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;6:358–366.**

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.

Table 1. Effect sizes (Cohen d) in large-scale flibanserin trials

Trial name, menopausal status	Group (n)	FSFI-d			FSDS-R			FSDS-R item 13		
	Scale range, best—worst	6.0—1.2			0—52			0—4		
		Baseline	Change at end	d	Baseline	Change at end	d	Baseline	Change at end	d
VIOLET,[12,13] pre-menopausal	Drug (290)	1.9 (0.7)	0.90	1.29	30.7 (9.8)	−8.86	0.90	3.2 (0.9)	−0.8	0.89
	Placebo (295)	1.8 (0.7)	0.55	0.79	30.2 (9.9)	−4.93	0.50	3.2 (0.8)	−0.5	0.62
	Drug—placebo			0.50			0.40			
DAISY,[12,14] pre-menopausal	Drug (395)	1.8 (0.7)	0.89	1.27	30.8 (9.4)	−7.77	0.83	3.3 (0.7)	0.7	1.00
	Placebo (398)	1.8 (0.7)	0.56	0.80	30.2 (10.0)	−5.22	0.52	3.2 (0.8)	0.5	0.62
	Drug—placebo			0.47			0.31			
BEGONIA,[15,16] pre-menopausal	Drug (542)	1.9 (0.7)	1.0	1.43	32.8 (9.0)	−9.4	1.04	3.4 (0.7)	−1.0	1.43
	Placebo (545)	1.9 (0.7)	0.7	1.0	32.5 (8.7)	−6.1	0.70	3.4 (0.7)	−0.7	1.00
	Drug—placebo			0.43			0.34			0.4
SNOWDROP,[17] post-menopausal	Drug (468)	1.8 (0.7)	0.7	1.0	30.5 (9.3)	−8.3	0.89	3.3 (0.8)	−0.8	1.00
	Placebo (481)	1.8 (0.7)	0.4	0.57	31.2 (9.1)	−6.3	0.69	3.3 (0.7)	−0.6	0.86
	Drug—placebo			0.43			0.20			0.14

FSDS-R = Female Sexual Distress Scale-Revised; FSFI-d = desire domain of the Female Sexual Function Index; item 13 = frequency bothered by low sexual desire.
Treatment (drug or placebo) is followed by the number of patients at baseline (n).
All other values except d are mean (SD).
Cohen d, calculated as mean at end of treatment minus mean at baseline, divided by SD at baseline.

To date, published phase 3 results (studies 301 and 302) have included only the mean changes on the 2 co-primary measures of FSFI-d and FSDS—desire/arousal/orgasm item 13 (distress about low sexual desire). Mean change with medication, mean change with placebo, and mean superiority of medication over placebo were similar to the values cited from the phase 2 trial.[22] Table 2 also includes speculative values of d for the phase 3 trials, each designated by a question mark. In the large-scale flibanserin trials (Table 1), with treatment group n = 280—545, the baseline SD for FSFI-d was identical at 0.7 in each of the 8 treatment groups, so it seems reasonable to assume the same value for the large-scale bremelanotide trials (mean n per treatment group, 300). Similarly, in the flibanserin trials, the baseline SD value for the FSDS desire distress item varied little, from 0.7—0.9, mean, 0.8, so a value of 0.8 was assumed for the phase 3 bremelanotide trials.

Testosterone Transdermal System

For the TTS, the available publications on phase 3a trial surgical menopause (SM)2[23,24] failed to include baseline SD. However, phase 3a trial SM1[25] and both phase 3b trials, naturally menopause (NM)1[26] and APHRODITE,[27] did include baseline variance data. For NM1, SD were not given, but SE and n were available for calculating SD as SE times the square root of n.

The main end point on sexual desire was the Profile of Female Sexual Function desire sub-scale (PFSF-desire).[28] The main end point on sexual distress was the Personal Distress Scale (PDS).[29] Of the 6 available results for TTS, the median was 0.9 (PFSF-desire: 0.75, 0.85, and 1.10; PDS: 0.85, 0.91, and 1.02). Corresponding effect sizes with placebo were a median of 0.5; range, 0.31—0.59 (PFSF-desire: 0.31, 0.43, and 0.49; PDS: 0.46, 0.57, and 0.59). In these 3 trials, TTS-placebo differences at end point were 5.12, 5.79, and 7.0 on the PFSF-desire and 7.24, 9.04, and 11 on the PDS. In the SM2 trial, the corresponding differences were similar: 4.95 on PFSF-desire and 7.24 on PDS.

