SEXUAL MEDICINE REVIEWS REVIEW

Effect Size in Efficacy Trials of Women With Decreased Sexual Desire

Robert E. Pyke, MD, PhD,[1] and Anita H. Clayton, MD[2]

ABSTRACT

Background: Regarding hypoactive sexual desire disorder (HSDD) in women, some reviewers judge the effect size small for medications vs placebo, but substantial for cognitive behavior therapy (CBT) or mindfulness meditation training (MMT) vs wait list. However, we lack comparisons of the effect sizes for the active intervention itself, for the control treatment, and for the differential between the two.

Aim: For efficacy trials of HSDD in women, compare effect sizes for medications (testosterone/testosterone transdermal system, flibanserin, and bremelanotide) and placebo vs effect sizes for psychotherapy and wait-list control.

Methods: We conducted a literature search for mean changes and SD on main measures of sexual desire and associated distress in trials of medications, CBT, or MMT. Effect size was used as it measures the magnitude of the intervention without confounding by sample size.

Outcomes: Cohen d was used to determine effect sizes.

Results: For medications, mean (SD) effect size was 1.0 (0.34); for CBT and MMT, 1.0 (0.36); for placebo, 0.55 (0.16); and for wait list, 0.05 (0.26).

Clinical Translation: Recommendations of psychotherapy over medication for treatment of HSDD are premature and not supported by data on effect sizes. Active participation in treatment conveys considerable non-specific benefits. Caregivers should attend to biological and psychosocial elements, and patient preference, to optimize response.

Conclusions: Few clinical trials of psychotherapies were substantial in size or utilized adequate control paradigms. Medications and psychotherapies had similar, large effect sizes. Effect size of placebo was moderate. Effect size of wait-list control was very small, about one quarter that of placebo. Thus, a substantial non-specific therapeutic effect is associated with receiving placebo plus active care and evaluation. The difference in effect size between placebo and wait-list controls distorts the value of the subtraction of effect of the control paradigms to estimate intervention effectiveness. **Pyke RE, Clayton AH. Effect Size in Efficacy Trials of Women With Decreased Sexual Desire. Sex Med Rev 2018;XX:XXX–XXX.**

Key Words: Hypoactive Sexual Desire Disorder; Flibanserin; Bremelanotide; Cognitive Behavior Therapy; Mindfulness; Effect Size

INTRODUCTION

Hypoactive sexual desire disorder (HSDD) in women is associated with decreased health-related quality of life, low satisfaction with partners, and depression, yet remains undertreated.[1]

In reviewing the clinical trials of treatments for HSDD in women, we have pointed out that statistical testing is the overwhelmingly favored method of reporting results in publications of such clinical trials, but fails to tell the clinical relevance of change with treatment. Our reviews have concentrated on *clinical significance*, using responder and remitter analyses.[2,3] This gives point estimates for the percent of treated patients with a clinically relevant level of improvement. This is obviously important to clinicians and patients weighing treatment decisions.

However, *effect size* is also important because it estimates the *strength* of the treatment statistically rather than giving an absolute value for improvement, such as mean change from baseline to end of treatment, or giving a value for the probability that the change with treatment (or difference from a control treatment) can be attributed to chance, or giving a simple point estimate of clinical significance (percent responders has varied from trial to trial with a given treatment). Statistical differences

Received December 5, 2017. Accepted January 19, 2018.

[1]Pykonsult LLC, New Fairfield, CT, USA;

[2]Departments of Psychiatry and Neurobehavioral Sciences, and Obstetrics and Gynecology, University of Virginia, Charlottesville, VA, USA

https://doi.org/10.1016/j.sxmr.2018.01.003

can be "highly statistically significant" even if the effect is trivial, if a very large sample size is used. The huge sample sizes in the phase 3 trials of medications for HSDD (about 300−500 per treatment) might indeed cast a shadow of doubt over the low *P* values found. Total reliance on *P* values rather than giving percent responders has put into question the utility of psychological treatments tested for women with HSDD.

What is effect size? Several types are recommended, but perhaps the most accepted is *Cohen d*.[4] Values for *d* can vary from infinitesimal for a treatment that gives virtually no benefit to about 2.0 (described as "huge").[5] Usual values for treatments vary from 0.2 ("small), to "medium" (0.5) to "large" (0.8)[4]; 1.20 is described as "very large."[5]

Some reviews of drug treatment trials for HSDD in women have concluded that the effect size, eg, for flibanserin and for the testosterone transdermal system (TTS), is small.[6−9] Other publications claim substantial effect sizes for cognitive behavior therapy (CBT) and mindfulness meditation training (MMT) for HSDD, based on uncontrolled or wait-list control trials.[2,10] In wait-list control trials, the active intervention is given (unblinded) to a sub-set of enrolled patients, usually half; the remainder of the enrolled patients are also screened actively but then are assigned to no visits/intervention for evaluation or care for the same duration, ie, assigned to wait for treatment.

