

Scoring and Interpretation of the FSFI: What can be Learned From 20 Years of use?

Cindy M. Meston, PhD,¹ Bridget K. Freihart, MSW,¹ Ariel B. Handy, BA,¹ Chelsea D. Kilimnik, MA,¹ and Raymond C. Rosen, PhD²

ABSTRACT

Introduction: Over the past 20 years, the Female Sexual Function Index (FSFI) has been considered the gold standard for the measurement of sexual function in women, with over 1,000 published manuscripts citing the article. Despite the measure's widespread usage and excellent psychometric properties, there has been some confusion over how to best implement and score the measure and interpret corresponding findings.

Aim: The aim of the current article is to provide guidance, drawing from 20 years of use, on how to best implement the FSFI in research settings and interpret results based on the validation studies that have been conducted to date.

Methods: The overview of scoring and interpretation procedures found in this article is drawn from a review of the published literature on the psychometric properties of the FSFI.

Main Outcome Measure: The measure of interest for the present review is the FSFI.

Results: This review article provides information about implementing, scoring, and interpreting the full-scale FSFI. Domain-level scoring and interpretation procedures are also discussed across the 5 domains of the FSFI: arousal, satisfaction, desire, pain, and lubrication. Additionally, guidance is provided for evaluating translated versions of the FSFI and using the measure to examine sexual function in culturally diverse populations.

Clinical Implications: Guidance on appropriately scoring and interpreting the FSFI has the potential to strengthen our empirical understanding of sexual function, and consequently, to guide theory-driven treatment development and clinical practice.

Strength & Limitations: The present review provides applied guidance for the appropriate use of the FSFI specifically, but does not cover other common measures of sexual function or adaptations of the original measure.

Conclusion: It is our hope that the guidance found in this review will ultimately lead to more rigorous and appropriate usage of the FSFI in research settings. **Meston CM, Freihart BK, Handy AB, et al. Scoring and Interpretation of the FSFI: What can be Learned From 20 Years of use? J Sex Med 2019;XX:XXX–XXX.**

Copyright © 2019, International Society for Sexual Medicine. Published by Elsevier Inc. All rights reserved.

Key Words: Female Sexual Function Index (FSFI); Sexual Function; Psychometric Review

INTRODUCTION

The study of women's sexual function has increased substantially over the past 2 decades. As a result, a number of new assessment instruments have been developed for use in diagnostic, epidemiological, and treatment outcome studies. Among

these, the Female Sexual Function Index (FSFI)¹ has become the most widely used screening tool and outcome measure of female sexual function, likely because of its clear wording and scale structure, excellent psychometric properties, and relative brevity. An English-only Google Scholar search using the terms "FSFI" and "Female Sexual Function Index" led to almost 10,000 hits as of May 1, 2019. The FSFI has been used in psychometric and translation studies, studies examining female sexual function and dysfunction, epidemiological studies on the prevalence of female sexual dysfunction (FSD) in specific geographic locations or populations of women with particular diseases or conditions, and treatment studies and clinical trials that assessed the efficacy of therapies and interventions for FSD. These studies total over 1,000 published reports, and include at least 20 translations into languages other than English.

Received June 29, 2019. Accepted October 5, 2019.

¹Department of Psychology, University of Texas at Austin, Austin, Texas, USA;

²Adjunct Professor of Psychiatry, Rutgers – Robert Wood Johnson Medical School, Piscataway, New Jersey, USA

Copyright © 2019, International Society for Sexual Medicine. Published by Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.jsxm.2019.10.007>

The FSFI¹ is a 19-item, self-report measure of female sexual function that provides scores on overall levels of sexual function as well as the primary components of sexual function in women, including sexual desire, arousal, orgasm, pain, and satisfaction. The instrument was developed in the year 2000 by a panel of 8 experts in the field of female sexuality who created the initial items based on an extensive literature review and their understanding of FSD at that time, which was influenced by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)² and International Classification of Diseases, Tenth Edition (ICD-10).³ An emphasis was placed on items related to Female Sexual Arousal Disorder (FSAD) given that, with the introduction of sildenafil by Pfizer just 2 years prior in 1998, there was an explosion of interest in developing drugs for the treatment of arousal disorders in women. Although the panel acknowledged the substantial overlap between disorders of desire and arousal, both in terms of factor analytic findings and clinical observations, they elected to maintain separate domains for disorders of desire and arousal, and to further differentiate arousal disorders into those that were genitally based (eg, lubrication) and those that were more subjective or cognitively based. Their reasoning was to provide a greater ability to assess treatment specificity. Interestingly, almost 2 decades later in 2019, a panel of 10 international experts on FSD funded by the International Society for the Study of Women's Sexual Health (ISSWSH) recommended maintaining separate diagnostic categories for disorders of desire and arousal in women, and proposed subdividing disorders of arousal into those that are genital (Female Genital Arousal Disorder) and those that are subjective or cognitive (Female Cognitive Arousal Disorder).⁴ The proposal to maintain separate diagnostic categories for desire and arousal disorders is consistent with the proposed ICD-11 diagnostic guidelines⁵ but in disagreement with the DSM-V⁶ merger of desire and arousal into a single diagnostic category called Female Sexual Interest/Arousal Disorder.

