
Nijland EA¹, Davis SR², Laan ETM³ and Weijmar Schultz WCM¹

1. University Medical Center Groningen, Department of Obstetrics and Gynaecology, PO Box 30001, 9700 RB, Groningen, The Netherlands
2. Monash University, Department of Medicine, CECS, Prahran, Victoria 3181, Australia
3. Academic Medical Center Amsterdam, Department of Sexology & Psychosomatic Obstetrics/Gynecology, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

J Sex Med 2006; 3: 763-777
Abstract

Introduction: A considerable number of double-blind randomized controlled studies investigating the effects of pharmaceutical intervention on female sexual functioning have been published in the recent years. However, a comparison between outcomes of various studies is difficult as no general accepted/correct approach to research has been established yet. To be able to translate trial results to daily clinical practice, current limitations and issues in drug intervention studies in Female Sexual Dysfunction need to be clarified.

Aim: To evaluate the needs in research into intervention in female sexual dysfunction (FSD) by reviewing published studies.

Methods: A systematic review of double blind randomized controlled intervention trials on FSD.

Main Outcome Measures: Definitions of study populations, inclusion and exclusion criteria, use of power calculations, outcome measures and treatment duration.

Results: 25 double-blind randomized controlled trials investigating the effects of pharmaceutical intervention on female sexual functioning have been published. Of these, 11 studies required the diagnosis of FSD as an inclusion criterion. A standard methodology for research in this field is lacking. Significant differences in population inclusion requirements and tools for the measurement of change in sexual functioning were identified as major limiting factors.

Conclusions: The investigation of FSD is an evolving area in that new definitions and a new model for female sexual functioning have been recently proposed. There is a need for experts in the field and regulating authorities to reach a consensus regarding appropriate inclusion and exclusion criteria for FSD trials and main outcome measures appropriate for the evaluation of drug interventions. This consensus should also determine which treatment effect is considered to be clinically relevant. Treatment efficacy and clinical relevance should be related to outcomes which are meaningful for affected women.
Introduction

The study of FSD - an evolving area of research

In recent years the study of Female Sexual Dysfunction (FSD) has received increasing attention. However, understanding of factors relevant to research into female sexual functioning has lagged behind research into male sexual health. In contrast to research in male sexual function, no gold standard self assessment instrument exists for research into female sexual function. Major reasons for the lack of standardized instruments are uncertainty as to the definition of ‘normal sexual function’ and the continuous evolution of the definition of FSD.

Normative data of sexual behavior across the adult female life-span is lacking. This contributes to difficulty in classifying a sexual problem as a sexual dysfunction. In the literature, much reference is made to a few population based studies which have indicated that the prevalence of sexual problems among women ranges from 9-43%.

However, the validity and reliability of these data is uncertain as epidemiological studies on female sexual function are constrained by the limited response rate, the limited use of validated instruments including addressing sexuality related personal distress and lack of information about the duration of sexual problems and the context of sexual problems.

The second reason for the lack of one or two generally accepted standardized questionnaires for research into FSD is that the classification system for sexual disorders according to the DSM-IV-R classification system has been revised repeatedly resulting in new definitions for FSD by subspecialty groups (American Foundation for Urologic Disease, AFUD classification 2000 and revision 2003). The proposed reclassification is based on a non-linear model of women’s sexual response that explicitly recognizes its contextual nature, the variable sequence of desire and arousal and the frequent lack of correlation between subjective and physical (genital) arousal. Subsequently, the criteria for performing intervention trials in FSD have been proposed by individual expert groups and authorities. However, no standardized approach to research has been established or meaningful endpoints agreed upon.

The Food and Drug Administration’s (FDA) 2000 draft guidance document for FSD clinical trials recommends the use of the change in the frequency of successful satisfactory sexual events recorded in a daily diary as the primary endpoint and self administered questionnaires (SAQ’s) as secondary endpoints. The applicability of this approach has been criticized by a number of experts in the field.

The complexity of the female sexual response and the strong impact of contextual factors, make the definition of endpoints and outcomes challenging. It has been proposed that the best predictors for changes in sexual response are contextual factors including previous sexual function, change of partner or feelings for partner. Sexual behavior is ultimately, social behavior (“it takes two to tango”) and where a loss of certain aspects of sexuality like arousal or interest does not automatically result in sexual problems or sexual dysfunction, an increase in sexual arousal or interest does not automatically result in sexual happiness and satisfaction.
Issues in FSD research that deserve consideration include selection of study populations and study design. **Aim:** to summarize the current limitations and issues in randomized controlled trials in FSD with regard to design and outcome measures. We have not reviewed the efficacy of the various pharmaceutical preparations on FSD. The relevance of the outcomes of studies in FSD with regard to clinical practice is reviewed, unresolved questions identified and proposals to optimize the design of future studies made.

