Journal of Sex & Marital Therapy, 31:153–166, 2005

ISSN: 0092-623X print
DOI: 10.1080/00926230590909989

Routledge
Taylor & Francis Group

Outcome Measurement in Female Sexual Dysfunction Clinical Trials: Review and Recommendations

STANLEY E. ALTHOF, Ph.D.
Professor of Psychology, Department of Urology, Case Western Reserve University School of Medicine, Director, Center for Marital and Sexual Health of South Florida
RAYMOND C. ROSEN, Ph.D.
Professor of Psychology, Department of Psychiatry, Robert Wood Johnson Medical School, Piscataway, New Jersey, Center for Sexual and Marital Health
LEONARD DEROGATIS, Ph.D.
Director, Johns Hopkins Center for Sexual Health and Medicine
ERIC CORTY, Ph.D.
Associate Professor of Clinical Psychology, School of Humanities and Social Sciences Penn State Erie, The Behrend College
FRANCES QUIRK, Ph.D.
Senior Lecturer, Schools of Psychology and Medicine, James Cook University, Townsville, QLD
TARA SYMONDS, Ph.D.
Worldwide Outcomes Research, Sandwich Laboratories, Pfizer Ltd, Kent, UK

Defining and measuring Female Sexual Dysfunction (FSD) is a complex and challenging task. Several factors have confounded the theory and measurement of FSD including: the use of an inappropriate male paradigm; difficulty in capturing the complexity of women's sexual response; an evolving but presently untested nosology; and the relative independence between subjective and objective aspects of women's sexual response. Each of these factors have contributed to the difficulty in developing meaningful and valid endpoints for clinical trials.

The Food and Drug Administration's (FDA) 2000 draft guidance document for female sexual dysfunction clinical trials

We gratefully acknowledge Cathy Spino, D.Sc. for providing the authors with the statistical analyses of Pfizer's Viagra and Lasofoxefine FSD trials.

Address correspondence to Stanley E. Althof, Ph.D., Center for Marital and Sexual Health of South Florida, 1515 N. Flagler Drive, Suite 540, West Palm Beach, Florida 33401. E-mail: sxa6@po.cwru.edu

recommended the use of daily diary measures as primary and self-administered questionnaires (SAQs) as secondary endpoints. Event logs or diary measures may be adequate for assessing aspects of male sexual performance (e.g., erectile function), or in other therapeutic areas with discrete and readily observable endpoints (e.g., incontinence). However, psychometric theory suggests that for female sexual dysfunction clinical trials, SAQ instruments may provide more sensitive and reliable measures of outcome. We offer an alternative set of recommendations in the hope that the FDA will reconsider its position and to serve as potential guidelines for non-industry sponsored research on female sexuality as well.

First, we propose that SAQs be elevated from their current status as secondary endpoints to be considered as potential primary endpoints in clinical trials of FSD. Second, we recommend that depending on the trial design and intervention under study, either an SAQ or diary measure (typically one or the other, and not both), might serve as a primary endpoint in a clinical trial. Third, SAQs and diaries should be employed, analyzed and interpreted in their particular areas of strength. Diaries are most useful for enumerating events and/or counting frequencies. SAQs are superior at gathering subjective data related to women's sexual function. Fourth, we believe there is a theoretical basis for considering SAQs to be superior measurement tools compared to diaries in assessing sexual dysfunction in women. At present, however there is insufficient objective data to fully support this opinion. Conversely, we do not anticipate either theoretical or objective evidence to support the alternative hypothesis (that diaries are superior to SAQs). If this proves to be correct in the future, diary measures may no longer be considered as primary endpoints for FSD clinical trials. Finally, we recommend that the FDA and/or other regulatory agencies reconsider the emphasis given to the number of successful or satisfactory sexual events over time as primary endpoints because they do not definitively demonstrate whether there has or has not been any improvement in the FSD endpoint under study (e.g., sexual desire). Successful and satisfactory encounters represent an amalgam of subjective assessments that are too far removed from the essential FSD component.

