significantly, from 23.0% during the run-in period (from intake to baseline) to 50.7% during the final 4 weeks of treatment; $\chi^2(3) = 61.961$, $P < 0.001$ (Figure 2). Across measurement intervals, SSE frequency in a given 4-week period was strongly and consistently correlated with the FSFI score at the end of the 4 weeks; Pearson r 's ranged from 0.47 to 0.64 and did not significantly differ in magnitude, $\chi^2(3) = 1.80$, $P = 0.61$.

In response to Question 4 on the SAR ("How much genital stimulation did you receive?"), overall a greater proportion of answers indicated "much" or "very much" genital stimulation with sexual activity during treatment compared with baseline, and a smaller proportion of answers indicated "a little" or "very little" genital stimulation with sexual activity during treatment as compared with baseline. The association of visit number and proportion of events coded as having "much" or "very much" genital stimulation was statistically significant, $\chi^2 = 19.186$, $P < 0.001$. As a group, participants endorsed having received a greater amount of genital stimulation during sexual activity while taking placebo tablets than during the 4-week baseline run-in period. Follow-up analyses suggested that this pattern was not explained by changes in masturbation or self-stimulation during sexual activities.

Multilevel Model Predicting FSFI Outcome

In addition to the apparent effect of time on outcomes, two significant predictors of treatment outcome emerged from bivariate analyses: a between-subjects variable assessed at a single time point (FSFI Satisfaction subscale score at baseline) and a within-subjects predictor assessed at each assessment interval during the study (SSEs). We evaluated a series of multilevel models to test the independent effects of time and SSE count, and to test whether baseline FSFI Satisfaction moderated either effect. First, we determined the best-fitting level-1 model as a function of time (i.e., treatment period: weeks -4 to 0, 0 to 4, 4 to 8, and 8 to 12). We modeled time as both a linear and a quadratic polynomial function to accommodate the curvilinear pattern of outcome scores. We then modeled the effect of SSEs during each 4-week treatment period. After final selection of level-1 parameters, we

expanded the model by testing the significance of baseline FSFI Satisfaction score as a potential between-subjects (level-2) predictor.

Summary of Model Fitting

The best-fitting level-1 model included fixed effects for time, time², and SSE, and random effects for SSE. The addition of level-2 parameters resulted in relatively little change to the overall model fit. However, the final model included level-1 (within-person) fixed slopes for time, time², and SSE on FSFI outcome as well as a level-2 (between-person) effect of baseline FSFI Satisfaction on initial status. A random intercept and random slope for SSE were also statistically significant (i.e., statistically significant between-person variation existed for the association of SSEs and FSFI outcomes). Parameter estimates and goodness-of-fit indices are displayed in Table 3.

In exploratory analyses we predicted outcome with other measures of sexual event frequency. We tested alternative level-1 models replacing the SSE variable with the frequency of *all* sexual events (regardless of satisfaction rating), sexual events only involving a partner, or all sexual events resulting in orgasm. We compared each of these models using the AIC (lower AIC values indicate better model fit). Although we found a statistically significant fixed effect for the total number of sexual events as a level-1 predictor, the original level-1 model including SSEs provided a better fit to the data (AICs = 1,094.2 vs. 1,047.9, respectively). The original model using SSE count also proved to be a better fit to the data than did alternative models predicting outcome from events resulting in orgasm (AIC = 1,073.4) and from all partnered sexual events (AIC = 1,094.3).

Discussion

In this secondary analysis of clinical trial data we evaluated the outcome of double-blind placebo treatment among 50 women diagnosed with FSAD, most of whom also met criteria for HSDD. In both completer and intent-to-treat samples, over half of participants exhibited a magnitude of change unlikely to be caused by measurement error alone, and approximately one-third of participants met an evidence-based criterion defining clinically significant change. Although the clinical trial specifically targeted women with sexual arousal problems, the outcomes of treatment were generalized such that, with the exception of sexual pain outcomes, effects were similar in magnitude across multiple domains of sexual function. These results are in line with previous trials for HSDD and FSAD that yielded evidence of a substantial clinical response in placebo recipients [1].