**Methods**

We included pharmaceutical intervention studies in healthy women identified as having FSD. FSD was defined according to the revised DSM-IV definitions or AFUD classification and included hypoactive sexual desire disorder (HSDD), female sexual arousal disorder (FSAD), female orgasmic disorder (FOD) and sexual pain disorders or a subgroup or combination of these.\(^5\)\(^\text{6,13}\) A literature search for relevant studies was performed using PubMed from 1980-2005 (key words: female sexual dysfunction, menopause, hormones, estrogens, testosterone, sildenafil, DHEAS, androgens). Reference lists from the retrieved articles were also examined. Only the following studies were included for this review: double-blind randomized controlled trials of women identified with FSD based on specific questions and/or an interview by an expert in sexuality. The studies were required to have included specific and comprehensive measures of sexual function.

We excluded studies in surgically menopausal women being “assumed to have FSD” as a result of the rapid and sudden decline in sex-steroids after oophorectomy as ovariectomy does not necessarily result in FSD.\(^14\)\(^\text{16}\)

We defined the outcomes of interest as definition of the study population, inclusion and exclusion criteria, use of a power calculation, outcome measures used (questionnaires and hormone measurements) and treatment duration.

**Results**

A total of 25 double blind randomized controlled trials investigating the effects of pharmaceutical intervention on female sexual functioning have been published.\(^14\)\(^\text{17-40}\) Of these, 11 studies were considered to meet the criteria for inclusion in this review and are listed in Table 1.

Four studies investigated the effects of sildenafil on sexual arousal disorder or combined sexual arousal and desire disorders\(^28\)\(^\text{31}\) and seven studies investigated the effects of various testosterone preparations on sexual desire disorder and/or "low libido".\(^32\)\(^\text{38}\) A recent study of Davis et.al was not included in spite of meeting the selection criteria for this review as this study was a rather mechanistic study investigating the effects of aromatase inhibition on testosterone treatment effects on FSD.\(^40\)
Defining the target population

As recognized screening tools do not exist for FSD, the trial populations for the included studies were defined by non-standardized methods resulting in heterogeneity of the populations studied, with possible effects on trial outcomes. Shifren et.al. included surgically menopausal women with “impaired sexual function” based on 3 qualitative questions. All questions if answered affirmatively could indicate FSD, but these questions were not diagnostic for FSD. To overcome this problem, women also completed the Brief Index of Sexual Functioning for women (BISF) and were included for participation only if their composite score was less than 33.6, which is the mean value for normal women. As there is a broad normative range of sexual functioning, a composite score below the mean score for normal women does not confirm FSD.

Some but not all of the subsequent studies included, in addition to a non-standardized interview, a variety of questionnaires assessing the degree of sexual function/dysfunction and the degree of sexuality related personal distress. Warnock et.al. defined the study population by a score below 8 on the Sexual Desire/Interest subscale of the Changes in Sexual Functioning Questionnaire (CFSQ-F-C) and Goldstat et.al. included only women with diminished sexuality according to the Sabbatsberg Self-rating Scale (score ≤ 42).

In the testosterone patch studies, surgically menopausal women with hypoactive sexual desire disorder (HSDD) were included. Women were identified if they had experienced a meaningful loss of sexual desire, a decrease in sexual activity after surgery and being bothered/concerned by this decrease in sexual desire. Basson et. al. aimed to include women specifically identified with genital sexual arousal disorder based on a semi-structured interview.

None of the studies defined their populations by the amount of sexuality related personal distress. Only four studies included sexuality-related personal distress as an outcome measurement: Basson et.al used a 4 point scale, rating from satisfied to distressed, to measure the level of personal distress with orgasm. In the studies of Buster, Simon, Braunstein the Personal Distress Scale (PDS-7 items) was used as an outcome measurement.

Inclusion and exclusion criteria

Inclusion criteria of the trials listed in table 1 were dependent upon the defined target population, that is, the specific sexual disorder (FSAD, HSDD, or “broad spectrum FSD”) under investigation.

Age and menopause status:

Four trials included women with considerably wide age limits ranging from 24 to over 70 and two trials have included women with both natural and surgical menopause. Basson et.al. included a combined population of premenopausal and postmenopausal women. The results were reported separately for estrogenized (pre-and postmenopausal women, age 18-58) and estrogen deficient (non HT suppled) women.
Table 1. Randomized controlled intervention trials in FSD