In the initial psychometric validation phase, the FSFI was administered to 131 control women and 128 age-matched women with DSM-IV diagnoses of FSAD established through a clinical interview. The inventory was then tested for construct validity (factorial, discriminant, and divergent) and reliability (internal consistency and test-retest reliability).¹ The results of these assessments indicated excellent internal reliability (Cronbach's alphas > 0.9 for all subscales) and good test-retest reliability (assessed across 2 to 4 weeks; $r = 0.79-0.88$). Construct validity was demonstrated by highly significant mean difference scores between the sexually functional and dysfunctional groups for each of the domains ($P \pm .001$). Modest correlations between the FSFI and the Locke-Wallace Marital Adjustment Test,⁷ a measure of marital satisfaction, provided further support for the construct validity of the FSFI. In 2003, Meston⁸ extended the FSFI validation in 71 women with a DSM-IV-TR⁹ diagnosis of Female Orgasmic Disorder (FOD), 44 women with Hypoactive Sexual Desire Disorder (HSDD), and 71 age-matched control

women by confirming the FSFI's discriminant validity, divergent validity, and reliability (internal consistency and test-retest reliability).

In addition to these 2 original FSFI validation studies, Wiegel et al¹⁰ added further construct and discriminant validity by combining the data sets used by Rosen et al¹ and Meston⁸ with data from an additional 123 subjects that included women with multiple sexual dysfunctions and sexual pain disorders. By using a sufficiently large and heterogeneous sample, Wiegel et al¹⁰ were able to confirm the FSFI's discriminant validity for a wide range of sexual dysfunctions and to perform a Classification and Regression Tree (CART) analysis, which demonstrated that an FSFI total score of 26.55 provides an optimal cutoff for differentiating women with and without FSD.¹⁰ In 2010, CART analyses were conducted on 8 datasets that yielded a total of 618 women, and a cutoff score of 5 was established on the Desire domain in order to differentiate women with and without HSDD.¹¹ Numerous additional psychometric studies have since been conducted on the FSFI (for an extensive review, see Neijenhuijs et al¹²).

Despite the FSFI's excellent psychometric properties and ease of use, a number of errors are frequently made in the scoring and interpretation of FSFI results. For example, there exists a plethora of published studies that have administered the FSFI to women who are not sexually active, that have failed to calculate domain scores without multiplying the domain sum by the domain factor, or have used a reference point other than the past 4 weeks. Failing to follow the published FSFI scoring procedure and scale requirements not only renders the psychometrics invalid, but leads to artificially low "or high scores and misdiagnoses," which, in turn, muddies the literature and our knowledge of women's sexual dysfunction. The aim of this article is to provide an in-depth explanation of the correct scoring procedure to be adopted when using the FSFI, to provide guidelines on how the FSFI domain and total scores can and should be interpreted, to point out common errors in the scoring and administration of this instrument, and to provide guidance for how the measure can be used across diverse populations of women.

SCORING AND INTERPRETATION

The 19 items of the FSFI use a 5-point Likert scale ranging from 1–5 with higher scores indicating greater levels of sexual functioning on the respective item.¹ To score the measure, the sum of each domain score is first multiplied by a domain factor ratio (0.6 for desire; 0.3 for arousal; 0.3 for lubrication; 0.4 for orgasm; 0.4 for satisfaction; and 0.4 for pain) in order to place all domain totals on a more comparable scale, and then subsequently summed to derive a total FSFI score.¹

15 of the items contain a zero option in the response set to indicate either "no sexual activity" (12 items) or "did not attempt intercourse" (3 items) within the past 4 weeks. A requirement for

using the FSFI to calculate a total sexual functioning score is that the participants have engaged in sexual activity and have attempted vaginal penetration over the past 4 weeks. As noted by Meyer-Bahlburg and Dolezal¹³ in their critique of the methodological employment of the FSFI in research, drug trials, and clinical practice, the zero category in the response set is conceptually distinct from the assessment of sexual functioning. The absence of sexual activity or intercourse is not necessarily attributable to sexual dysfunction. Thus, the incorrect use of the zero category in calculating FSFI domain and total scores would drastically underestimate women's sexual functioning scores, increase the variance of total and domain scores, potentially inflate FSFI score differences between groups with and without FSD, and undermine the assessment of sexual dysfunction when using established clinical cutoffs. In such cases, one indication that the FSFI has been scored incorrectly is a total score falling outside of the possible range of 2–36.

In an effort to avoid use of the zero categories, many researchers have implemented exclusion criteria for women who have not been sexually active within the inventory's 4-week reference period. This is an acceptable option for ensuring against erroneous zero scores being added into the calculation of FSFI domain and total scores. For studies that do not exclude sexually inactive women, or for studies that use the FSFI for clinical screening purposes, the calculation of total FSFI scores and relevant domain scores (on all but the desire domain) should be limited to those who have not indicated a zero score on any of the FSFI items. This will ensure against incorrectly biasing the inventory's results toward dysfunctional scores. Alternatively, a lifelong version of the FSFI has been validated and may be an appropriate alternative to the original FSFI in cases where using the 4-week reference period presents logistical or empirical challenges.¹⁴

The original FSFI publication and several subsequent studies have demonstrated statistically significant mean differences between women with and without FSD, FSAD, and FOD using between-group mean comparisons on total FSFI scores and respective domain scores.^{1,8} Other research has used established clinical cutoffs for the FSFI to classify groups of women with and without FSD or to flag those likely to meet diagnostic criteria for clinical levels of sexual dysfunction in screening procedures.¹⁵ As such, clinical trials or assessments of treatments for FSD have 2 valid approaches to assess the effect of a treatment on women's sexual functioning: (1) changes in FSFI scores over the course of treatment and (2) changes in the number of women meeting the clinical cutoff criteria for FSD.