All differential effect sizes favored TTS, varying from 0.39—0.61; median, 0.4; values, 0.32, 0.39, 0.42, 0.44, 0.44, and 0.61.

Table 3 shows that the baseline SD for the PFSF-desire measure is quite constant across the 6 groups cited, never varying more than from 12.36—13.72, with an overall mean of 13.04. The range for baseline SD for the PDS is also narrow, 24.1—25.76; mean, 25.125. If these mean values for SD were used to estimate Cohen d (values followed by a question mark in Table 3), the values for TTS would be 0.91 and 0.94; the values for placebo would be 0.53 and 0.65; and the differential effect size values would be 0.38 and 0.29.

CBT and MMT

No treatment trials of CBT or MMT newer than those in the 2015 review[2] could be found by MEDLINE search. Effect sizes for CBT and MMT are shown in Table 4. A measure of sexual

Table 2. Effect sizes (Cohen d) in bremelanotide trials

	Scale range, best–worst	FSFI-d			FSDS-DAO total score			FSDS-DAO item 13		
		6.0–1.2			0–60			0–4		
Study	Group (n)	Baseline	Change at end	d	Baseline	Change at end	d	Baseline	Change at end	d
54, phase II*	Drug (149)	2.5 (0.9)	0.8	1.11	33.1 (13.1)	−11.1	0.85	2.6 (1.0)	−0.9 (1.1)	0.90
	Placebo (91)	2.4 (1.0)	0.4	0.40	32.9 (13.0)	−6.8	0.52	2.6 (1.0)	−0.6 (1.3)	0.60
	Drug–placebo		0.4	0.71			0.33		−0.3	0.30
301†	Drug (circa 300)	(0.7?)‡	0.54	0.77§				(0.8)‡	−0.74	0.92§
	Placebo (circa 300)	(0.7?)‡	0.24	0.34§				(0.8)‡	−0.35	0.44§
	Drug–placebo	(0.7?)‡	0.30	0.43*				(0.8)‡	−0.39	0.49§
302†	Drug (circa 300)	(0.7?)‡	0.63	0.90§				(0.8)‡	−0.71	0.89§
	Placebo (circa 300)	(0.7?)‡	0.21	0.30§				(0.8)‡	−0.41	0.51§
	Drug–placebo	(0.7?)‡	0.42	0.60§				(0.8)‡	−0.30	0.37§

FSDS-DAO = Female Sexual Distress Scale-desire/arousal/orgasm; FSFI-d = desire domain of the Female Sexual Function Index; item 13 = frequency bothered by low sexual desire.
Treatment (drug or placebo) is followed by the number of patients at baseline (n).
All other values except d are mean (SD).
Cohen d, calculated as mean at end of treatment minus mean at baseline, divided by SD at baseline.
*125-mg and 1.75-mg Dose groups pooled.
†Treatment group sizes have not been disclosed, but the total population consisted of 1202 women and randomization was 1:1.
‡Expected SD from baseline values in similar-sized flibanserin trials.
§Speculative value (see text).

desire was included in each trial, but no measure of sexual distress in 1 trial.[31] Active treatment was associated with a median effect size of 0.91; range, 0.43–1.47; the 9 values were 0.43, 0.64, 0.65, 0.81, 0.91, 1.22, 1.30, 1.40, and 1.47. Of the 7 available wait-list results,[30–33] the median effect size was 0.13; range, −0.49 to .032; values were −0.49, 0.1, 0.05, 0.13, 0.16, 0.19, and 0.32. Differential effect size, favoring active intervention in each trial, varied from 0.11–1.79; median, 0.72; values were 0.11, 0.53, 0.64, 0.72, 1.2, 1.42, and 1.79. In 1 of the 4 trials, the measure of sexual desire was the clinician-rated Sexual Interest and Desire Inventory-Female[34] instead of the self-rated FSFI-d.