<table>
<thead>
<tr>
<th>Author</th>
<th>Study</th>
<th>Treatment Time</th>
<th>N</th>
<th>Dose</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caruso et al.</td>
<td>RCT placebo cross-over vs sildenafil</td>
<td>Three 4-week periods</td>
<td>53</td>
<td>Sildenafil 25 mg and 50 mg</td>
<td>Fertile women 22-28 years</td>
</tr>
<tr>
<td>Basson et al.</td>
<td>RCT double blind placebo vs sildenafil</td>
<td>12 weeks after a 4-week run in</td>
<td>781</td>
<td>Sildenafil 10 mg, 50 mg, 100 mg</td>
<td>577 estrogenized women (pre-and postmenopausal) age 18-55 and 204 estrogen deficient women age 45-70 years</td>
</tr>
<tr>
<td>Basson et al.</td>
<td>RCT double blind placebo cross over vs sildenafil</td>
<td>1 hour prior to vibro and audiovisual erotic stimulation</td>
<td>34</td>
<td>Sildenafil 50 mg</td>
<td>Postmenopausal estrogenized women 40-78 years</td>
</tr>
<tr>
<td>Berman et al.</td>
<td>RCT double blind placebo vs sildenafil</td>
<td>12 weeks after a 4-week run-in period</td>
<td>202</td>
<td>Sildenafil 50 mg</td>
<td>Postmenopausal women 30-74 years</td>
</tr>
<tr>
<td>Shifren et al.</td>
<td>RCT placebo vs testosterone</td>
<td>12 weeks after a 4 week run in period</td>
<td>75</td>
<td>Testosterone matrix patch 150 μg or 300 μg daily</td>
<td>Estrogenized surgically postmenopausal women 31-56 years</td>
</tr>
<tr>
<td>Lobo et al.</td>
<td>RCT double blind placebo vs methyl testosterone</td>
<td>16 weeks after a 2-week run in period</td>
<td>218</td>
<td>Methyltestosterone 1.25 mg p.o. daily</td>
<td>Estrogenized postmenopausal women 40-65 years</td>
</tr>
<tr>
<td>Goldstat et al.</td>
<td>RCT placebo cross over vs testosterone</td>
<td>12 weeks+ 12 weeks</td>
<td>49</td>
<td>Testosterone 1% cream 10 mg daily</td>
<td>Premenopausal women 30-45 years</td>
</tr>
<tr>
<td>Warnock et al.</td>
<td>RCT double blind placebo vs methyl testosterone</td>
<td>8 weeks</td>
<td>102</td>
<td>Methyltestosterone 2.5 mg p.o daily</td>
<td>Estrogenized surgically postmenopausal women 33-62 years</td>
</tr>
<tr>
<td>Buster et al.</td>
<td>RCT placebo vs testosterone</td>
<td>24 weeks after a 8 week run in period</td>
<td>533</td>
<td>Testosterone matrix patch 300 μg twice weekly</td>
<td>Estrogenized surgically postmenopausal women 20-65 years</td>
</tr>
<tr>
<td>Braunstein et al.</td>
<td>RCT placebo vs testosterone</td>
<td>24 weeks</td>
<td>447</td>
<td>Testosterone matrix patch 150, 300 and 450 μg twice weekly</td>
<td>Estrogenized surgically postmenopausal women 24-70 years</td>
</tr>
<tr>
<td>Simon et al.</td>
<td>RCT placebo vs testosterone</td>
<td>24 weeks after a 8 week run in period</td>
<td>562</td>
<td>Testosterone 300 μg twice weekly</td>
<td>Estrogened surgically postmenopausal women 26-70 years</td>
</tr>
</tbody>
</table>