Interest in developing treatments for female sexual dysfunction (FSD) has gained impetus from the impressive success of oral treatments for erectile dysfunction. This achievement fostered a more open climate for discussion of sexual problems and gave hope to many women who suffer from sexual

disorders that a similar solution would soon be available to them. To date, these expectations have largely not been realized.

A recent example is the Pfizer announcement on sildenafil (Viagra) for women. Despite an intensive research and development program involving several large-scale clinical trials, Pfizer announced in April 2004 that the company had discontinued its program of clinical research on sildenafil as a medical treatment for female sexual arousal disorder (FSAD), (Harris, 2004). Inconsistent results had been obtained both within and across studies (e.g., Basson et al., 2002a; Berman et al., 2003; Caruso et al., 2003). Several factors most likely contributed to this inconsistency and lack of efficacy, including the overlap between Hypoactive Sexual Desire Disorder (HSDD) and FSAD diagnoses, a high placebo response rate in most studies, and clinical endpoints which yielded inconsistent and at times contradictory results. These factors are equally problematic for non-industry sponsored studies as well.

Several cogent reasons may account for why treatments for female sexual dysfunction may not have followed the same trajectory as those for men. To begin with, there is widespread lack of agreement regarding the choice of model or paradigm for understanding women's sexual response. The Masters and Johnson linear model of male sexual response, i.e., that of sexual desire leading to arousal followed by orgasm, has come under criticism for its neglect of women's sexual priorities and experiences (Tiefer et al., 2002). New, more suitable non-linear models recently proposed make clear that female sexual response is not a mirror image of male response subjectively, objectively or developmentally (Basson, 2002b). Thus, if a theoretical model does not accurately reflect women's sexual responses, methodologies and outcome measures based on this model are unlikely to be successful in capturing the emotional/subjective and physiological/objective aspects of women's sexual response.

A second important factor is the lack of consensus regarding diagnostic definitions and classification of sub-types of FSD. Specifically, previous diagnostic classifications have failed to address both subjective and objective aspects of women's sexual response.

IMPACT OF NEW CLASSIFICATION OF WOMEN'S SEXUAL DYSFUNCTION

With increasing concerns about the DSM IV-TR (APA, 2002) or ICD-10 (WHO, 1989) diagnostic categories and their diagnostic accuracy or reliability, a consensus conference of experts was convened recently to make recommendations for change in the definitions of women's sexual dysfunction (Basson et al., 2004). In addition to proposals for definitional change and development

of new categories of dysfunction, the new revisions emphasize the essential differences between subjective and physiological aspects of women's sexual arousal disorder. Thus, the new category of women's sexual arousal disorder now includes three distinct sub-categories: (i) physical/genital sexual arousal disorder; (ii) psychological sexual arousal disorder; (iii) combined physical and psychological sexual arousal disorder (Basson et al., 2004). These new categories are defined in the consensus recommendations as follows:

PHYSICAL/GENITAL SEXUAL AROUSAL DISORDER

Complaints of absent or impaired physical sexual arousal. Self-report may include minimal vulval (clitoral), swelling or vaginal lubrication from any type of sexual stimulation and reduced sexual sensations from sexually caressing genitalia, breasts and other areas. (Subjective sexual excitement and sexual pleasure still occurs, typically from non genital stimulation)

PSYCHOLOGICAL SEXUAL AROUSAL DISORDER

Absence of or markedly diminished subjective sexual excitement and awareness of sexual pleasure from any type of sexual stimulation. (Vaginal lubrication or other signs of physical response still occur.)

COMBINED PHYSICAL AND PSYCHOLOGICAL AROUSAL DISORDER

Absence of or markedly diminished subjective sexual excitement and feelings of sexual pleasure from any type of sexual stimulation as well as complaints of absent or impaired physical sexual arousal. (Basson et al., 2004)

Contextual factors need to be taken into account in diagnosing the disorder, according to the proposed definitions. According to the new classification, women whose life circumstances may not be conducive to sexual interaction or enjoyment, such as exists with disordered relationships or economic poverty, may feel disinclined to enter into a sexual relationship or engage in specific sexual activity (Bancroft et al., 2003). Thus, the new classification is both more differentiated in describing response and more specific in assessing context than previous consensus definitions (Basson et al., 2002c), or the DSM-IV-TR (APA, 2002) or ICD-10 (WHO, 1989) diagnoses.