Our analyses indicated that the number of SSEs during the preceding 4-week period was significantly associated with the FSFI Total score. Although total sexual event frequency returned to baseline levels by post-treatment, the proportion of sexual events labeled as satisfactory was markedly higher at post-treatment. Thus, changes in sexual behavior, and particularly changes in sexual behaviors perceived as satisfying, partially accounted for variation in outcome across participants. Models predicting outcome from SSEs provided a better fit to the data than did models that predicted outcome from all sexual events, sexual events involving a partner, or sexual events resulting in orgasm. However, the presence of significant unexplained variability in the relationship between SSEs and sexual function scores suggests that the frequency of SSEs is more influential in the sexual functional outcomes of some women than others. Neither participant age nor participants' baseline sexual satisfaction appeared to moderate this relationship. A limitation of this study was a lack of viable between-person variables that could explain individual variability in the influence of sexual event frequency on outcome. Another possibility is that SSE frequency was merely a proxy for another variable, such as changes in general relationship functioning, which might better explain changes in sexual function during the trial. Finally, it is

noteworthy that participants in this trial were rigorously screened and highly motivated, as evidenced by their willingness to engage in regular sexual activity as a condition of enrollment. Thus, the relationship between SSEs and sexual function outcomes presumes several clinical and motivational characteristics that may not be typical of all women with FSAD.

Our study was observational, and thus we can only hypothesize about potential causal explanations for our findings. Prior work suggests that the observed relationship between changes in sexual behavior and outcome in placebo treatment may reflect the influence of expectancies [22]. It is also possible that behavior change was enhanced through unintended consequences of clinical trial procedures. During both the 4-week run-in period and each 4-week segment of the treatment phase, participants were asked to attempt sexual activity at least three times. This alone might have brought about positive changes by decreasing avoidance of sexual activity and generating opportunities for rewarding sexual encounters. Furthermore, completing detailed questions about each sexual encounter might have prompted participants to notice patterns of sexual response and actively consider how they might improve their sexual lives. In effect, the collection of sexual activity data might have constituted a type of self-monitoring intervention [23]. Interestingly, participants reported a greater amount of genital stimulation with sexual activities recorded during treatment than with sexual activities during the baseline run-in period. Behavioral sex therapy interventions are known to be effective in treating a variety of sexual problems [24], and compliance with behavioral homework exercises is an important predictor of treatment outcome [9,24]. To the extent that these clinical trial procedures mimicked certain aspects of behavioral therapy, a substantial clinical response is not surprising.

Although the use of existing data allowed us to analyze outcomes that otherwise may not have been feasible to collect, this methodology entails several major limitations. Most importantly, in this retrospective analysis it was not possible to directly manipulate variables of interest. We were therefore limited to our observations of a single group and the predictors of change *within* that group. In other words, our research question addressed systematic variability in symptom change within the placebo arm, not whether all the change in that group was caused by placebo administration per se. Therefore, it remains unknown whether trial participants who received no treatment would have reported the same effects as participants who received placebo. This uncertainty reflects two fundamental questions:

1. To what extent might procedures other than placebo administration have contributed to the results? In accord with recent work, we have conceptualized placebo response as a reaction to contextual elements of the clinical encounter. This position, in fact, asserts that contextual elements are necessary (though perhaps not entirely sufficient) for placebo response. Hypothetically, effects could have arisen from a variety of circumstances or events that occurred during the trial. For instance, participants interacted intensively with several clinicians during the intake visit, and a recent experimental study suggests that patient-provider interaction is a potent component of placebo response [25]. Also, as mentioned earlier, the instructions to attempt sexual activity at least three times per month is a potentially powerful manipulation that might have promoted changes in sexual function. On the other hand, we observed that sexual distress changed little during the 4-week run-in period following the intake process, and changed most rapidly during the first 4 weeks of placebo treatment. Although the timing of symptom changes may have been coincidental, it appears in this case that placebo administration was an important element, though certainly not the only or even most important one. Much remains to be learned about the effective contextual components of placebo therapy and their complicated relationship to the administration of the placebo itself. At

present, a working conceptualization is that clinical trial experiences provide meaningful information about the patient's symptoms and/or the treatment itself in a manner that facilitates improvement. Experimental studies, in which these various factors are manipulated, are clearly necessary to arrive at more definitive conclusions. A recent experimental study in men with erectile disorder tested the influence of several types of false treatment allocation feedback on placebo response and, surprisingly, found no effect of this manipulation on sexual outcomes [26]. However, it may be more profitable to focus on the effects of specific behavioral interventions embedded in clinical trial procedures.