FSAD: Female Sexual Arousal Disorder, HSDD: Hypoactive Sexual Desire Disorder, FOD: Female Orgasm Disorder, VPA: Vaginal Pulse Amplitude, PEQs: Personal Experience Questionnaire, GEG: Global Efficacy Questions (“Did treatment improve physical response during sexual activity?” and “Did treatment improve ability to participate in sexual intercourse?”), LSC: Life Satisfaction Questionnaire, DSFI: Derogatis Sexual Functioning Inventory,
<table>
<thead>
<tr>
<th>Inclusion</th>
<th>Outcome Measures</th>
<th>Sexual behaviour</th>
<th>Adequate Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSAD</td>
<td>PEQS</td>
<td>↑ arousal, orgasm, enjoyment, frequency of sexual fantasies and intercourse</td>
<td>No</td>
</tr>
<tr>
<td>Lubrication problems, FSAD, HSDD, FOD or combinations Score ≥5 on LSC</td>
<td>2 Global Efficacy Questions</td>
<td>No change in sexual response</td>
<td>Yes</td>
</tr>
<tr>
<td>i.e. satisfactory sex life</td>
<td>LSC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SFQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSAD and impaired orgasm</td>
<td>Orgasm latency time</td>
<td>No effect on orgasm latency and intensity.</td>
<td>No</td>
</tr>
<tr>
<td>Various levels of distress by orgasmic functioning (rating from satisfied</td>
<td>Orgasm intensity</td>
<td>No change in subjective sexual arousal or genital arousal. Low VPA predictive for</td>
<td></td>
</tr>
<tr>
<td>to distressed)</td>
<td>DSI</td>
<td>orgasmic response on treatment</td>
<td></td>
</tr>
<tr>
<td>Minimal serum levels of estrogen and testosterone required (no estrogen</td>
<td>VPA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and/or androgen insufficiency)</td>
<td>Beck anxiety inventory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary diagnosis FSAD, secondary HSDD allowed</td>
<td>FIEI</td>
<td>In women without concomitant HSDD:</td>
<td>Yes</td>
</tr>
<tr>
<td>Minimal serum levels of estrogen and testosterone required (no estrogen</td>
<td>SFQ</td>
<td>↑ genital sensation during intercourse/stimulation.</td>
<td></td>
</tr>
<tr>
<td>and/or androgen insufficiency)</td>
<td>Sexual event logs</td>
<td>↑ satisfaction with intercourse and/or foreplay.</td>
<td></td>
</tr>
<tr>
<td>Impaired sexual function after hysterectomy/oophorectomy</td>
<td>BISF-W</td>
<td>No change in number of satisfying sexual events</td>
<td>No</td>
</tr>
<tr>
<td>HSDD with the onset of menopause. Score ≤ 3 on the thoughts/</td>
<td>MSIQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>desire dimension BISF Deficit of ≥2 points from premenopausal level of</td>
<td>BISF-W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>desire on SIQ</td>
<td>Serum hormone concentrations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low libido (SSSR scale score &lt; 42), BDI-I score &lt; 28 (no depression),</td>
<td>SSSR Scale</td>
<td>↑ in composite scores PGWBI and SSSR</td>
<td>No</td>
</tr>
<tr>
<td>early morning testosterone serum level &lt; 2.2nmol/L</td>
<td>PGWBI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSDD with the onset of menopause. Score ≤ 8 on sexual desire</td>
<td>CSFQ-F-C</td>
<td>No change in CSFQ-F-C scores</td>
<td>No</td>
</tr>
<tr>
<td>/interest dimension of the CSFQ-F-C and a SES score ≤ 3</td>
<td>CSES</td>
<td>↑ subscore MSlQ (sexual desire/interest)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSlQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WHQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Serum hormone concentrations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSDD with the onset of surgical menopause</td>
<td>SAL</td>
<td>↑ satisfying sexual activity</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>PFSF</td>
<td>↑ sexual desire</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PDS</td>
<td>↑ personal distress</td>
<td></td>
</tr>
<tr>
<td>HSDD with the onset of surgical menopause menopause BDI-II score ≥ 14</td>
<td>SAL</td>
<td>↑ satisfying sexual activity</td>
<td>Yes</td>
</tr>
<tr>
<td>(no depression)</td>
<td>PFSF</td>
<td>↑ sexual desire</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PDS</td>
<td>↑ personal distress</td>
<td></td>
</tr>
<tr>
<td>HSDD with the onset of surgical menopause. BDI-II score ≥ 14 (no</td>
<td>SAL</td>
<td>↑ satisfying sexual activity</td>
<td>Yes</td>
</tr>
<tr>
<td>depression)</td>
<td>PFSF</td>
<td>↑ sexual desire</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PDS</td>
<td>↑ personal distress</td>
<td></td>
</tr>
<tr>
<td>SFQ: Sexual Function Questionnaire, FIEI: Female Intervention Efficacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>index, BISF-W: Brief Index Sexual Functioning-Women, MSlQ: Sexual Interest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaire, SSSR: Sabbatsberg Sexual Self Rating scale, PGWBI:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological General Well-Being Index, BDI: Beck Depression Inventory,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSFQ-F-C: Changes in Sexual Functioning Questionnaire-Female, CSES:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Changes Sexual Energy Scale , WHQ: Women's Health Questionnaire</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Duration of the relationship
All the listed studies included only women reporting to be in a stable sexual relationship (of at least 6 months to 1 year).

Frequency of sexual activity
All subjects included in the identified studies were required to have a minimal frequency of sexual activity of at least once a month. Sexual activity prior to the trial was not restricted to a maximum, so women may have been included with a sexual activity of more than once a day during the pre-treatment period.

Concurrent health conditions and disallowed medication
Exclusion criteria in the reviewed trials were comparable and contained the following: sensitivity/contraindications to estrogens and/or androgen therapy, use of androgens in the preceding 3 months, a history of estrogen related cancer, illnesses known to influence sexual function (psychiatric disorders, neurological disease, gynecological cancer), endocrine disorders with systemic disease which would impair overall health and well being (treated hypothyroidism or hyperthyroidism usually allowed), uncontrolled hypertension, sexual or phobic disorders unrelated or related to sexual abuse, medications known to interfere with the trial medication, medications known to impair sexual function (anti hypertensive drugs including $\beta$-blockers, anti-depressants), body mass index (BMI) within certain limits, dyspareunia as the primary cause of sexual dysfunction, physical limitations interfering with normal sexual function, angina pectoris and or myocardial infarction in the past (sildenafil trials) and lipid lowering drugs (androgen trials). Some of the trials included women with a BMI above 30 kg/m$^2$ with a maximum of BMI 35 kg/m$^2$.

Concomitant use of estrogen containing preparations
In the studies of Buster$^{36}$ and Simon$^{38}$, women were stratified by oral or transdermal estrogen use. The majority of subjects used oral estrogens (75% in the Simon study and 80% in the Buster study). Goldstat$^{34}$ included women using oral contraceptives (OAC), but the results were not stratified by OAC use.