It is uncertain what effect, if any, the proposed new classification will have on outcomes assessment in clinical trials of FSD. On the positive side, it is possible that by categorizing women according to the proposed new nosology, diagnostic accuracy could be improved, leading in turn to improved treatment efficacy and more clinically relevant outcomes. New study measures or endpoints might also be developed to assess sub-category-specific outcomes, such as increased subjective arousal (assessed by SAQ or diary endpoints) in the absence of changes in physiological response (assessed by physiologic or self-report measures). Essentially, the decision on selection of study endpoints for research on women's sexual arousal should take into

account the potential sub-categories of diagnosis as proposed in the new classification, once they have been confirmed as valid.

What will be the impact ultimately of this new diagnostic sub-typing? It could be argued that more precise sub-typing, in and of itself, might lead to increased specificity in the assessment of treatment outcomes and accordingly, more precise measurement. It will certainly lead to reduction of within-groups error variance associated with diagnostic mis-assignments. On the other hand, if the categories themselves are not reliable or valid, the proposed new classification would not enhance outcome assessment, and could serve to expand both within-groups error and measurement error. Additionally, it is uncertain what impact, if any, this new classification will have on future regulatory or scientific recommendations in this area.

OBJECTIVE VERSUS SUBJECTIVE ASPECTS OF SEXUAL RESPONSE

Men and women differ markedly in their physiological responses and subjective awareness of sexual desire and arousal, necessitating assessment approaches tailored to each gender's unique pattern of sexual response (Rosen & Beck, 1988). Unlike men's visual and palpable awareness of erection, women's perceptions of genital arousal are more subtle and indistinct. Further complicating accurate measurement of sexual response, is the significant variability between women, particularly in their experience of bodily sensations. These physical sensations, even when consciously attended to; do not necessarily predict a woman's subjective sexual experience (Laan & Everaerd, 1995). In contrast, studies in men typically show substantial agreement between subjective and objective measures of arousal. Men have visual and tactile feedback of their arousal, and this may account for the high correlation observed between objective and subjective measures of male arousal. Men are able to report reliably the occurrence and relative firmness of erections, and these reports typically form the basis of event log or daily diary measures in male ED trials. This has not been accomplished to date for female arousal. Moreover, erectile function in men is a relatively uni-dimensional domain of response, whereas sexual arousal in women is generally viewed as a complex, multi-dimensional response.

Furthermore, hypoactive sexual desire disorder (HSDD) in both men and women is defined entirely according to subjective criteria. Although sexual desire may be associated with certain physiological or hormonal changes (e.g., decreased androgen) these changes are not reliable within or across individuals, and have not been used as efficacy measures in male or female studies of HSDD. Instead, studies in this area have relied on the individual's subjective assessment of their interest in or response to sexual activity (Lobo et al., 2003; Shifren et al., 2000).

In essence, evidence from numerous sources indicates the importance, or primacy, for most women of their *subjective experience*, whether or not this correlates with objective or physiological measures. Subjective experiences are intrinsically difficult to measure, require an appreciation of the context in which the response occurs, and are usually best captured in SAQs. Despite certain shortcomings, SAQs remain the cornerstone of measurement relative to the more complex aspects of human experience (Derogatis & Coons, 1993).

THE FDA'S DRAFT GUIDANCE FOR FSD CLINICAL TRIALS

With the advent of industry-sponsored clinical trials in female sexual dysfunction, the fundamental issue of delineating psychometrically adequate and clinically sensitive measurement endpoints becomes paramount. The FDA is the first agency to date to formulate draft guidance to serve as a basis for dialogue with sponsors/investigators in developing new medicines in this complex area. Professional groups, such as the International Society for Sexual Medicine and the International Society for the Study of Women's Sexual Health are also considering the development of research standards, including recommendations for endpoints, for female sexual dysfunction research.