Table 3. Effect sizes (Cohen d) in testosterone transdermal system phase 2 and 3 trials

	Group (n)	PFSF-desire			PDS		
	Scale range, best–worst	100–0			0–100		
Trial		Baseline	Change at end	d	Baseline	Change at end	d
APHRODITE[27]	Drug (267)	19.6 (12.36)	13.6*	1.10	65.6 (25.52)	26.0*	1.02
	Placebo (277)	20.2 (13.4)	6.6*	0.49	66.2 (25.76)	14.7*	0.57
	Drug–placebo		7.0	0.61		11	0.44
NM1[24]	Drug (270)	20.35 (13.0)	9.79	0.75	60.67 (24.1)	−20.49	0.85
	Placebo (269)	20.45 (12.8)	4.00	0.31	59.39 (25.1)	−11.45	0.46
	Drug–placebo		5.79	0.44		−9.04	0.39
SM1[23]	Drug (269)	19.9 (12.96)	11.06	0.85	64.7 (24.92)	−22.77	0.91
	Placebo (269)	20.7 (13.72)	5.94	0.43	62.4 (25.35)	−15.07	0.59
	Drug–placebo		5.12	0.42		−7.70	0.32
SM2[21]	Drug (269)	19.79 (13.04†)	11.85	0.91†	64.78 (25.125?)	−23.55	0.94†
	Placebo (269)	20.82 (13.04†)	6.90	0.53†	62.57 (25.125?)	−16.31	0.65†
	Drug–placebo		4.95	0.38†		−7.24	0.29†

PDS = Personal Distress Scale; PFSF-desire = Profile of Female Sexual Function, desire sub-scale.
Treatment (medication or placebo) is followed by the number of patients at baseline (n).
All other values except d are mean (SD).
Cohen d, calculated as mean at end of treatment minus mean at baseline, divided by SD at baseline.
*Value interpolated from bar graph.
†Speculative value (see text).

Table 4. Effect sizes (Cohen d) in controlled trials of cognitive behavioral therapy or mindfulness meditation training to treat female sexual dysfunction

		FSFI total			FSDS or FSDS-R			SIDI-F[30]			FSFI-d		
Scale range, best—worst		36—6			52—0			51—0			6.0—1.2		
Study	Group (n)	Baseline	Treated	d	Baseline	Treated	d	Baseline	Treated	d	Baseline	Treated	d
Revive[31]	CBT (26)	19.78 (5.76)	27.30 (3.95)	1.30							2.45 (0.83)	3.65 (0.83)	1.4
	Wait list (27)	21.90 (6.51)	18.69 (8.21)	−0.49							2.68 (1.08)	2.85 (0.96)	0.16
Treatment—wait list				1.79									1.2
Pursuing Pleasure[31]	CBT (26)	21.91* (not given)	27.98* (not given)	n.av.	33.32 (9.65)	19.13 (10.54)	1.47				2.67 (0.86)	3.55 (0.64)	0.91
	Wait list (31)	21.92* (not given)	21.75* (not given)	n.av.	30.14 (10.03)	29.61 (8.99)	0.05				2.54 (1.37)	2.74 (1.11)	0.19
Treatment—wait list							1.42						0.72
Cancer[32]	Mindfulness training (31)	18.36 (6.57)	26.13 (5.01)	1.18	23.19 (10.42)	14.71 (10.74)	0.81				1.82 (0.92)	2.94 (1.41)	1.22
	Wait list (9)	15.83 (1.07)	Not given	n.av.	25.44 (10.26)	Not given	n.av.				1.87 (0.82)	Not given	n.av.
Treatment—wait list				n.av.			n.av.						n.av.
HSDD[30]	Mindfulness training (68)†	19.55 (5.35)	23.05 (6.07)	0.65	29.82 (9.32)	25.78 (11.38)	0.43	16.06 (7.01)	21.56 (9.74)	0.64	Not given		
	Wait list (49)	18.94 (4.88)	19.01 (5.11)	0.01	30.11 (9.01)	27.24 (10.98)	0.32	16.02 (7.50)	16.96 (8.91)	0.13			
Treatment—wait list				0.64			0.11			0.53			

CBT = cognitive behavioral therapy; FSDS = Female Sexual Distress Scale; FSDS-R = Female Sexual Distress Scale-Revised; FSFI = Female Sexual Function Index; FSFI-d = desire domain of the Female Sexual Function Index; HSDD = hypoactive sexual desire disorder; n.av. = not available; SIDI-F = Sexual Interest and Desire Inventory-Female.
Treatment (psychotherapy or wait list) is followed by the number of patients at baseline (n).
All other values except d are mean (SD).
Cohen d, calculated as mean at end of treatment minus mean at baseline, divided by SD at baseline.
Blanks indicate that a measure was not used in that study.
*Hand-calculated from published domain means.
†Includes women moved early from delayed-treatment group to mindfulness meditation training intervention.