Power Calculation
Most but not all of the trials included in this review provided a primary objective and some form of sample size calculation. Only three studies based their sample size calculation on a predefined minimal difference in effect between treatment-groups considered to be clinically relevant.$^{36-38}$ These studies included the number of satisfactory sexual events as (co)primary endpoint. The power calculation was based on a difference of 0.34 sexual events per week between the placebo arm and the treatment arm. Caruso$^{28}$ based his sample size calculation on differences found on the erectile dysfunction score in a trial in men.

Other studies included as the primary efficacy parameter the change from baseline on certain scales or subscales (domains) of sexual functioning. In these studies, an adequate power calculation was either not described or was not based on the methods.
recommended by Cohen. This is important as most instruments used in these trials were not developed for the measurement of a treatment effect. Scales measuring complex behavior, like sexual functioning, have by definition a low predictive power and briefer questionnaires further compound this problem by their relatively low reliability. Thus only a small to medium effect size for total or composite scores on the sexual function scales can be assumed.

**Outcome Measures**

*Patient reported outcomes*

The selection of self-administered sexuality questionnaires included in the reviewed publications is included in Table 1.

Various instruments to assess the level of sexual functioning are currently available and several are considered to be comprehensive and useful and recommended for inclusion in FSD trials. The following questionnaires have demonstrated acceptable reliability and validity: the Golombok Rust Inventory of Sexual Satisfaction (GRISS; 1987), the Brief Index of Sexual Functioning (BISF; 1994), the Changes in Sexual Functioning Questionnaire (CSFQ; 1997), the Derogatis Interview for Sexual Functioning (DISF-SR; 1997), the Index of Female Sexual Function (IFSF; 1999), the Female Sexual Function Index (FSFI; 2000), the Female Intervention Efficacy index (FIEI; 2001), the Sexual Function Questionnaire (SFQ; 2002) and the Female Sexual Distress Scale (FSDS; 2002).

The Profile of Female Sexual Function (PFSF; 2004) which is used in three of the included trials was developed very recently and is thus not (yet) included in the recommendations for clinical trials in FSD. However, it shows robust psychometric properties including good inter-item correlations, discriminating ability, test/retest reliability, internal consistency and construct validity. The Menopausal Sexual Interest Questionnaire (MSIQ; 2004) was used as one of the primary efficacy outcome measures in the trials of Lobo and Warnock prior to having proven its sensitivity and construct validity. Subsequently, the validity of the MSIQ has been published.

Of note: all instruments are still based on the traditional linear model of female sexual function underlying the DSM-IV-R classification which depicts discreet phases and presumes initial or spontaneous desire and ignores subjective arousal. Most questionnaires do not provide clear cut-off scores discriminating women with FSD from women without FSD. Of the available instruments, only the BISF, the SFQ, the FSDS, the FSFI, the FIEI and the PFSF have been developed, or subsequently validated, to be sensitive to a treatment effect.

Of the self-reported questionnaires used in the reviewed studies, only the SFQ was included as a secondary endpoint in phase II studies in order to assess whether a treatment effect could be measured with the scale prior to use it in larger phase III trials as a (co)primary endpoint.

For the other relatively new questionnaires, the intervention study itself was used to investigate whether the scale would respond to change and to further validate the questionnaire (inclusion as secondary endpoint).

Some questionnaires were specifically developed to measure certain aspects of sexual dysfunction like the SFQ (Female Sexual Arousal Disorder) and the PFSF (Hypoactive
Sexual Desire Disorder). These questionnaires are not automatically applicable for trials evaluating a broad spectrum of FSD.

Subject satisfaction with the sexual relationship is only addressed in the BISF and FSFI. The PFSF is the only questionnaire addressing subjective experience of being receptive to sexual cues other than to sexual activity only. None of the questionnaires includes questions pertaining to subject satisfaction with a particular result after treatment. The frequency of successful satisfactory sexual events, one of the FDA’s recommended primary endpoints, was included in 4 of the 11 reviewed trials\textsuperscript{31-38} and was measured by a one week recall diary, the Sexual Activity Log (SAL).\textsuperscript{59}

**Hormone Measurements**

As depicted in table 1, all studies of testosterone therapy in this review have included serum hormone measurements to monitor safety. One study used baseline estradiol and androgen measures as inclusion criteria to exclude androgen and/or estrogen insufficiency prior to being randomized to sildenafil.\textsuperscript{31} Two studies reported significant positive correlations between changes in serum total, bio-available and free testosterone levels at week 24 and an improvement in measures of sexual function.\textsuperscript{36,38} In the testosterone patch studies total testosterone was measured by radioimmunoassay, bio T by ammonium sulfate precipitation technique, free T by equilibrium dialysis and SHBG by immunoassay.

**Study duration and time to treatment effect**

For sildenafil, the treatment duration in the studies included varied from 1 hour prior to sexual activity (erotic video) to 12 weeks. For testosterone, the treatment duration varied from 8 weeks to 24 weeks.

If a significant improvement in sexual functioning was identified during the trial period, this was mostly observed after at least 12 weeks of therapy. In the studies of Simon\textsuperscript{38} and Buster\textsuperscript{36}, a statistical significant difference in the frequency of satisfactory sexual events and sexual desire between the two study arms was found after 5 (Buster) to 8 (Simon) weeks and was maintained over the total study period (24 weeks).