The FDA's draft Guidance Document on FSD (2000) recommends the use of daily diary measures as primary endpoints, and self-administered questionnaires (SAQs) as secondary efficacy measures. Event logs or diary measures may provide sensitive and reliable measures of efficacy in studies of erectile dysfunction or premature ejaculation in men, or in other therapeutic areas with discrete and observable endpoints (e.g., incontinence). However, we present evidence based on psychometric theory that suggests for female sexual dysfunction clinical trials, SAQ measures such as the Sexual Function Questionnaire (Quirk et al., 2002) or Female Sexual Function Index (Rosen et al., 2000) are superior in gathering subjective, more multidimensional information than diaries or event logs. The approach employed in studies of male sexual dysfunction cannot be assumed to be readily transferable to women with the same level of reliability and sensitivity to treatment that has been evident in male trials of erectile dysfunction.

The lack of concordance between diary measures and SAQs has been noted in several large-scale studies in women. For example, in the widely-cited study by Shifren et al. (2000) of testosterone replacement in oophorectomized women, marked discrepancies were noted between a telephone-based, daily diary measure of sexual desire and results obtained from the Brief Index of Sexual Functioning (Taylor et al., 1994) self-administered questionnaire measure. In this study the telephone-based diary was found to be both more intrusive and less sensitive to treatment effects.

Compliance difficulties and missing data were additional problems with the diary measure (Shifren et al., 2000). The authors also reported that the high placebo response observed in younger women in the trial may have obscured positive treatment effects. These authors concerns about the use of daily diaries or event logs in clinical trials of FSD, e.g., intrusiveness, lack of compliance and potential insensitivity to treatment need to be carefully considered in planning future trials in this area.

In writing about clinical endpoints the FDA's draft guidance stated:

> "Primary endpoints for trials of drug products to treat FSD should be clinically meaningful and specifically related to the component or components of FSD being studied in the trials. These endpoints should be based on the number of successful and satisfactory sexual events or encounters over time. The determination of *successful* and *satisfactory* should be made by the woman participating in the trial, as opposed to her partner."

A theoretical basis may be lacking for the FDA's emphasis on the *number of successful and satisfactory sexual events or encounters over time* as being the most important primary outcome endpoint for all FSD clinical trials. While we agree that it is important to ascertain the number of successful satisfactory sexual events, this criterion by itself represents a one item/multiple concept question. By this we mean that there may be several related or independent factors that blend together under the rubric of either successful or satisfactory sexual events. Counting events is simple; what determines a woman's satisfaction or success with an event is less clear. Moreover, as this judgment is entirely subjective, it is more nuanced and best captured by an SAQ.

Additionally, the concept of satisfactory or successful sexual encounter is several steps removed from the components under study, e.g., desire, arousal, orgasm or pain. Women may experience significantly improved sexual desire or arousal; they may not necessarily choose to engage in sexual behaviors or, may not describe an event as satisfactory or successful for reasons that have little to do with their desire or arousal. Conversely, they may not experience any improvement in desire or arousal but report a successful or satisfactory event due to interpersonal or contextual reasons. What is being measured, successful or satisfactory sexual events, is too distal to the targeted FSD component.

Based on study findings to date and new conceptualizations of FSD, we question the rationale for the FDA's draft guidance recommendations on clinical endpoints for FSD trials. In the next section, we consider the psychometric advantages and major limitations of daily diary and SAQ methods, and conclude with a set of recommendations for revision of the current FDA draft guidance.

PSYCHOMETRIC THEORY

Measurement scales have various levels of precision. In ascending order of sophistication the four measurement scales are: nominal, ordinal, interval and ratio scales. *Nominal* scales are the most primitive or crudest form of measurement. Many measurement experts consider these as "premeasurement" devices, since their operations involve enumeration or counting of events. They are typically employed to categorize or label phenomena (e.g., male/female; presence/absence). Diaries are typically forms of nominal measurement asking yes/no type questions (e.g., Was your erection hard enough for penetration?).

Ordinal scales which, are more sophisticated, possess the characteristic of "greater than" and are essentially a mechanism to rank data (e.g., ranks in a beauty contest, best selling books). Rank order statistics can be applied to ordinal scales.

More sophisticated than ordinal measures are *interval scales,* which not only reflect an ordinal (greater than) relationship between numbers of the scale, but also boast the characteristic of possessing equal intervals between scale numbers (e.g., the distance between "1" and "2" is equal to that between "21" and "22"). Fahrenheit and Centigrade temperature scales and IQ tests are examples of interval scale measures.