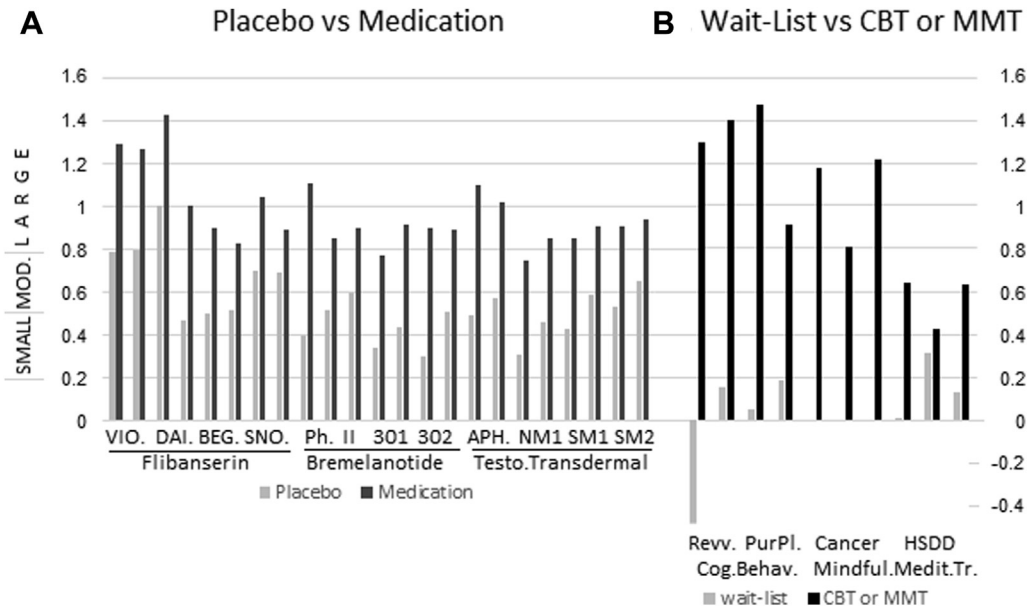

Figure 1. Effect sizes (Cohen d) for placebo vs medication (A) compared to wait list vs cognitive behavior therapy (CBT) or mindfulness meditation training (MMT) (B). Medications: flibanserin (VIOLET [VIO], DAISY [DAI], BEGONIA [BEG], SNOWDROP [SNO]), bremelanotide (phase [Ph] II study 54, Ph III studies 301 and 302), testosterone transdermal system (APHRODITE [APH]). Psychotherapies: CBT (Revive study [Revv], Pursuing Pleasure study [PurPl]); MMT (cancer patients [no wait group data], hypoactive sexual desire disorder [HSDD] patients). Desire at left, distress at right, in the same order as in Tables 1–4.

Comparison of Treatments

All values for d are summarized in Figure 1. Part A shows Cohen d for medication trials. Mean (SD) for medication was 0.97 (0.34); and for placebo was 0.55 (0.16). Part B shows Cohen d for psychotherapy trials. Mean (SD) for CBT or MMT was 1.00 (0.36); and for wait list was 0.05 (0.26). Thus, the effect sizes for medication (0.97) and CBT or MMT (1.00) were almost identical. The mean (0.55) and median (0.52) effect size for placebo were moderate, whereas the mean (0.05) and median (0.13) effect size for wait list were slightly below the threshold for small (0.20). By subtraction, the differential effect size for medication vs placebo was 0.42 and the differential effect size for CBT or MMT was 0.95. Comparing part A to part B, the apparent advantage of CBT or MMT over medication is contingent on the difference in effect size of the control group rather than on the effect size of the active treatment.

DISCUSSION

Why calculate Cohen d or effect size? Why is determination of statistical significance or the *P* value inadequate? Statistical significance is the probability that the observed difference is due to chance, and depends on both sample size and effect size, whereas effect size is the quantitative measure of the strength of a phenomenon without confounding by sample size to determine the standardized difference or magnitude of the effect. Cohen d is simply the mean change with treatment divided by the SD at baseline (SD_{bsln}). The innate value of *d* comes from SD_{bsln}. SD_{bsln} estimates the range of values constituting moderate to mild disease (severity from the 50th–84th percentile) or constituting severe to moderate disease (severity from the 16th–50th percentile), as shown in Figure 2.