Both the study of Shifren\textsuperscript{32} as well as the study of Goldstat\textsuperscript{34} found a significant improvement at 12 weeks. Mixed results were found in the study of Warnock\textsuperscript{25}, which can be partially due to the treatment period which was only 8 weeks for the total study period. In the study of the Lobo\textsuperscript{33}, a statistical significant improvement in sexual desire over placebo was only found after 16 weeks.

**Discussion**

In the past 6 years, both expert groups and authorities have given guidance to clinicians and researchers how to define FSD (including a new classification system) and how to perform clinical trials in FSD.\textsuperscript{6,7,9,11} As the field of FSD is an evolving area of research, continuous adaptations and revisions to definitions and assessment measures have
been made parallel to the conduct of several major trials which in turn have increased our understanding of female sexual function. This review has identified a number of limitations and issues in drug intervention studies in FSD which should be taken into consideration by anyone doing future research in FSD and which might be addressed in updated guidelines about how to perform intervention trials in FSD.

Defining the population

In spite of a better understanding of female sexual function, the diagnosis of FSD is still very challenging and remains difficult due to its complex nature including several subjective aspects and contextual factors. There is no internationally accepted screening tool for FSD such that the diagnosis of FSD is based on non-standardized interviews by an expert in the field. In clinical research this can become a problem as due to multiple sites participating, not every center has the same expertise. Recently, cut-off scores to discriminate between normal sexual function and FSD women have been provided for the SFQ and FSFI. In an effort to standardize the diagnosis of FSD and to prevent a wide variability of subjects included in clinical trials, a structured diagnostic method (SDM) has been developed. The SDM can be used by trained interviewers to consistently, and when compared to experts, correctly diagnose FSD status. The SDM consists of four self-reported questionnaires, followed by a structured face-to-face interview. Despite showing high convergent validity and intrarater reliability, this method needs further validation in larger and different patient populations. For sexual functioning, specifically the SFQ and FSDS are included in the SDM. Of note, the SFQ is designed to measure a treatment effect of interventions with a focus on genital arousal disorder.

The FDA draft guidelines require that for inclusion in clinical trials, women must have sexuality related personal distress and can be assessed by the Female Sexual Distress Scale (FSDS) or another validated scale. However, none of the studies in this review used sexuality personal distress as an inclusion criterion. One of the reasons might be that it is difficult to determine what should be the level of distress for an individual woman especially if recruited for a trial, where women are approached via advertisements and not spontaneously seeking medical help for a problem. Not all women who are concerned about their level of sexual desire or arousal have associated feelings of guilt and inferiority (factors in the FSDS) in relation to their sexual concerns. The criterion is somewhat odd as men with erectile dysfunction (ED) are not required to demonstrate distress about their ED to merit therapy. In daily practice it should suffice if a woman is sufficiently concerned about her sexual function that she seeks medical advice.

The selection of questionnaires for the SDM is based on the AFUD classification of 2000 which is still based on the traditional linear model of female sexual response. Since then, various papers have been published which have recommended changes and recently the AFUD revised her classification system which will influence the design of future questionnaires and diagnostic methods. To define FSD for inclusion in clinical trials, we recommend reconsidering the selection of (new to be developed) questionnaires included in the SDM based on the current insights about female sexual functioning.
Inclusion and exclusion criteria

Heterogeneity in the study population of FSD trials may influence the outcomes. Allowing broad age limits may complicate the detection of small treatment effects as there is a normative broad range of sexual functioning and a normative lowering with age which causes large variation in sexual functioning at baseline. In addition, the clinical issues and sexual health needs for young women are likely to be very different from those of older women and this is likely to be further complicated by including young women who have undergone a surgical premature menopause, in whom sexuality might be influenced by other factors such as the premature loss of fertility. The length of the relationship is predictive for sexual behavior with recent change of partner being a positively influence. Thus women included in trials with a relationship of more than 10 years in duration are likely to respond differently to intervention for their sexual functioning than women who have had a more short term relationship and report dissatisfaction with their sexlife.

In contrast to men with ED, women will continue to be sexually active even though they lack desire or arousal. Thus a minimal frequency of sexual activity at screening has been required for inclusion in most FSD trials, where as few trials have had an upper limitation. This requirement is based on the assumption that women who have abstained from sexual activity because they have lost interest are unlikely to recommence activity with therapy. There is no data to support this assumption. It is also unclear whether women with a greater baseline frequency of activity are more responsive or refractory to therapy. This needs to be investigated more carefully.

Including women who are obese (BMI greater than 30 kg/m²) warrants consideration as these women may have the metabolic syndrome and metabolize sex steroids differently when compared to women with a lower BMI. To circumvent this problem a range of SHBG for inclusion should be considered.