Self administered sexuality questionnaires are designed with both ordinal and interval scales. They detect subtle differences in degree at a sophisticated level of measurement. A diary question (nominal measurement) might ask—"Are you satisfied with this treatment?" to which the subject responds, "Yes" or "No". In contrast to diary assessment, SAQs using ordinal/interval measurement allow investigators to capture more complexity and nuances of the subject's experience e.g., changes in desire or subjective arousal levels. As an alternative to the previous diary question on treatment satisfaction an SAQ asks, "Overall, how satisfied are you with this treatment?" to which subjects may choose one of several response options, including: "Very satisfied, Somewhat satisfied, Neither satisfied or dissatisfied, Somewhat dissatisfied, or Very dissatisfied". In addition to detecting subtle differences, Likert-type SAQ data are amenable to parametric tests and other more complex and powerful forms of statistical analysis.

The most sophisticated measurement scales are *ratio scales* (see above), which in addition to possessing equal intervals between scale points, also possess a true zero allowing equivalence of ratios throughout the scale continuum (e.g., distance, weight). Scale sophistication directly affects precision of measurement and is inversely related to error of measurement. Through this mechanism, the sensitivity of an outcomes measure to therapeutic effects can be affected, and positive outcomes in clinical trials can be muted or obscured by imprecise measures.

The FDA draft Guidance document (2000) for clinical trials of Female Sexual Dysfunction (FSD) recommends sexual event logs or "diaries" as *primary* endpoints, while self-administered questionnaires (SAQs) are recommended as *secondary* endpoints. Unfortunately, this strategy has intrinsic limitations since it requires that the least sophisticated measures available, viz., nominal scale diaries, be treated as primary outcomes measure in trials, while more psychometrically advanced SAQs, including both ordinal or interval forms of measurement, are relegated to ancillary or secondary status as outcomes measures. Ultimately, this approach serves to decrease the power of clinical trials and increase the probability of Type II measurement errors by increasing the error of measurement of the primary outcomes measure, and thereby decreasing the sensitivity of the trial to potential therapeutic efficacy (Derogatis, 2001).

Diaries can provide clinically-meaningful information when used appropriately. They offer the investigator a reliable means of documenting whether an activity took place (was intercourse attempted—yes/no), when a pill was ingested (e.g., 7:30 pm) or, when a sexual encounter occurred. The latter two questions can provide the investigator with the temporal relationship between dosing and the sexual event. Also, diary data may be of value in communicating with physicians or patients about a drug's effectiveness, (e.g., 68% of patients taking the drug increased intercourse attempts to at least 3 times per month). Diaries or event logs also allow for multiple recordings of events, thus offering researchers the benefit of aggregation of data. However, they are not a substitute for more sophisticated or accurate forms of measurement.

However, daily completion of a diary may create a skewing or biasing of the data. By completing a daily diary subjects are forced to reflect on their level of desire, arousal or satisfaction which may act as an intervention contributing, in part, to the high placebo response rates seen in FSD studies (Clayton, personal communication).

Although diaries are typically designed to be completed daily, or immediately after a sexual encounter, evidence suggests that patient burden and lack of compliance are significant limitations, undermining the argument that they are less dependent on the subject's recall or the contamination of recall through the lens of time and context (Shifren et al., 2000).

This was recently demonstrated in a cleverly designed study by Stone et al. (2003) in which patients were randomly assigned to complete either paper or electronic daily diaries. Unbeknownst to the participants in the paper diary arm was that the investigators had embedded an electronic device in the diary to record the actual dates and times they opened the diaries. These actual times were compared to the times subjects reported completing the diaries.

Subjects indicated that they complied with the investigator's instructions to complete the daily diaries 91% of the time. However, their actual

compliance was only 11% indicating that 79.5% of the data had been "faked". The author's state, "For most subjects, the records were marked by long periods-often several consecutive days or a full week—during which the diary binders were not opened at all, even though diary cards had been filled out for those periods (p. 196)." Additionally, in addition to backfilling the records, subjects also forward filled diary cards for upcoming future assessment points.