If Cohen d is 1.0, the simplest interpretation is that all patients got an effect as big as the difference from the 50th percentile (moderate disease) to the 84th percentile (mild disease). Cohen d of 1.0 may also be found when half the patients reported a change from the 16th percentile (severe disease) to the 84th percentile (mild disease). No wonder Cohen calls a d of 1.0 "large."

Much more usually in medicine, values of *d* are closer to 0.5. A simple way to interpret $d = 0.5$ is that it is as if half of the patients got an effect and that the effect is as big as moving from the 50th

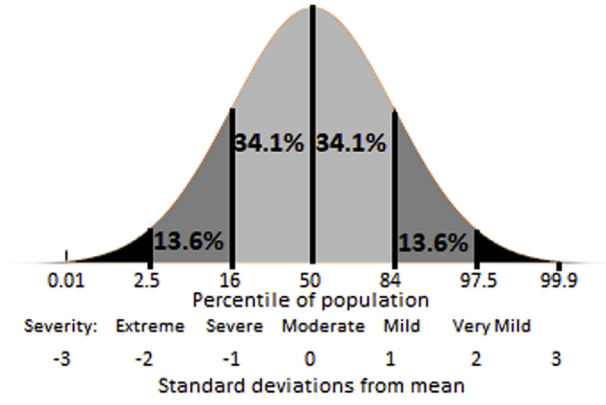

Figure 2. Normal distribution of the severity of an illness.

percentile (moderate disease) to the 84th percentile (mild disease). This shows why Cohen called that value for d "moderate."

In the case where the disorder is HSDD, the usual measure of sexual desire is the FSFI-d, range 1.2 (worst) to 6.0 (supra-normal). The usual baseline mean is about 2.0, SD_{bsln} is about 0.7, and the remitter criterion is >3.3.[12–18] The value of 1.0 for Cohen d that we found for each form of treatment can be simply interpreted based on if all patients improve by 1 SD or 0.7 points. The responder value for this measure is 0.6.[3,12] Thus, *the average change is slightly better than if all patients became responders* (a large effect indeed).

In the real world, of course, the distribution of responses includes treatment failure, moderate response, and remission. The occurrence of remission, eg, improving by 2.4 instead of 0.6, skews the post-treatment mean upward. Thus, even a Cohen d value as large as 1.0 does not mean that all patients actually become responders: if some patients remitted, some must have failed to improve if the average improvement was 0.7.

We found that medications and psychotherapies had large and very similar effect sizes around a Cohen d value of 1.0. Placebo had roughly 4 times the treatment effect of wait-listing in clinical trials of women with HSDD. The mean (and median) effect sizes were 0.55 (0.52) and 0.05 (0.13), respectively. Proof that the medications tested (flibanserin, bremelanotide, or testosterone) provide significant relief for women with HSDD was provided by double-blind placebo-controlled trials.[35,36] Proof of whether CBT or MMT provides adequate relief depends in part on the question of the adequacy of unblinded testing and wait-listing as a control treatment. The latter issues have been discussed previously,[2] concluding that the clinical trials of these psychotherapies to date provide inadequate proof of efficacy, and suggesting that some of the observed treatment effect size for CBT or MMT may depend on non-specific effects due to active, repeated interviewing and other forms of participation that tend to establish a therapeutic alliance. A very recent, authoritative review cites similar effect size values and concludes that the efficacy of CBT and MMT are supported at evidence level = 2.[10]

Therapeutic alliance is the best predictor of results in many therapies, eg, in the related fields of depression, sexual abuse, and couple's counseling.[37,38] Wait-listing not only avoids therapeutic alliance-building but also fails to provide the non-specific gain due to the Hawthorne effect, ie, an awareness of being observed improves behaviors and outcomes, as is well known in industrial psychology.[39]

In practice, the expected value of a treatment to the patient depends not only on results from adequately controlled clinical trials but also on open-label studies, because the patient will receive her treatment unblinded. We have shown open-label results with CBT and MMT, but the question remains, how does response to medication given open-label compare to response when the treatment is given double-blind vs placebo?