At least 4 questions arise from the apparent discrepancy in effect sizes. (1) What is the effect size of the active intervention itself? (2) What is the effect size of the control treatment? (3) What is the *differential* effect size between active and control? (4) Are trials using a wait list adequately controlled, or how much non-specific therapeutic effect is associated with participation when placebo-taking subjects receive as much active care and evaluation as those who receive pharmacologically active treatment?

AIMS

For efficacy trials of HSDD in women, we sought to compare effect sizes for medications (TTS, flibanserin, and bremelanotide) and placebo vs effect sizes for psychotherapy and wait-list control.

METHODS

MEDLINE was searched for recent reviews and all publications on studies incorporating "female sexual dysfunction" and "clinical trial" through November 1, 2017. Trials studying dysfunctional sexual desire were included even if HSDD was not the only sexual problem.

The currently accepted ways to show benefits with a treatment for HSDD rely on ameliorating the 2 defining features of the disorder: loss of sexual desire and the associated significant distress.[7,11] Change in sexual desire was tabulated using the main measure(s) of sexual desire and the main measure(s) of sexual distress in the study. Values for Cohen d were calculated as the mean change with treatment, divided by the baseline SD of the respective treatment group.[4] Values from 0.2−0.4 were designated as small, values 0.5−0.7 were designated as moderate, and values of 0.8 or more were designated as large.

Where SD were unavailable, they were estimated by multiplying the SE by the square root of n, or by reference to similar trials.

No published studies were excluded.

RESULTS

Flibanserin

Effect sizes for medication and for placebo in the 4 published flibanserin trials are shown in Table 1. The desire domain of the Female Sexual Function Index (FSFI-d)[18] was tabulated as the measure of sexual desire, though a non-validated daily measure had also been used in 2 of the studies.[12] Change in the Female Sexual Distress Scale (FSDS)-Revised (FSDS-R)[19] was tabulated as the main measure of sexual distress. Change in the scale's item 13, bother about low desire, appears to be a closer measure of sexual distress related to low desire, but lack of precision was notable in the references. In the Evaluation of the Impact on Sexuality with Evening Treatment (VIOLET) trial, mean change with drug was −0.8 vs −0.5 with placebo, yet placebo-subtracted change was listed as 0.4.[13] In the Dose Ascending Study Over Half a Year (DAISY) trial, the corresponding values were −0.7 and −0.5, yet placebo-subtracted change was −0.3.[14] The authors' calculated value of d for placebo on FSDS-R item 13 was also anomalously large, a mean of 0.77 vs 0.61 for placebo on the FSDS-R total. Therefore, the FSDS-R *total* score was used instead as a more conservative estimate.

Of the 8 results for medication, the median effect size was 1.0; range, 0.83−1.43 (FSFI-d: 1.0, 1.27, 1.29, and 1.43; FSDS-R: 0.83, 0.89, 0.90, and 1.04). Of the 8 placebo results, the median effect size was 0.70; range, 0.57−1.0 (FSFI-d: 0.57, .79, .80, and 1.0; FSDS-R: 0.50, 0.52, 0.69, and 0.70). Each differential effect size (effect size for drug minus effect size for placebo) favored drug. Values were 0.20, 0.31, 0.34, 0.40, 0.43, 0.43, 0.47, and 0.50. Thus, the median differential effect size was 0.4 favoring drug over placebo.

Bremelanotide

Data for bremelanotide from the phase 2 and phase 3 trials are shown in Table 2. The sexual distress end point emphasized in the phase 3 results was the FSDS−desire/arousal/orgasm[20] item 13, which is identical to the FSDS-R item 13. Data on the 2 largest doses were pooled in the phase 2 study publication.[21] Of the 2 results for medication, the median effect size was approximately 0.9; values were 0.85, 0.90, and 1.11. Of the 3 available placebo results, the median effect size was approximately 0.4; values were 0.3, 0.4, and 0.52. Differential effect sizes (effect size with drug minus effect size with placebo) favored drug by a median of approximately 0.3; range, 0.30−0.71; values, 0.3, 0.33, and 0.71.