The FSFI clinical cutoff score, established by Wiegel et al¹⁰ in 2005, has been widely used as an index to aid in distinguishing between women with and without clinical sexual dysfunction. The authors used both response operator characteristic (ROC) curves and CART procedures to extract the optimal score to balance both specificity and sensitivity in the detection of FSD. They fit ROC curves to the data of a large sample of women with and without sexual dysfunctions to determine the variable (ie,

from the various domains and the total score) with the largest area under the ROC curve (ie, the best sensitivity to 1 – specificity profile; area = 0.899, $P < .001$). They then used CART procedures (a process of developing a classification algorithm to determine the optimal variable(s) and cutoff point to yield the most accurate classification of subjects into categories see Breiman et al¹⁶ 1984 for further information) to determine which was the best decision-basing variable in the model. The findings of both the ROC curve and CART procedures indicated that the FSFI total score variable was the most predictive classification variable. Specifically, they concluded that scores of ≤ 26.55 most accurately captured clinically relevant sexual dysfunction (specificity = 0.733; sensitivity = 0.889).¹⁰ Although the results indicated that the FSFI total score yielded the most accurate results, the authors also discuss the added utility of the individual domain scores (both within the model and conceptually) for estimating differential diagnoses across the various types of FSD.

Despite the FSFI's utility as a quantifiable measure of clinical sexual dysfunction, it is not a clinically diagnostic tool. The FSFI alone cannot be used to make a diagnosis of a sexual dysfunction as there is no measure of distress within the scale and the DSM-V⁶ and both the ICD-10³ and the proposed ICD-11⁵ diagnostic criteria require sexual complaints to be associated with significant amounts of distress. One option to further the diagnostic applicability of the FSFI is to administer an accompanying assessment of sexual function-related distress, such as the Female Sexual Distress Scale.¹⁷ The combined assessment of FSFI total scores, FSFI domain scores, and levels of sexual distress would allow for a more substantiated argument for a proxy clinical diagnosis of sexual dysfunction. Of note, Stephenson et al¹⁸ demonstrated the clinical utility of the FSFI in their report of moderate to large correlations between FSFI total and domain scores and clinical interview data. However, the absence of a distress assessment within the FSFI precludes a truly diagnostic evaluation of FSD. It should also be noted that a clinical diagnosis of FSD cannot be made without ruling out other medical diagnoses and medications that can impact sexual functioning, and the period of time assessed by the FSFI is not fully representative of the period of time required to determine a clinical diagnosis. The FSFI uses a time reference of the past 4 weeks, whereas the DSM-V uses a reference period of the past 6 months and the proposed ICD-11 uses a reference period of “at least several months.”^{5,6}

Although no self-report questionnaire can claim to fully replace the “gold standard” clinical interview, the FSFI does provide insight into the degree and type of sexual dysfunction experienced, and can be used to guide further assessment within the spectrum of FSD diagnoses. As discussed by Wiegel et al,¹⁰ the domain scores are useful for interpreting differential diagnoses. Yet, many researchers publish FSFI total scores without also publishing the domain scores. In the absence of domain scores, it is unclear which specific aspects of sexual functioning (eg, desire, arousal, orgasm, and pain) account for the lower FSFI total scores. Notably, the FSFI clinical cutoff score cannot be

used to determine the presence or absence of a specific type of sexual dysfunction. Wiegel et al¹⁰ validated the clinical cutoff in a sample of women with mixed sexual dysfunctions (ie, FSAD, FOD, HSDD, and dyspareunia/vaginismus) and, as such, the total cutoff score only relates to general (ie, nonspecific) FSD.

It is important to note that having a clinically validated cutoff score is not the same as being validated for use within a certain population. To validate a questionnaire for use within a certain population, it must be able to successfully discriminate between populations while also demonstrating good psychometric properties. The FSFI has been validated for use in healthy controls, women with FSAD, FOD, chronic pelvic pain (CPP), and both premenopausal and postmenopausal women with HSDD.^{1,8,19,20} This does not, however, indicate that clinical cutoff scores exist for each of these populations on the full scale or for any particular domain.

DOMAIN SCORING AND INTERPRETATION

Desire

The FSFI Desire domain reflects a motivation to engage in sexual activity.^{21,22} Items in the Desire domain examine a woman's frequency and degree of experiencing sexual desire. Domain scores range from 1.2–6. Of the 6 sexual function domains included in the FSFI, the 2-item Desire domain is the only domain that can be used independently. Through ROC analyses on 2 independent populations of women with and without HSDD, Gerstenberger et al¹¹ found that a cutoff score of 5 on the Desire domain maximized diagnostic sensitivity (ie, correctly classifying women *with* HSDD) and specificity (ie, correctly classifying women *without* HSDD). This indicates that women with scores of 5 or less on the Desire domain likely meet diagnostic criteria for HSDD, and women with scores greater than 5 likely do not meet criteria. However, as noted earlier, the FSFI does not examine distress—a key diagnostic component included in both the DSM and ICD diagnostic systems. As such, it is recommended that a validated scale assessing distress be administered along with the FSFI Desire domain when making clinical inferences.