Only 1 open-label trial, of flibanserin, relates directly and fairly to the question, because most open-label trials of medication have reported only safety results. The Researching Outcomes on Sustained Efficacy (ROSE) study tested 738 pre-menopausal patients with HSDD for 24 weeks open-label (before randomizing responders to continue flibanserin or switch to placebo).[40] Values for FSFI-d results were not given, but the daily sexual desire measure increased by a mean of 13.2 from 12.2 (SD, 9.5) at baseline, and FSDS-R scores decreased by a mean of 9.5 from 30.2 (SD, 9.1) at baseline. Thus, values for Cohen d were 1.39 and 1.04, respectively. FSFI total score (d = 0.9) and FSDS-R item 13 (d = 1.3) also supported a large effect size. The proportion who reported improvement on the Patient Global Impression was 65%; in the double-blind trials, 50% so reported.[13–17] Thus, at least for flibanserin, open-label results, including a mean and median *d* value of 1.2 (a "very large" effect size),[5] may be somewhat more encouraging than double-blind placebo results, as expected.

On the other hand, neither medication nor psychotherapy in the form of CBT or MMT has provided remission for a substantial majority of women with distressing loss of sexual desire in clinical trials to date.[2,7–9,13–17,23–27,30–33,41] Thus, care for patients with HSDD requires attention to biological *and* psychosocial elements to optimize response.[10,42,43]

CONCLUSIONS

In women with distressing loss of sexual desire, the 3 active blinded medications that have been tested extensively (TTS, oral flibanserin, sub cutaneous bremelanotide) were each associated with large treatment effect sizes, while placebo treatment was associated with a moderate treatment effect size. In trials of CBT or MMT, unblinded active treatment was associated with large treatment effect sizes. The wait-list control in the latter trials had a less-than-small treatment effect size. Comparisons between types of trial, though less valid, shows that wait-listing provided, on average, about one-third of the treatment effect of placebo. Medication-placebo comparisons showed small to moderate *differential* effect size favoring medication. Psychotherapy and wait-list comparisons showed moderate differential effect size favoring psychotherapy. The inadequacy of wait-listing to provide non-specific treatment effects associated with active participation in a trial apparently may have contributed substantially to the differential effect favoring psychotherapy over control.

Once one understands how meaningful effect size is, applying it to treatments for HSDD in women improves our understanding of the substantial overall impact of these treatments.

Corresponding Author: Robert E. Pyke, MD, PhD, Pykonsult LLC, 23 Eastview Drive, New Fairfield, CT 06812, USA. Tel: 203-399-2390; Fax: 203-746-9152; E-mail: robertepyke@gmail.com

Disclosure: Anita Clayton acknowledges that she has received the following: grants—Axsome, Allergan, Endoceutics, Janssen, Palatin Technologies, Sage Therapeutics Inc, Takeda; advisory board fee/consultant fee—Fabre-Kramer, Palatin Technologies, S1 Biopharma Inc, Sprout—a division of Valeant; royalties/copyright—Ballantine Books/Random House, Changes in Sexual Functioning Questionnaire, Guilford Publications; shares/restricted stock units—Euthymics; S1 Biopharma Inc. Robert Pyke acknowledges a potential conflict of interest as a member of Pykonsult LLC, as a stockholder in S1 Biopharma Inc, and a consultant for Endoceutics and Olive Therapeutics.

Funding: None.

STATEMENT OF AUTHORSHIP

Category 1

(a) Conception and Design
Robert E. Pyke
(b) Acquisition of Data
Robert E. Pyke; Anita H. Clayton
(c) Analysis and Interpretation of Data
Robert E. Pyke; Anita H. Clayton

Category 2

(a) Drafting the Article
Robert E. Pyke
(b) Revising It for Intellectual Content
Robert E. Pyke; Anita H. Clayton