Unfortunately, little is known about the influence of many prescribed and non prescribed medications on sexual functioning. Most information is derived from experience from trials of male ED. It is questionable if causalities for ED are applicable for FSD. For instance, the use of thiazide diuretics is highly associated with ED and the use of this drug has not been an exclusion criterion in trials for women. The more selective β1-blockers are not associated with ED [2nd international Consultation on Sexual Dysfunctions-Paris 2003]. From clinical experience, it is assumed that narcotics, including the regular use of codeine (e.g. in flu tablets) influences sexual functioning, however, this has not been a standard exclusion criterion in the reviewed trials. From a clinical practical point of view, it is difficult, and may be inappropriate to exclude all women who use medication which is very common in certain age groups (like anti-hypertensive drugs, i.e. β-blockers) as trial outcomes will then never be applicable to the more general population.

The role of the use of hormone therapy (HT) prior to inclusion in the trials has not been clearly defined. Estrogens appear to be important for normal sexual functioning not only for maintaining structural and functional integrity of the vaginal tissues but also for their central brain effects. As a result, there is a prevailing belief that the diagnosis FSD should ideally only be made if a woman is adequately estrogenized. However, this is based on
expert opinion and not on research data. If studies currently underway demonstrate that testosterone therapy in postmenopausal women not on concurrent estrogen is both effective and safe, the issue of requiring a woman to be estrogenized before diagnosing or treating FSD will need to be seriously reviewed.

After menopause, sexual functioning can be negatively influenced by a mixture of climacteric symptoms and it is likely that an improvement in general well-being achieved with symptom amelioration by HT, will positively influence sexual functioning. In clinical trials assessing the efficacy of a certain treatment on sexual function this is important as a real placebo arm will quickly unblind the trial.

On the other hand, oral HT increases SHBG which in turn decreases bio-available testosterone, which might negatively influence sexual functioning. There is also data to suggest that oral conjugated equine estrogen therapy (CEE) interferes with the efficacy of exogenous testosterone. All trials included in this review involved women either with normal estrogen levels or being adequately estrogenized prior to inclusion and many involved women taking oral CEE. It can be discussed whether the oral contraceptives act similar as anti-androgens because they increase SHBG levels resulting in less bio-available testosterone.

**Treatment effect**

The impetus for treating women who complain of decreased sexual desire or arousal with pharmaceutical interventions like sildenafil or testosterone comes from the assumption that these compounds play a similar role in women’s sexuality as it does in men’s. However, female sexual experience is mediated to a certain extent by expectations which are the product of prevailing societal values, familial values and partner expectations. So it can be questioned what exactly is measured in trials attempting to determine the extent to which a single factor, such as drug intervention, modulates sexual function as long as sexual functioning is not placed in an adequate female sexual bio-psycho-social model that includes the role of androgens.

In four trials included in this review for which power was inadequate, (partially) negative results have been found. This does not necessarily exclude an association between the treatment and a parameter of sexual function investigated but may be the result of type 1 error (sample size not large enough to detect a minimal difference).

In a small lab based trial including women diagnosed with genital arousal disorder based on a clinical assessment only, no treatment effect could be found. Also the large trial of Basson et.al., including women with broad spectrum FSD, including FSAD, did not find any treatment effect. Apart from the heterogenous population in the latter trial, which has most likely contributed to the negative results, another explanation might be that the relationship between subjective sexual arousal and genital sexual arousal is poor and that the overlap between desire and arousal may require desire to improve for arousal to improve. This might be due to the fact that a woman’s subjective experience of sexual arousal is determined less by feedback from her genitals and more by the intensity and appraisal of the sexual stimulus.

The diagnosis genital arousal disorder can only be made if certain conditions, like the presence of adequate sexual stimulation, are met. It is virtually impossible for a clinician
to judge this based on an interview only and the diagnosis genital arousal disorder can possibly only be made under laboratory conditions (erotic video and VPA), where the photoplethysmograph might be useful in further characterization of the genital arousal disorder.\textsuperscript{30,77} In clinical practice, the place of vaginal plethysmography is unclear and in most doctor’s offices not available.

In seven trials included in this review\textsuperscript{28-30,32-35} an improvement in certain aspects of female sexual function was found based on self administered sexuality questionnaires. Whether a statistically significant improvement is clinically meaningful remains a topic of discussion for future trials as currently, a minimal important difference in effect size has not been defined yet for intervention trials in FSD. Another limitation is that the outcomes of most trials in this review are concluded from questionnaires still based on the traditional linear sexual response model. For instance, Shifren et.al.\textsuperscript{32} measured an “improvement in sexual functioning,” Basson et.al.\textsuperscript{30} found an “increased subjective sexual arousal and a reduced latency to orgasm” and de Lobo\textsuperscript{33} found an increase in sexual desire and interest scores. In spite of the results being interesting and contributing to our understanding of female sexual functioning, it is unclear what these outcomes mean to the individual patient. First of all, a woman’s motivation for sexual activity is frequently for reasons other than sexual desire and/or sexual thoughts and fantasies and its absence does not equate to dysfunction.\textsuperscript{6,78} Second, emotional well-being and positive emotional responses during sexual activity are reported to contribute more to sexual satisfaction than physical or genital aspects of sexual response.\textsuperscript{70,78} So, to maintain a focus on initial or spontaneous sexual desire, fantasies, thoughts when assessing sexual function/satisfaction is not evidence based.