The Desire cutoff score can be used for screening women with/without desire concerns for clinical trials, as outcome criteria for assessing changes in desire in clinical trials, and for determining population estimates of HSDD. In outcome studies, clinical improvements can be inferred by examining whether domain scores move from below to above the cutoff score over the course of treatment. This could be achieved either by looking at the change in average domain score for all women undergoing a specific treatment or by examining individual scores and determining how many women no longer fall below the clinical cutoff. A shift from below to above the clinical cutoff score indicates the woman would likely no longer meet clinical criteria for HSDD. Similarly, if using the FSFI Desire domain for participant selection, when combined with a validated measure of distress, the Desire cutoff score of 5 could be used as a proxy indicator of clinical status.

Arousal

Distinction Between Arousal and Desire

There has been substantial confusion and debate in the literature as to what exactly arousal is, and whether it should be considered a manifestation of desire. Much of this confusion stems from diagnostic changes incurred with the transition from the fourth to fifth edition of the DSM. Within both the DSM-4-TR⁹ and the ICD-10,³ disorders of arousal and desire had been recognized as distinct disorders warranting unique diagnoses. These 2 disorders were merged into 1 diagnosis, Female Sexual Interest/Arousal Disorder, in the DSM-V. This was due, in part, to what was viewed as a meaningful overlap between desire and arousal. Indeed, this overlap is reflected in the factor structure of the FSFI itself. In the original validation article of the FSFI, the factor analytic structure supported either a 5-factor (desire and arousal combined) or 6-factor (desire and arousal distinct) solution. A 6-factor solution was ultimately chosen in order to better assess treatment specificity. Support for both a 5-factor and 6-factor solution was recently replicated in a systematic review of 83 studies investigating the measurement properties of the FSFI.¹²

Domain intercorrelations published in the initial validation study of the FSFI in a combined group of women with and without FSAD indicate a shared variance of only 58% between the domains of Desire and Arousal. This demonstrates overlap between the constructs of desire and arousal but also substantial distinction. Further evidence for maintaining separate domains for Desire and Arousal is provided by Althof et al.²¹ In addition, many organizations (ie, ISSWSH, the committee for ICD-11) have opted to keep these 2 diagnostic categories distinct.^{4,5}

Distinction Between Cognitive Arousal and Genital Arousal

The FSFI was validated for use in women with a DSM-IV-TR diagnosis of FSAD,¹ indicating that it can detect mean differences between women with and without a DSM-IV-TR diagnosis of FSAD. It is important to note that FSAD as delineated in the DSM-IV-TR refers to a lack of genital arousal, namely a lubrication-swelling response. As such, the FSFI Lubrication domain closely taps into the DSM-IV-TR definition of FSAD. However, more recently, FSAD has been conceptualized more broadly as a disorder that includes both a subjective or cognitive component and a genital or physiological component. Recently, a panel of 10 experts, funded by the ISSWSH convened with the goal of developing specific nosology and nomenclature for arousal disorder in women. The outcome of this consensus meeting was the proposal of an overarching category of FSAD that includes 2 subtypes of arousal disorders: Female Cognitive Arousal Disorder (FCAD) and Female Genital Arousal Disorder (FGAD).⁴ Using this more comprehensive and specific conceptualization of FSAD, both the FSFI Arousal and Lubrication domains are required for assessment. FCAD is most closely assessed using the FSFI Arousal domain, and FGAD is most closely assessed using the FSFI Lubrication domain.

FSFI Arousal Domain

The FSFI Arousal domain contains 4 items assessing the frequency of arousal, intensity of arousal, confidence in one's ability to become aroused, and frequency of feeling satisfied with one's arousal response. Domain scores range from 0–6. As noted earlier, the Arousal domain reflects *cognitive* arousal, or a woman's experience of being mentally "turned on." Although the FSFI Arousal domain closely maps onto the FCAD subtype of arousal disorders recently proposed,⁴ it does not assess for the presence or absence of distress, which is a requirement for the clinical diagnosis of FCAD. It is also important to note that the FSFI was validated among women using the DSM-IV-TR criteria for FSAD and not the newly proposed criteria for FCAD. Studies are required to determine whether mean scores on the FSFI Arousal domain differ significantly between women who have been clinically diagnosed as having FCAD and women who have not.

Unlike the FSFI Desire domain, clinical cutoff scores for the Arousal domain have not yet been determined. As such, the Arousal domain may be used by comparing mean scores from pretreatment to post-treatment to provide an indication of how markers of arousal have responded to treatment, but strong inferences about clinically relevant changes cannot be made.

Lubrication

4 items comprise the FSFI Lubrication domain. The items assess the frequency and difficulty of attaining and maintaining lubrication during sexual activity; domain scores range from 0–6. The Lubrication domain most closely aligns with the DSM-IV-TR⁹ diagnosis of FSAD and the newly proposed FSAD subtype of FGAD.⁴ It does not, however, assess for the presence or absence of distress, which is a requirement for the clinical diagnosis of both FSAD and FGAD. It is also important to note that lubrication is only 1 aspect of genital arousal. Parish et al⁴ propose the following definition of FGAD: "FGAD is characterized by the distressing difficulty or inability to attain or maintain adequate genital response, including vulvovaginal lubrication, engorgement of the genitalia, and sensitivity of the genitalia associated with sexual activity, for a minimum of six months."

As the FSFI was validated among women using the DSM-IV-TR criteria for FSAD and not the criteria for FGAD, future studies are required to determine whether mean scores on the FSFI Lubrication domain differ significantly between women who have been diagnosed as having FGAD and those who have not. As was the case with the FSFI Arousal domain, a clinical cutoff score has not yet been determined for the Lubrication domain. As such, the Lubrication domain can be used for examining treatment changes in genital lubrication but cannot be used to diagnose either FSAD or FGAD.