Category 3

(a) Final Approval of the Completed Article
Robert E. Pyke; Anita H. Clayton

REFERENCES

1. Parish SJ, Hahn SR. Hypoactive sexual desire disorder: a review of epidemiology, biopsychology, diagnosis, and treatment. **Sex Med Rev 2016;4:103-120.**
2. Pyke RE, Clayton AH. Psychological treatment trials for hypoactive sexual desire disorder: a sexual medicine critique and perspective. **J Sex Med 2015;12:2451-2458.**
3. Pyke RE, Clayton AH. Measures for clinical trials of HSDD in women: endpoints, desire-related behavior, and clinical significance. **Sex Med Rev 2018 Jan 19. pii: S2050-0521(17)30148-8. doi: 10.1016/j.sxmr.2017.11.008. [Epub ahead of print].**
4. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
5. Sawilowsky SS. A different future for social and behavioral science research. **J Mod Appl Stat Methods 2003;2:128-132.**
6. Jaspers L, Feys F, Bramer WM, Franco OH, Leusink P, Laan ET. Efficacy and safety of flibanserin for the treatment of hypoactive sexual desire disorder in women: a systematic review and meta-analysis. **JAMA Intern Med 2016;176:453-462.**
7. FDA. Office Director Memo: Benefit-Risk Assessment, Benefit-Risk Summary and assessment, Addyi, 2015. Available at: https://google2.fda.gov/search?q=cache:xmhxfIshJoQJ:www.accessdata.fda.gov/drugsatfda_docs/nda/2015/022526orig1s000odmemo.pdf+addyi+advisory+committee&client=FDAgov&site=FDAgov&lr=&proxystylesheet=FDAgov&output=xml_no_dtd&ie=UTF-8&access=p&oe=UTF-8. Accessed March 15, 2018.
8. European Medicines Agency. Intrinsa: European public assessment report, scientific discussion. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion/human/000634/WC500034387.pdf. Accessed November 5, 2017.
9. Intrinsa (testosterone trasdermal system). NDA 21-769, review by the Division of Reproductive and Urologic Drug Products, November 3, 2004, page 32. Available at: https://www.fda.gov/ohrms/dockets/ac/04/briefing/2004-4082B1_02_B-FDA-Intrinsa-Medical-Review.pdf. Accessed March 15, 2018.
10. Kingsberg SA, Althof S, Simon JA, et al. Female sexual dysfunction—medical and psychological treatments, committee 14. **J Sex Med 2017;12:1463-1491.**
11. FDA. 2014: Patient-focused drug development public meeting and scientific workshop on female sexual dysfunction, October 28th scientific workshop. Available at http://wayback.archive-it.org/7993/20161023010508/http://www.fda.gov/downloads/Drugs/NewsEvents/UCM423111.pdf. Accessed March 15, 2018.
12. FDA. Background document for meeting of advisory committee for reproductive health drugs: flibanserin. May 20, 2010.
13. Derogatis LR, Komer L, Katz M, et al; VIOLET Trial Investigators. Treatment of hypoactive sexual desire disorder in premenopausal women: efficacy of flibanserin in the VIOLET study. **J Sex Med 2012;9:1074-1085.**
14. Thorp J, Simon J, Dattani D, et al; DAISY Trial Investigators. Treatment of hypoactive sexual desire disorder in premenopausal women: efficacy of flibanserin in the DAISY study. **J Sex Med 2012;9:793-804.**
15. Sprout Pharma. Briefing document for FDA. June 2015.
16. Katz M, DeRogatis LR, Ackerman R, et al. Efficacy of flibanserin in women with hypoactive sexual desire disorder: results from the BEGONIA trial. **J Sex Med 2013;10:1807-1815.**
17. Simon JA, Kingsberg SA, Shumel B, Hanes V, Garcia M, Sand M. Efficacy and safety of flibanserin in postmenopausal women with hypoactive sexual desire disorder: results of the SNOWDROP trial. **Menopause 2014;21:633-640.**
18. Wiegel M, Meston C, Rosen R. The Female Sexual Function Index (FSFI): cross-validation and development of clinical cutoff scores. **J Sex Marital Ther 2006;13:46-56.**
19. Derogatis L, Clayton A, Lewis-D'Agostino D, Wunderlich G, Fu Y. Validation of the Female Sexual Distress Scale-Revised for assessing distress in women with hypoactive sexual desire disorder. **J Sex Med 2008;5:357-364.**
20. DeRogatis LR, Edelson J, Revicki DA. Reliability and validity of the Female Sexual Distress Scale—desire/arousal/orgasm instrument in a phase 2B dose-ranging study of bremelanotide.

Paper presented at: 167th Annual Meeting of the American Psychiatric Association. May 3–7, 2014; New York, NY.