Recently, more technologically sophisticated diaries using handheld computers and telephone call-in methodologies obviate certain compliance problems, although accuracy and honesty of reporting remain concerns. In Stone et al.'s study (2003), compliance with the electronic diary was 94%. However, as the complexity and subjectivity of the inquiry increases, as is particularly the case in studies of HSDD, diaries become increasingly less useful. Some diaries are not validated measures; therefore, one can not be certain that they are measuring what the investigator hopes they are measuring. Simply because the question appears straight-forward and simple, i.e., has face validity, does not necessarily mean that it possesses any of the more compelling forms of validity—construct, discriminant, convergent, or divergent. Though diaries may deviate from counting or recording of data to ask questions that employ Likert-type response sets (e.g., very satisfied, somewhat satisfied, neither satisfied nor unsatisfied, somewhat unsatisfied, and very unsatisfied) diaries are not usually used for this purpose.

By tallying events or encounters, investigators are essentially engaged in counting, which is a primitive and relatively insensitive form of measurement. Such a model presumes a large and salient effect size to be present (e.g., erectile dysfunction and PDE5 inhibitors). FSD studies, however, tend to have high placebo response rates which obscure treatment effects unless they are quite robust. Given the complexity and diversity of women's sexual response, researchers are more likely to see moderate effect sizes. This approach runs a high risk of Type II statistical error: missing effective therapeutic interventions because our primary endpoint measures are too crude or insensitive to detect them.

In contrast, validated SAQs allow the investigator a greater sense of certainty that what they are measuring is a valid operational definition of the construct of interest (e.g., sexual desire, sexual arousal). To be psychometrically, sound SAQs must demonstrate reliability, validity (face, construct, criterion, convergent and discriminant validity), sensitivity, specificity and responsiveness. Measures that meet or surpass the thresholds for the aforementioned psychometric concepts insure the quality of the data collected.

Self-administered questionnaires are self-report data with the limitations that a person may not want to report accurately on her experiences or that she may not be a veridical observer of her own experiences (Funder, 2004). Another major limitation of self-administered questionnaires is their reliance on the subject's recall over a period of several weeks. It is

unclear how recall over time may influence these responses, although strong correlations have been observed between and within SAQ responses over time (Quirk et al., 2000; Rosen et al., 2000). The current FDA's draft guidance may reflect some current misconceptions regarding the objectivity of diaries versus SAQs. Diaries are ultimately no more "objective" than other types of self-report data. Anecdotes abound among investigators concerning patients who completed diaries in the parking lot or waiting room, or subjects who fear a negative reaction from their partners, representing events with positive distortions. Additionally, because diary data tends to be relatively crude and insensitive, it is not amenable to the more advanced analytical techniques that are at investigators' disposal to analyze ordinal and interval measures to demonstrate meaningful differences between groups.

Also, in women, sexual behaviors per se are often a poor proxy for underlying biological/motivational states (e.g., levels of sexual desire or arousal). Women engage in sexual activities for a variety of reasons (e.g., intimacy, pleasure, sense of duty, guilt, personal expectations, and marital harmony) therefore these activities may not be valid reflections of underlying sexual desire or arousal. These behaviors can be mediated by multiple factors that are unrelated to the core biological events that the interventions are designed to effect, thereby introducing unknown levels of distortion into the outcomes assessments.

An alternative strategy for clinical trial outcomes assessment in FSD would be to employ either event logs/diaries and validated psychometric outcomes measures (i.e., self-report inventories and rating scales) as primary endpoints. Typically one, not both, would be designated as the primary endpoints and gather data appropriate to each instrument's psychometric strengths.

Finally, it is worth noting that for over four decades in psychiatric clinical trials involving psychotropic medication, SAQs have long been utilized as primary endpoints, not to the exclusion of behavioral data, but concomitant with it (Meltzer, 1987). It appears that such an approach would result in superior clinical trials that would not only possess the capacity to identify ineffective treatments early on, but would also possess the sensitivity to maximally identify promising new agents and interventions.