Orgasm

The Orgasm domain of the FSFI contains 2 items pertaining to a woman's ability to reach orgasm characterized by

self-reported frequency of orgasm and difficulty reaching orgasm, and one item that pertains to a woman's satisfaction with the experience of orgasm. Domain scores range from 0–6. A clinical cutoff score has not yet been determined for differentiating women with and without FOD, and, thus, the domain cannot be used to infer clinical diagnoses. The FSFI has, however, been validated for use in this population, which means that the measure produced significant mean differences between women with and without a clinical diagnoses of FOD.⁸ This validation suggests that full scale FSFI scores have the sensitivity to detect clinically significant orgasm dysfunction. Although the orgasm domain cannot be used for diagnostic purposes, it can be used for patient selection, tracking treatment progress, and clinical inference in the same manner as described for the Arousal and Lubrication domains.

Satisfaction

The Satisfaction domain includes 3 items assessing satisfaction: 2 that are partner relevant, and 1 that assesses satisfaction with "overall sex life." For the 2 items that assess partner variables, 1 pertains to the general sexual relationship with one's partner and the other specifically examines the relationship with one's partner *during sex*. This added specificity differs from most questionnaires that assess sexual or relationship satisfaction in that it allows for a greater understanding of how satisfied women are with their actual sexual experiences. Scores on this domain range from 0.8–6.

It is important to note that the Satisfaction domain is different from a diagnostic notion of distress. Satisfaction generally refers to a subjective experience of well-being,^{23,24} whereas distress refers to negative emotionality, such as worry, frustration, or anxiety.^{17,25} Although research has shown that distress and satisfaction are positively correlated,²⁶ using the Satisfaction domain to make clinical inferences about distress is not advised. In a sample of women with and without FSAD, Stephenson and Meston²⁶ found that, although changes in satisfaction and distress tend to be associated, the relationship between these 2 constructs differs by clinical status. For women with FSAD, distress seems closely linked with sexual function. In sexually healthy women, however, satisfaction may be more closely linked with sexual function than is distress. Satisfaction and distress, although closely linked, are not interchangeable and do not represent opposite poles of the same continuum. The Satisfaction domain within the FSFI should, therefore, not be used as an indicator of distress, but rather as an indicator of women's subjective experience of sexual well-being.

Pain

The Pain domain of the FSFI captures genital pain that is elicited during or after vaginal penetration. This domain contains 3 items, and domain scores range from 0–6. There has not yet been a clinical cutoff score developed for the Pain domain and, as such, it cannot be used for clinical diagnoses. The FSFI has been

validated for use in women with CPP, indicating the FSFI is able to significantly discriminate between women with and without clinical diagnoses of CPP.¹⁹ Masheb et al²⁷ demonstrated strong reliability and discriminant validity of the FSFI total and domain scores in a sample of 42 women diagnosed via gynecologic evaluations with chronic vulvar pain (vulvodynia). More recently, Smith et al^{28,29} have shown the FSFI to differ significantly between women with and without gynecologist-diagnosed provoked vestibulodynia, although discriminant validity has yet to be established. As with the other FSFI domains that do not have established clinical cutoff scores, the Pain domain cannot be used to make strong clinical inferences or diagnoses. It can, however, be used for patient selection and for tracking treatment progress by examining changes in the mean domain score.

FSFI USE ACROSS POPULATIONS

Translations Across Languages

A large component of the FSFI's widespread usage can be attributed to translated versions, which have been developed for a range of languages, including Arabic,³⁰ Chinese,³¹ Filipino,³² French,³³ German,³⁴ Greek,³⁵ Indonesian,³⁶ Italian,³⁷ Iranian,³⁸ Japanese,³⁹ Korean,⁴⁰ Malay,⁴¹ Persian,⁴² Polish,⁴³ Portuguese,⁴⁴ Spanish,⁴⁵ Swedish,⁴⁶ Taiwanese,⁴⁷ Turkish,⁴⁸ and Urdu.⁴⁹ Although translations of the FSFI have facilitated advances in our understanding of sexual function cross-culturally, there are variations in the degree to which such versions adhere to appropriate procedures for translation and validation. As such, it is important to examine the development procedures and measurement properties of individual FSFI translations when implementing such measures and interpreting corresponding findings.

For a version of the FSFI to be appropriately translated and validated, it should follow the following standard procedures: (1) the original FSFI should undergo a forward translation (ie, translation of the FSFI into the language of interest), (2) the translated version should then be subject to a back translation (ie, the translated version should be translated back into English by someone uninvolved in the forward translation), (3) researchers should then assess the degree to which the back translation matches the original, English version of the FSFI, with steps 1–3 completed in an iterative process until the back translation closely matches the original, (4) the translated version of the FSFI should then be administered to a sample of women who have also completed a clinical interview for FSD, and (5) researchers should determine whether there are statistically significant differences between the scores of women with and without FSD (as determined by a clinical interview).

Several translations of the FSFI can be considered quite valid and reliable given their close adherence to the above procedures. For instance, the Iranian version of the FSFI was developed using all of these procedures, and included additional steps to increase evidence for the measure's reliability, including a 4-week

follow-up to assess test-retest reliability.³⁸ Other versions have been subject to even more rigorous analyses to increase their specificity for use in different cultural contexts. For instance, several translated versions have not only been appropriately validated, but have also been used to develop clinical cutoff scores that are population-specific (eg, Zachariou et al³⁵ and Filocamo et al³⁷). In such cases where the FSFI has been properly translated and validated, the resulting measure can be reasonably trusted to accurately assess sexual function within a given population.