2. To what extent might women have improved with no intervention (i.e., how much variability in outcome is attributable to the natural history of the condition)? There is little empirical data about the typical course of sexual desire and arousal disorders, particularly when the symptoms are not lifelong (as in the present study) [27], so the natural course of the disorder cannot be ruled out as a possible contributor to the observed outcomes. Moreover, knowledge of the natural history of a condition entails measurement, and therefore the question is very difficult, if not impossible, to answer without also considering the additional factor of measurement error in repeated assessment of sexual function. In women *without* sexual dysfunction, sexual function scores tend to be stable in the short term. Flory, Bissonnette, Amsel, and Binik [28] found that a control group of 40 healthy women with no sexual dysfunction reported virtually no change in symptoms over a 6-month period using a validated sexual function interview. In the validation of the FSFI, Rosen et al. [11] also reported high test-retest reliability (0.91) over 2–4 weeks in a sample of 101 women with no sexual dysfunction. Less is known about the stability of symptom severity in women with sexual problems. In the validation study of the FSFI, the test-retest reliability in a group of 97 women with FSAD (0.70) was appreciably lower than that of the control group [11]. Thus, measurement effects, whether they are fundamental to the condition or caused by weaknesses in assessment methods, may have contributed to the magnitude of change we observed. Another limitation of this study, and of most clinical trials in this area, is the lack of long-term follow-up to assess the extent to which symptom reduction was maintained after the intervention.

A further limitation of this work was the focus on behavior change and other correlates of response exclusively within the placebo arm of the trial. Analytic strategies generally assume that the magnitude of placebo response is comparable in active treatment and placebo groups, and the difference between the two is caused by the “true effect” of the active treatment. This convention is seldom questioned, though it is speculative to assume that placebo responses are identical in active and placebo treatment conditions [29]. For example, if persons who receive an active treatment can accurately detect their treatment assignment (e.g., by observing side effects), this might generate highly confident, positive expectancies that are not necessarily shared with persons who receive placebo. The experience of receiving an active treatment, and perceiving one's treatment as such, may in turn shape behaviors that are consistent with the person's desired treatment outcome. Hence, the placebo response “mechanism” associated with an active treatment may be distinguishable from that of the actual placebo treatment, and therefore our findings may not be generalizable even to other conditions within the same trial. An interesting direction for future research would be to examine between-group differences in behavior change and other processes in persons who do and do not accurately detect their treatment assignment.

Despite notable weaknesses, this study also has several important strengths. First, the sample represents clinical populations of women from multiple geographic locations in the United States who were diagnosed according to established criteria. Second, the conditions of

treatment, including a thorough medical examination in an established women's health center, commercially manufactured and packaged placebo tablets, and a recognizable trial sponsor are likely to have enhanced the credibility of the intervention. Third, compliance with treatment was relatively high, as 86% of the women in this sample completed the 12-week protocol, and dropouts did not appear to appreciably influence response rates.

The placebo effect does not exist outside of a therapeutic context, nor is it limited to the specific effects of taking a drug or undergoing a procedure [3,4,30]. A contextualized view of placebo response, in which factors both internal and external to the patient promote change in symptoms, provides a broad framework for understanding placebo response in the treatment of sexual dysfunction in women. The existence of what appears to be a large placebo response in this population reflects an opportunity to understand fundamental processes involved in symptom reduction.

Conclusion

Our work suggests that there is a close relationship between increased satisfying sexual behavior and sexual function outcomes in women receiving placebo treatment for sexual dysfunction, although there appears to be individual variation in the strength of this relationship. Promising targets for future study include evaluation of outcomes in waitlist/natural history vs. placebo treatment groups, testing of psychosocial predictor variables as moderators of placebo response, the influence of partner behavior and expectancies on outcomes, and the effects of behavior change manipulations on outcomes.

Acknowledgments

The authors wish to thank Eli Lilly/ICOS for sharing the data set used in this study. This manuscript is substantially based upon a portion of the first author's doctoral dissertation. A version of this manuscript was presented at the 2010 Annual Meeting of the International Society for the Study of Women's Sexual Health and awarded the Prize Essay Award in the Clinical Psychology category. This work was supported in part by the Houston VA HSR&D Center of Excellence (HFP90-020). The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. This research was also supported, in part, by Grant Number R01 HD51676-3 from the National Institute for Child Health and Human Development to Cindy M. Meston. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute for Child Health and Human Development.