CONCLUSION

With the resurgence of interest in FSD research guidelines that apply to industry and non-industry sponsored research need to be refined. As the FDA develops their guidance from the current draft version to a final version, it is expected that they will take into account the evolving science/experience accruing in the realm of female sexual dysfunction. We hope this paper may serve as one meaningful advance in thinking and the ideas contained herein

will be given thoughtful consideration for possible incorporation into their next guidance document.

We believe there is insufficient data to support the FDA draft guidance that in FSD clinical trials diary measures should serve as primary endpoints while self-administered questionnaires be designated "a priori" as secondary endpoints. As we have described, diary-based measures have important disadvantages when used in this area. SAQs can potentially provide more sensitive and accurate information in the measurement of complex and subjective aspects of women's sexual function.

We are also concerned with the FDA's emphasis on the number of satisfactory or sexual events over time as the primary endpoint for all FSD clinical trials. The FDA advocates counting events; however, this crude measure obtained by diary or event log methods, may not capture the subtle, complex and subjective aspects that comprise the essential nature of the women's sexual difficulties. SAQs are more suited for reliably measuring these changes.

The lack of concordance or agreement between different measures of outcome poses a serious problem for investigators in interpreting and understanding the results of their studies. However, because the diary is the more crude or insensitive measure, and is best used for counting events having taken place, the SAQ data are more likely to reflect the true outcome where a multidimensional assessment can be made.

At present, there is insufficient objective evidence to make definitive claims for one form of measurement exclusively over another in this emerging new area of research. However, as stated above, the diary-based measures have essential limitations that should be given careful consideration in the design of future trials of FSD. In particular, we recommend that the FDA reconsider its position on the use of diary measures as primary outcomes for all clinical trials of FSD. We offer an alternative set of recommendations in the hope that the FDA will reconsider its position. As we have discussed, self-administered questionnaires, being more sophisticated measurement tools, are well suited to assess complex subjective aspects of sexual response that can capture a multidimensional overview of outcome.

Based on this review we offer the following recommendations for revision of the FDA's current draft guidance. First, we recommend that SAQs be elevated from their current status as secondary endpoints and be considered as potential primary endpoints in clinical trials of FSD. Second, we recommend that depending on the trial design and intervention under study, that either an SAQ or diary (typically one or the other, and not both), might serve as primary endpoints in the trial. Third, SAQs and diaries should be employed, analyzed and interpreted in their particular areas of strength. Diaries are most useful for enumerating events and/or counting frequencies. SAQs are superior at gathering subjective data related to women's sexual function. Fourth, we believe there is a theoretical basis for considering SAQs to be superior measurement tools compared to diaries. At present, there is

however insufficient objective data to justify this opinion. Conversely, we do not anticipate either theoretical or objective evidence to support the alternative hypothesis (that diaries may be superior to SAQs). If our belief proves to be correct, in the future diary measures will no longer be considered as primary endpoints for FSD clinical trials. Finally, we recommend that the FDA reconsider the emphasis given to the number of successful or satisfactory sexual events over time because it represents an amalgam of subjective assessments that is too distal to the FSD component under study.

REFERENCES

American Psychiatric Association. (2002). *Diagnostic and Statistical Manual of Mental Disorders*—IV-TR (4th edition-Text revision). American Psychiatric Association, Washington, D.C.

Bancroft, J., Loftus, J., & Long, J. S. (2003). Distress about sex: A national survey of women in heterosexual relationships. *Archives of Sexual Behavior, 32*, 193–208.

Basson, R., McInnes, R., Smith, M. D., Hodgson, G., & Koppiker, N. (2002a). Efficacy and safety of sildenafil citrate in women with sexual dysfunction associated with female sexual arousal disorder. *Journal of Women's Health and Gender Based Medicine, 11*(4), 367–377.

Basson, R. (2002b). A model of women's sexual arousal. *Journal of Sex and Marital Therapy, 28*, 1–10.

Basson, R. (2002c). Are our definitions of women's desire, arousal and sexual pain disorders too broad and our definitions of orgasmic disorders too narrow. *Journal of Sex and Marital Therapy, 28*, 289–300.