This is not, however, the case for translated versions that have not been developed or validated using appropriate procedures. For instance, a validation study for the Spanish version of the FSFI followed many of the procedures outlined above but did not include structured clinical interviews.⁴⁵ Although researchers were able to conclude that the overall factor structure of the original measure held in the translated version, they were unable to determine whether the new version produced meaningful differences between populations of women with and without FSD. Because this was a preliminary validation study, results from studies using this measure or other translated measures that have not been fully validated should be interpreted with caution until such a time that the measures can undergo more rigorous evaluation.

Validation Across Populations

Cross-Cultural Validation

There is an important distinction between the *use* of a scale cross-culturally and the *validation* of a scale on a specific population. Although the original FSFI is appropriate for usage across a wide-range of cultural contexts, there are cases in which validation on a specific population of interest is advised. Here, validation refers specifically to (1) demonstrating that the factor structure of the measure holds for the specific population and (2) finding significant differences between the FSFI scores of women with and without FSD in that population.

In most instances, validation of the FSFI for a specific population is not necessary for effective use. For instance, the FSFI does not need to be validated for use on women with depression, as the scale is specifically designed to examine differences in sexual function across diverse groups. It becomes more important to ensure that the scale has been validated on a population of interest, however, when their cultural context varies significantly from the population on whom the FSFI was originally validated. For instance, one study validated the FSFI on an English-speaking population in Holland, a cultural context that may be notably different than the United States (see Wouters,⁵⁰ 2013, for an eloquent review of increased sexually liberal dynamics in the Netherlands vs the United States), where the FSFI was originally validated.⁵¹ Similarly, when administering translated versions of the FSFI in languages spoken across culturally diverse countries, it may be important to ensure they have been validated

for the specific population of interest (ie, ensuring the Spanish version of the FSFI, which was developed in Colombia, is valid in a Mexican population).

Use with Sexual Minority Women

The FSFI was specifically written to be applicable to both heterosexual and sexual minority women. Across all of the domains except for pain, items use wording such as “sexual activity or intercourse” or “sexual stimulation or intercourse” in order to be inclusive of a wide range of sexual contexts, including non-penetrative sex. In the Pain domain, the word “penetration,” is used rather than “sexual activity,” given that sexual pain almost always occurs in response to vaginal insertion of some kind (ie, penile, digital, etc.). Again, the word “penetration” was chosen over the term “intercourse” to be inclusive of sexual pain occurring outside of the context of penile-vaginal sex. Although the FSFI was written such that it could be *used* in research settings with lesbian and bisexual women, the measure has not been specifically *validated* in this population. In order to better assess sexual function among sexual minority women, one study slightly modified the items of the FSFI by omitting the term “intercourse” and instead using the term “penetration” and initial evidence for sound psychometric properties.^{50,52} Using the altered wording and original scoring approach, Cronbach alpha’s for domain and total scores were excellent and ranged from 0.89–0.95. Further work should be done to validate that modified version with clinical interviews in order to determine whether the measure can reliably produce mean group differences between women with and without FSD. Additionally, future research should seek to elucidate whether the original FSFI itself is a valid measure for assessing sexual function in sexual minority women, and, if so, to develop clinical cutoff scores specific to this population.

SUMMARY AND CONCLUSIONS

The FSFI is a 19-item, multidimensional, self-report measure of sexual function in women, which was developed 2 decades ago in response to burgeoning interest in new pharmacologic and behavioral treatments for FSD. Since its initial publication in 2000, widespread adoption of the FSFI has led its inclusion in hundreds of clinical and nonclinical studies worldwide, and more than a thousand published reports at the time of writing. What accounts for the high level of professional and scientific acceptance and broad adoption of the FSFI? Certainly, availability of an extraordinarily large database of published studies, including results from multiple observational and interventional studies in diverse study populations is a significant advantage of the measure to both investigators and granting agencies. The instrument is also relatively brief, can be completed in 5 minutes or less on average, and has a straightforward and practical scoring system. Additionally, the underlying domain structure, which

was designed to assess sexual desire and arousal separately, in addition to other components of sexual function in women, has been validated and replicated in multiple studies.

As noted throughout our review, proper scoring and interpretation of the measure necessitates: a) inclusion of a distress measure for diagnosing sexual dysfunction clinically, as defined by current standards; and b) correction for sexual activity status and adapted scoring for sexually inactive respondents. Of note, despite these considerations and potential limitations, the FSFI was endorsed by a consensus of experts as the preferred self-report measure for assessing sexual function in women by a select panel of expert advisors to the U.S. Food and Drug Administration in mid-2014.⁵³ See the following link for full details of the panel recommendations: (<http://www.fda.gov/media/92963/download>).

Despite the broad acceptance of the measure, additional validation studies are needed in at least 3 areas: (1) the item wording and domain scoring in minority populations needs further validation, particularly in gay and bisexual women or institutionalized women without access to a sexual partner; (2) linguistic adaptation and further cultural validation studies are needed to assess the generalizability of the scoring approach and clinical norms across countries and cultures, and to assess the relative correlates of sexual dysfunctions in other cultures and settings; (3) to investigate further the domain structure of the scale and the relative contributions of desire and arousal difficulties to sexual distress and help-seeking in premenopausal and postmenopausal women with various subtypes of sexual dysfunction. Hopefully, further studies will shed light on these important issues in years to come.