21. Clayton AH, Althof SE, Kingsberg S, et al. Bremelanotide for female sexual dysfunctions in premenopausal women: a randomized, placebo-controlled dose-finding trial. **Womens Health (Lond) 2016;12:325-337.**
22. Palatin Technologies Inc. Bremelanotide meets co-primary endpoints in Palatin's phase 3 trials for hypoactive sexual desire disorder. Available at: https://www.prnewswire.com/news-releases/bremelanotide-meets-co-primary-endpoints-in-palatins-phase-3-trials-for-hypoactive-sexual-desire-disorder-300355401.html. Accessed November 3, 2017.
23. Buster JE, Kingsberg SA, Aguirre O. Testosterone patch for low sexual desire in surgically menopausal women: a randomized trial. **Obstet Gynecol 2005;105:944-952.**
24. Davis SR, Braunstein GD. Efficacy and safety of testosterone in the management of hypoactive sexual desire disorder in postmenopausal women. **J Sex Med 2012;9:1134-1148.**
25. Simon J, Braunstein G, Nachtigall L, et al. Testosterone patch increases sexual activity and desire in surgically menopausal women with hypoactive sexual desire disorder. **J Clin Endocrinol Metab 2005;90:5226-5233.**
26. Shifren JL, Davis SR, Moreau M, et al. Testosterone patch for the treatment of hypoactive sexual desire disorder in naturally menopausal women: results from the INTIMATE NM1 study. **Menopause 2006;13:770-779.**
27. Davis SR, Moreau M, Kroll R, et al; APHRODITE Study Team. Testosterone for low libido in postmenopausal women not taking estrogen. **N Engl J Med 2008;359:2005-2017.**
28. Derogatis L, Rust J, Golombok S, et al. Validation of the profile of female sexual function (PFSF) in surgically and naturally menopausal women. **J Sex Marital Ther 2004;30:25-36.**
29. DeRogatis LR, Graziottin A, Bitzer J, Schmitt S, Koochaki PE, Rodenberg C. Clinically relevant changes in sexual desire, satisfying sexual activity and personal distress as measured by the profile of female sexual function, sexual activity log, and personal distress scale in postmenopausal women with hypoactive sexual desire disorder. **J Sex Med 2009;6:175-183.**
30. Brotto LA, Basson R. Group mindfulness-based therapy significantly improves sexual desire in women. **Behav Res Ther 2014;57:43-54.**
31. Hucker A, McCabe MP. Incorporating mindfulness and chat groups into online cognitive behavioral therapy for mixed female sexual problems. **J Sex Res 2015;52:627-639.**
32. Brotto LA, Erskine Y, Carey M, et al. A brief mindfulness-based cognitive behavioral intervention improves sexual functioning versus wait-list control in women treated for gynecologic cancer. **Gynecol Oncol 2012;125:320-325.**
33. Jones LM, McCabe MP. The effectiveness of an Internet-based psychological treatment program for female sexual dysfunction. **J Sex Med 2011;8:2781-2792.**
34. Clayton AH, Segraves RT, Bakish D, et al. Cutoff score of the sexual interest and desire inventory-female for diagnosis of hypoactive sexual desire disorder. **J Womens Health (Larchmt) 2010;19:2191-2195.**
35. Fisher WA, Pyke RE. Flibanserin efficacy and safety in premenopausal women with generalized acquired hypoactive sexual desire disorder. **Sex Med Rev 2017;5:445-460.**
36. Dooley EM, Miller MK, Clayton AH. Flibanserin: from bench to bedside. **Sex Med Rev 2017;5:461-469.**
37. Schramm E, Hautzinger M, Zobel I, Kriston L, Berger M, Haerter M. Comparative efficacy of the cognitive behavioral analysis system of psychotherapy versus supportive psychotherapy for early onset chronic depression: design and rationale of a multisite randomized controlled trial. **BMC Psychiatry 2011;11:134.**
38. Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. **Cognit Ther Res 2012;36:427-440.**
39. Hawthorne Effect. Wikipedia, 2018. Available at: https://en.wikipedia.org/wiki/Hawthorne_effect. Accessed March 15, 2018.
40. Goldfischer ER, Breaux J, Katz M, et al. Continued efficacy and safety of flibanserin in premenopausal women with hypoactive sexual desire disorder (HSDD): results from a randomized withdrawal trial. **J Sex Med 2011;8:3160-3172.**
41. DeRogatis LR, Rosen RC, Edelson J, Jordan R, Greenberg S, Portman DJ. Subcutaneous bremelanotide for female sexual dysfunctions in premenopausal women: responder analyses from a phase 2b dose-ranging study. **Obstet Gynecol 2014. Available at: http://journals.lww.com/greenjournal/Abstract/2014/05001/Bremelanotide_for_Female_Sexual_Dysfunctions_.53.aspx. Accessed March 15, 2018.**
42. Nappi RE, Martini E, Terreno E, et al. Management of hypoactive sexual desire disorder in women: current and emerging therapies. **Int J Womens Health 2010;2:167-175.**
43. Thomas HN, Thurston RC. A biopsychosocial approach to women's sexual function and dysfunction at midlife: a narrative review. **Maturitas 2016;87:49-60.**