Measuring improvements of sexual functioning in questionnaires showing an increased sexual desire and/or arousal gives no full answer on whether a treatment has been efficacious or not as a higher or lower degree of sexual desire does not form a measure for sexual satisfaction.

The later published studies\textsuperscript{36-38} all used a combination of Self Administered Questionnaires (SAQ’s) to measure qualitative aspects of sexuality and an event log, like the Sexual Activity Log (SAL), to measure the frequency of successful and satisfactory sexual events. Using the frequency of satisfactory sexual events as an endpoint, the woman herself reports, based on a personal judgment, if she has experienced a certain sexual event (including oral sex, masturbation, intercourse, orgasm) as being satisfactory for her or not.

As this is a subjective assessment, it does not matter what has determined her satisfaction with the event, whether it is due to the improvement in desire or other for instance sexual relationship related or contextual reasons. As in most studies in this review overall sexual satisfaction and satisfaction with the sexual relationship was not evaluated it might be useful, however, to measure patient satisfaction with the sexual relationship and with a particular treatment in future trials.

A limitation of the use of a diary or event log measure is that they are typically completed at home and susceptible to various forms of response bias or error. Other authors
have commented on this and provided suggestions how to improve compliance issues with diaries by using more technically sophisticated methods like electronic diaries or phone call-in methodologies. In addition, diary data are restricted to the scope of measurement and do not provide the broad multidimensional assessment of the response.

A combined approach (using both SAQ’s and diaries) seems to be a valid way to assess overall treatment effects of pharmaceutical intervention as both subjective and qualitative aspects of sexuality like desire and arousal are addressed and also the effect of these aspects on sexual function is measured.

For practicing physicians, it would be helpful to understand what a particular significant result (increase in score) means for clinical practice as most trials in this review used different questionnaires which are difficult to compare in the clinical setting.

To inform both women and health care practitioners better what they can expect from a certain treatment, we suggest that a measure like “what percentage of women is satisfied with a specific result or how many of the treated women experienced a 50% or greater improvement in their satisfaction with sexual activity and/or relationship”. Goldstat et.al elegantly described this in her paper by translating the statistical significant improvement of 15.7 units on the Sabbatsberg Sexual Self Rating Scale to terms of meaning for clinical practice which is: 46% of the women experienced a 50% or greater increase in their total sexual self rating score with testosterone treatment.

Another important issue is the cultural acceptance and understanding of questionnaires. Most questionnaires used in the described trials are validated in women in English speaking countries only (U.S, U.K, and Australia). The PFSF did undergo linguistic validation in several European languages. However in addition to linguistic validation, cultural applicability must also be determined, particularly for application in non-western cultures as beliefs with respect to “normalcy” of certain sexual behaviors are highly sensitive to societal beliefs.

For future trials a universally accepted selection of questionnaires and event logs should be used which are based on the current insights and understanding of female sexual functioning.

The patient reported outcome measures (PRO’s) should be able to measure a treatment effect, not only in frequency but also in intensity and should be applicable for different cultures. A minimal important difference that should be seen as change over time from baseline between the placebo arm and treatment arm and time to expected treatment effect should be discussed upfront between expert groups and authorities.

**Hormone measurements**

Although it is widely accepted that sex hormones exert an important influence on sexual functioning, the biomechanism(s) by which hormones operate is poorly understood. There is no evidence that measurement of steroid hormones assist with the diagnosis of FSD. Measurement of hormonal levels are primarily indicated to monitor therapeutic safety. One of the difficulties in assessing androgen serum levels is that the range of testosterone to measure in women is very narrow and distinguishing
significant differences is difficult. Free T by equilibrium dialysis is regarded as the gold standard whereas calculation of free T (or calculation of free androgen index) by measuring total T after organic solvent extraction or RIA has been validated in several reports and is recommended as a possible option as well.\textsuperscript{80-82}

As intracrinology plays a pivotal role in androgen metabolism, it is likely that behavioral aspects of androgens by tissue metabolism cannot be measured peripherally. It is possible that androgen derivatives measured in serum or urine, are a better reflection of what is happening at a tissue level. This is a subject for future investigation.

Conclusion

In order to be able to measure a treatment effect of the intervention which is clinically meaningful for a woman the following issues need to be resolved prior to initiating new intervention trials in FSD.

A universally accepted method for defining the FSD population is needed and as a result a consensus should be reached for appropriate inclusion and exclusion criteria for FSD trials. Similarly main outcome measures appropriate for the evaluation of drug intervention should be defined. Trial endpoints need to reflect clinically relevant outcomes and should be embedded in a bio-psycho-social sexual response model for women.
References


74. Hill DA, Weiss NS, LaCroix AZ. Adherence to postmenopausal hormone therapy during the year after the initial prescription. AJOG 2000; 182(2): 270-276.