Corresponding Author: Cindy M. Meston, PhD, Department of Psychology, University of Texas at Austin, 108 East Dean Keeton Street, Stop A8000, Austin, TX 78712, USA. Tel: (512) 232-4644; Fax: 512-471-5935; E-mail: meston@psy.utexas.edu

Conflict of Interest: Cindy Meston is a paid consultant for S1 Pharmaceuticals and Strategic Sciences and Technologies. Raymond Rosen is a paid consultant for Strategic Sciences and Technologies. The other authors report no conflicts of interest.

Funding: None.

STATEMENT OF AUTHORSHIP

Category 1

(a) **Conception and Design**
Cindy M. Meston

Category 2

(a) **Drafting the Article**
Cindy M. Meston; Bridget K. Freihart; Ariel B. Handy; Chelsea D. Kilimnik; Raymond C. Rosen
(b) **Revising It for Intellectual Content**
Cindy M. Meston

Category 3

(a) Final Approval of the Completed Article

Cindy M. Meston; Bridget K. Freihart; Ariel B. Handy; Chelsea D. Kilimnik; Raymond C. Rosen

REFERENCES

- Rosen R, Brown C, Heiman J, et al. The female sexual function index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. *J Sex Marital Ther* 2000;26:191-208.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition. American Psychiatric Association; 1994.
- World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. Geneva: WHO; 1992.
- Parish SJ, Meston CM, Althof SE, et al. Toward a more evidence-based nosology and nomenclature for female sexual dysfunctions—part III. *J Sex Med* 2019;16:452-462.
- Reed GM, Drescher J, Krueger RB, et al. Disorders related to sexuality and gender identity in the ICD-11: Revising the ICD-10 classification based on current scientific evidence, best clinical practices, and human rights considerations. *World Psychiatry* 2016;15:205-221.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington, DC: American Psychiatric Publishing; 2013.
- Locke HJ, Wallace KM. Short marital-adjustment and prediction tests: Their reliability and validity. *Marriage Fam Living* 1959;21:251.
- Meston CM. Validation of the Female Sexual Function Index (FSFI) in women with female orgasmic disorder and in women with hypoactive sexual desire disorder. *J Sex Marital Ther* 2003;29:39-46.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. TR. Washington, DC: American Psychiatric Publishing; 2000.
- Wiegel M, Meston C, Rosen R. The Female Sexual Function Index (FSFI): Cross-validation and development of clinical cutoff scores. *J Sex Marital Ther* 2005;31:1-20.
- Gerstenberger EP, Rosen RC, Brewer JV, et al. Sexual desire and the female sexual function index (FSFI): A sexual desire cutpoint for clinical interpretation of the FSFI in women with and without hypoactive sexual desire disorder. *J Sex Med* 2010;7:3096-3103.
- Neijenhuijs KI, Hooghiemstra N, Holtmaat K, et al. The Female Sexual Function Index (FSFI)—A systematic review of measurement properties. *J Sex Med* 2019;16:640-660.
- Meyer-Bahlburg HFL, Dolezal C. The Female Sexual Function Index: A methodological critique and suggestions for improvement. *J Sex Marital Ther* 2007;33:217-224.
- Burri A, Cherkas L, Spector T. Replication of psychometric properties of the FSFI and validation of a modified version (FSFI-LL) assessing lifelong sexual function in an unselected sample of females. *J Sex Med* 2010;7:3929-3939.
- Handy AB, Stanton AM, Pulverman CS, et al. Differences in perceived and physiologic genital arousal between women with and without sexual dysfunction. *J Sex Med* 2018;15:52-63.
- Breiman L, Friedman JH, Olshen RA, et al. Classification and Regression Trees. Routledge. Chapman and Hall CRC; 2017.
- Derogatis LR, Rosen R, Leiblum S, et al. The Female Sexual Distress Scale (FSDS): Initial validation of a standardized scale for assessment of sexually related personal distress in women. *J Sex Marital Ther* 2002;28:317-330.
- Stephenson KR, Toorabally N, Lyons L, et al. Further validation of the Female Sexual Function Index: Specificity and associations with clinical interview data. *J Sex Marital Ther* 2016;42:448-461.
- Verit FF, Verit A. Validation of the Female Sexual Function Index in women with chronic pelvic pain. *J Sex Med* 2007;4:1635-1641.
- Revicki DA, Margolis MK, Bush EN, et al. Content validity of the Female Sexual Function Index (FSFI) in pre- and postmenopausal women with hypoactive sexual desire disorder. *J Sex Med* 2011;8:2237-2245.
- Althof SE, Meston CM, Perelman MA, et al. Opinion Paper: On the diagnosis/classification of sexual arousal concerns in women. *J Sex Med* 2017;14:1365-1371.
- Parish SJ, Goldstein AT, Goldstein SW, et al. Toward a more evidence-based nosology and nomenclature for female sexual dysfunctions-part II. *J Sex Med* 2016;13:1888-1906.
- Lawrance K, Byers ES. Development of the interpersonal exchange model of sexual satisfaction in long-term relationships. *Can J Hum Sex* 1995;1:123-128.
- Byers ES. Relationship satisfaction and sexual satisfaction: A longitudinal study of individuals in long-term relationships. *J Sex Res* 2005;42:113-118.
- Meston C, Trapnell P. Development and validation of a five-factor sexual satisfaction and distress scale for women: The sexual satisfaction scale for women (SSS-W). *J Sex Med* 2005;2:66-81.
- Stephenson KR, Meston CM. Differentiating components of sexual well-being in women: Are sexual satisfaction and sexual distress independent constructs? *J Sex Med* 2010;7:2458-2468.
- Masheb RM, Lozano-Blanco C, Kohorn E, et al. Assessing sexual function and dyspareunia with the Female Sexual Function Index (FSFI) in women with vulvodinia. *J Sex Marital Ther* 2004;30:315-324.
- Smith KB, Pukall CF, Chamberlain SM. Sexual and relationship satisfaction and vestibular pain sensitivity among women with provoked vestibulodynia. *J Sex Med* 2013;10:2009-2023.
- Smith KB, Sadownik LA, Dargie E, et al. Multidisciplinary treatment for provoked vestibulodynia. *Clin J Pain* 2019;35:335-344.
- Anis TH, Gheit SA, Saied HS, et al. Arabic translation of Female Sexual Function Index and validation in an Egyptian population. *J Sex Med* 2011;8:3370-3378.