Basson, R., Leiblum, S., Brotto, L., Derogatis, L., Fourcroy, J., Fugl-Meyer, K., Graziottin, A., Heiman, J., Lann, E., Meston, C., Schover, L., van Lankveld, J., & Schultz, W. W. (2004). Definitions of women's sexual dysfunction reconsidered: Advocating expansion and revision. *Journal of Psychosomatic Obstetrics and Gynecology, 24*, 221–229.

Berman, J. R., Berman, L. A., Toler, S. M., Gill, J., & Haughie, S. (2003). Safety and efficacy of sildenafil citrate for the treatment of female sexual arousal disorder: A double-blind, placebo-controlled study. *Journal of Urology, 170*, 2333–2338.

Caruso, S., Intelisano, G., Farina, M., Di Mari, L., & Agnello, C. (2003). The function of sildenafil on female sexual pathways: A double-blind, cross-over, placebo-controlled study. *European Journal of Obstetrics and Gynecology and Reproductive Biology, 110*, 201–206.

Clayton, A. (2004). Personal communication.

Derogatis, L. R. (2001). Sexual Function and Quality of Life: Endpoints and Outcomes. *Journal of Gender Specific Medicine, 4*, 35–42.

Derogatis, L. R., & Coons, H. L. (1993). Self report measures of stress. In L. Goldberger & S. Breznitz (Eds.), *Handbook of Stress: Theoretical and Clinical Aspects*. New York: The Free Press-Macmillan.

FDA Guidance Document, Guidance for Industry: Female Sexual Dysfunction: Clinical Development of Drug Products for Treatment (2000). www.fda.gov/cder/guidance/3312dft.htm

Funder, D. C. (2004). *The Personality Puzzle* (3rd ed.). New York: W. W. Norton.

Harris, G. (2004). Pfizer gives up testing Viagra on women. New York Times, February 28th, Section C, page 1, column 5.

Laan, E., Everaerd, W., van der Velde, J., & Geer, J. (1995). Determinants of subjective experience of sexual arousal in women: feedback from genital arousal and erotic stimulus content. *Psychophysiology, 32*, 444–451.

Lobo, R., Rosen, R., Yang, H., Block, B., & Van Der Hoop, R. (2003). Comparative effects or oral esterified estrogens with and without methyltestosterone on endocrine profiles and dimensions of sexual function in postmenopausal women with hypoactive sexual desire. *Fertility and Sterility, 79*, 1341–1352.

Meltzer, H. Y. (1987). *Psychopharmacology: The Third Generation of Progress*. Raven Press, New York (LRD).

Quirk, F. H., Heiman, J. R., Rosen, R. C., Laan, E., Smith, M. D., & Boolell, M. (2002). Development of a sexual function questionnaire for clinical trials of female sexual dysfunction. *Journal of Womens Health Gender Based Medicine, 11*, 331–333.

Rosen, R. C., & Beck, J. G. (1988). *Patterns of Sexual Arousal: Psychophysiological Processes and Clinical Applications*. New York: Guilford Press.

Rosen, R. C., Brown, C., Heiman, J., Leiblum, S., Meston, C., Shabsigh, R., Ferguson, D., & D'Agostino. (2000). The Female Sexual Function Index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. *Journal of Sex and Marital Therapy, 26*, 191–208

Shifren, J., Glenn, B. D., Simon, J. A., Casson, P. R., Buster, J. E., Redumond, G. P., Burki, R. E., Ginsburg, E. S., Rosen, R. C., Leiblum, S. R., Caramelli, K. E., & Mazur, N. A. (2000). Transdermal testosterone treatment in women with impaired sexual function after oophorectomy. *The New England Journal of Medicine, 343*, 682–688

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials,* (24), 182–199.

Taylor, J. F., Rosen, R. C., & Leiblum, S. R. (1994). Self report assessment of female sexual function: Psychometric evaluation of the Brief Index of Sexual Functioning for Women. *Archives of Sexual Behavior, 23*, 627–643.

Tiefer, L., Hall, M., & Tavris, C. (2002). Beyond dysfunction: A new view of women's sexual problems. *Journal of Sex and Marital Therapy, 23*, 225–232.

World Health Organization (WHO). (1989). International Classification of Diseases, 10th Revision.