References

1. Bradford A, Meston CM. Placebo response in the treatment of women's sexual dysfunctions: A review and commentary. *J Sex Marital Ther.* 2009; 35:164–181. [PubMed: 19360517]
2. Ernst E, Resch KL. Concept of true and perceived placebo effects. *BMJ.* 1995; 311:551–553. [PubMed: 7663213]
3. Price DD, Finniss DG, Benedetti F. A comprehensive review of the placebo effect: Recent advances and current thought. *Annu Rev Psychol.* 2008; 59:565–590. [PubMed: 17550344]
4. Papakostas YG, Daras MD. Placebos, placebo effect, and the response to the healing situation: The evolution of a concept. *Epilepsia.* 2001; 42:1614–1625. [PubMed: 11879377]
5. Kleinman, A.; Guess, HA.; Wilentz, JS. An overview. In: Guess, HA.; Kleinman, A.; Kusek, JW.; Engel, LW., editors. *The science of placebo: Toward an interdisciplinary agenda.* London: BMJ Books; 2002. p. 1-32.
6. Bradford A, Meston CM. Correlates of placebo response in the treatment of sexual dysfunction in women: A preliminary report. *J Sex Med.* 2007; 4:1345–1351. [PubMed: 17666035]
7. Bootzin, RR.; Caspi, O. Explanatory mechanisms for placebo effects: Cognition, personality and social learning. In: Guess, HA.; Kleinman, A.; Kusek, JW.; Engel, LW., editors. *The science of placebo: Toward an interdisciplinary agenda.* London: BMJ Books; 2002. p. 108-132.

8. Stewart-Williams S. The placebo puzzle: Putting together the pieces. *Health Psychol.* 2004; 23:198–206. [PubMed: 15008665]
9. Hawton K, Catalan J. Prognostic factors in sex therapy. *Behav Res Ther.* 1986; 24:377–385. [PubMed: 3741303]
10. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th edition. Washington: American Psychiatric Association; 1994.
11. Rosen R, Brown C, Heiman J, Leiblum S, Meston CM, Shabsigh R, Ferguson D, D'Agostino R Jr. The Female Sexual Function Index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. *J Sex Marital Ther.* 2000; 26:191–208. [PubMed: 10782451]
12. Meston CM. Validation of the Female Sexual Function Index (FSFI) in women with female orgasmic disorder and in women with hypoactive sexual desire disorder. *J Sex Marital Ther.* 2003; 29:39–46. [PubMed: 12519665]
13. Rellini A, Meston C. The sensitivity of events logs, self-administered questionnaires and photoplethysmography to detect treatment-induced changes in female sexual arousal disorder (FSAD) diagnosis. *J Sex Med.* 2006; 3:283–291. [PubMed: 16490021]
14. Center for Drug Evaluation and Research. Guidance for industry: Female sexual dysfunction: Clinical development of drug products for treatment. Silver Spring, MD: Food and Drug Administration, US Department of Health and Human Services; 2000 [accessed August 2010]. Available at: <http://www.fda.gov/ScienceResearch/SpecialTopics/WomensHealthResearch/ucm133202.htm>
15. Derogatis LR, Rosen R, Leiblum S, Burnett A, Heiman J. The Female Sexual Distress Scale (FSDS): Initial validation of a standardized scale for assessment of sexually related personal distress in women. *J Sex Marital Ther.* 2002; 28:317–330. [PubMed: 12082670]
16. Cumming, G. Exploratory software for confidence intervals [computer program] 2001. Melbourne, Australia: La Trobe University; [accessed July, 2008]. Available at: <http://www.latrobe.edu.au/psy/esci>
17. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991; 59:12–19. [PubMed: 2002127]
18. Wiegel M, Meston C, Rosen R. The female sexual function index: Cross-validation and development of clinical cut-off scores. *J Sex Marital Ther.* 2005; 31:1–20. [PubMed: 15841702]
19. Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2nd edition. Thousand Oaks, CA: Sage Publications; 2002.
20. Singer, JD.; Willett, JB. Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford University Press; 2003.
21. Francis DJ, Fletcher JM, Stuebing KK, Davidson KC, Thompson NM. Analysis of change: Modeling individual growth. *J Consult Clin Psychol.* 1991; 59:27–37. [PubMed: 2002140]
22. Ross M, Olson JM. An expectancy-attribution model of the effects of placebos. *Psychol Rev.* 1981; 88:408–437. [PubMed: 7025076]
23. Beck, AT.; Rush, AJ.; Shaw, BF.; Emery, G. Cognitive therapy of depression. New York: Guilford Press; 1979.
24. Sarwer DB, Durlak JA. A field trial of the effectiveness of behavioral treatment for sexual dysfunction. *J Sex Marital Ther.* 1997; 23:87–97. [PubMed: 9230489]
25. Kaptchuk TJ, Kelley JM, Conboy LA, Davis RB, Kerr CE, Jacobson EE, Kirsch I, Schyner RN, Nam BH, Nguyen LT, Park M, Rivers AL, McManus C, Kokkotou E, Drossman DA, Goldman P, Lembo AJ. Components of placebo effect: Randomised controlled trial in patients with irritable bowel syndrome. *BMJ.* 2008; 336:999–1003. [PubMed: 18390493]
26. de Araujo AC, da Silva FG, Salvi F, Awad MC, da Silva EA, Damião R. The management of erectile dysfunction with placebo only: Does it work? *J Sex Med.* 2009; 6:3440–3448. [PubMed: 19758285]
27. Rosen RC, Connor MK, Maserejian NN. The HSDD Registry for Women: A novel patient registry for women with generalized acquired hypoactive sexual desire disorder. *J Sex Med.* 2010; 7:1747–1756. [PubMed: 20233298]

28. Flory N, Bissonnette F, Amsel RT, Binik YM. The psychosocial outcomes of total and subtotal hysterectomy: A randomized controlled trial. *J Sex Med.* 2006; 3:483–491. [PubMed: 16681474]
29. Barrett B, Muller D, Rakel D, Rabago D, Marchand L, Scheder J. Placebo, meaning, and health. *Perspect Biol Med.* 2006; 49:178–198. [PubMed: 16702703]
30. Spiro, H. Clinical reflections on the placebo phenomenon. In: Harrington, A., editor. *The placebo effect: An interdisciplinary exploration.* Cambridge: Harvard University Press; 1997. p. 37-55.

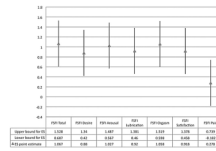


Figure 1. Effect sizes (ES) and 95% confidence intervals for Female Sexual Function Index (FSFI) Total and Domain scores (pre- to post-treatment).

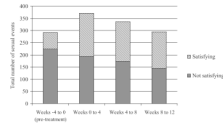


Figure 2. Sexual event frequency (all participants) by time and satisfaction rating.

Table 1

Demographics and health-related characteristics of the sample

Variable	Mean (standard deviation) (range)	N (%)
Age (years)	41.98 (4.22) (35.78–50.09)	
Ethnicity		
Black/African American		3 (6%)
East Asian		1 (2%)
Hispanic/Latina		2 (4%)
White/Caucasian		44 (88%)
Marital status		
Divorced		5 (10%)
Married		36 (72%)
Never married		8 (16%)
Separated		1 (2%)
Body mass index	27.13 (5.88) (19.68–42.97)	
Systolic blood pressure	118.48 (12.32) (96–157)	
Diastolic blood pressure	74.36 (7.55) (58–90)	
Current alcohol use		40 (80%)
Current tobacco use		10 (20%)

Table 2

Correlation of baseline predictors and Female Sexual Function Index (FSFI) Total change score (pre- to post-treatment; N = 43)

Variable	Correlation with FSFI Total change score (Pearson <i>r</i>)	<i>P</i>
Age	0.06	0.72
Baseline sexual function	-0.27	0.08
Baseline sexual satisfaction	-0.31	0.04
Baseline sexual distress	-0.03	0.85

Table 3

Parameter estimates of fixed and random effects and goodness-of-fit indices for selected models predicting Female Sexual Function Index (FSFI) Total score

Parameter	Model 1	Model 2	Model 3	Model 4
Fixed effects				
Level 1				
Intercept	19.56** (0.57)	18.39** (0.64)	17.35** (0.70)	12.34** (1.98)
Time	2.08** (0.33)	5.87** (0.82)	4.06** (0.72)	4.09** (0.72)
Time ²	—	-1.27** (0.22)	-0.92** (0.23)	-0.93** (0.22)
SSE	—	—	0.83** (0.12)	0.78** (0.10)
Level 2				
FSFI-Satisfaction/intercept	—	—	—	1.72 [†] (0.65)
Random effects				
Level 1				
Residual	14.06 (2.10)	11.82 (1.77)	8.98 (1.15)	8.98 (1.13)
Level 2				
Intercept	7.81 [†] (3.89)	8.93** (3.71)	14.85** (4.05)	12.32** (3.49)
Time	1.88 [†] (1.10)	1.77* (0.96)	—	—
Satisfying sexual events	—	—	0.09 [†] (0.07)	0.03 [†] (0.04)
No. of estimated parameters	6	7	8	9
Deviance	1,105.8	1,083.3	1,031.9	1,025.7
Pseudo-R ²	—	0.16	0.36	0.36

** $P < 0.0001$,

* $P < 0.01$,

[†] $P < 0.05$.

Models did not include random effects for time² or FSFI-Satisfaction score caused by negligible effects on goodness-of-fit. Pseudo-R² is based on the comparison of the unconditional growth model (Model 1) to more parameterized models [20].