31. Sun X, Li C, Jin L, et al. Development and validation of a Chinese version of Female Sexual Function Index in a Chinese population—A pilot study. *J Sex Med* 2011;8:1101-1111.
32. Honolina M, Gomez S, Rillon-Tabil N, et al. Translation and validity of the Female Sexual Function Index Filipino version (FSFI-Fil). *Philipp J Intern Med* 2013;51:1-11.
33. Wylomanski S, Bouquin R, Philippe H-J, et al. Psychometric properties of the French Female Sexual Function Index (FSFI). *Qual Life Res* 2014;23:2079-2087.
34. Berner MM, Kriston L, Zahradnik H-P, et al. Validity and reliability of the German Female Sexual Function Index (FSFI-d). *Geburtshilfe Frauenheilkd* 2004;64:293-303.
35. Zachariou A, Filiponi M, Kirana PS. Translation and validation of the Greek version of the Female Sexual Function Index questionnaire. *Int J Impot Res* 2017;29:171-174.
36. Nuring P, Budi Iman S, Denny A, et al. Validation test of Indonesian Female Sexual Function Index (Indonesian FSFI). *Bali Med J* 2018;8:164-168.
37. Filocamo MT, Serati M, Li Marzi V, et al. The Female Sexual Function Index (FSFI): Linguistic validation of the Italian version. *J Sex Med* 2014;11:447-453.
38. Fakhri A, Pakpour AH, Burri A, et al. The Female Sexual Function Index: Translation and validation of an Iranian version. *J Sex Med* 2012;9:514-523.
39. Takahashi M, Inokuchi T, Watanabe C, et al. The Female Sexual Function Index (FSFI): Development of a Japanese version. *J Sex Med* 2011;8:2246-2254.
40. Lee Y, Lim MC, Joo J, et al. Development and validation of the Korean version of the Female Sexual Function Index-6 (FSFI-6K). *Yonsei Med J* 2014;55:1442.
41. Sidi H, Abdullah N, Puteh SEW, et al. The Female Sexual Function Index (FSFI): Validation of the Malay version. *J Sex Med* 2007;4:1642-1654.
42. Ghassamia M, Asghari A, Shaeiri MR, et al. Validation of psychometric properties of the Persian version of the Female Sexual Function Index. *Urol J* 2013;10:878-885.
43. Nowosielski K, Wróbel B, Sioma-Markowska U, et al. Development and validation of the Polish version of the Female Sexual Function Index in the Polish population of females. *J Sex Med* 2013;10:386-395.
44. Thiel Rdo R, Dambros M, Palma PC, et al. [Translation into Portuguese, cross-national adaptation and validation of the Female Sexual Function Index]. *Rev Bras Ginecol Obstet* 2008;30:504-510 [in Portuguese].
45. Vallejo-Medina P, Pérez-Durán C, Saavedra-Roa A. Translation, adaptation, and preliminary validation of the female sexual function index into Spanish (Colombia). *Arch Sex Behav* 2018;47:797-810.
46. Ryding EL, Blom C. Validation of the Swedish version of the Female Sexual Function Index (FSFI) in women with hypoactive sexual desire disorder. *J Sex Med* 2015;12:341-349.
47. Chang SR, Chang TC, Chen KH, et al. Developing and validating a Taiwan version of the Female Sexual Function Index for pregnant women. *J Sex Med* 2009;6:1609-1616.
48. Aygin D. The Turkish adaptation of the Female Sexual Function Index. *Turkiye Klin J Med Sci* 2005;24:393-399.
49. Rehman KU, Asif Mahmood M, Sheikh SS, et al. The Female Sexual Function Index (FSFI): Translation, validation, and cross-cultural adaptation of an Urdu version. *Sex Med* 2015;3:244-250.
50. Wouters C. No sex under my roof: Teenage sexuality in the USA and in the Netherlands since the 1880s. *Política y Sociedad* 2013;50:421-452.
51. Ter Kuile MM, Brauer M, Laan E. The Female Sexual Function Index (FSFI) and the Female Sexual Distress Scale (FSDS): Psychometric properties within a Dutch population. *J Sex Marital Ther* 2006;32:289-304.
52. Boehmer U, Timm A, Ozonoff A, et al. Applying the female sexual functioning index to sexual minority women. *J Women's Health* 2012;21:401-409.
53. U.S. Food and Drug Administration. Center for Drug Evaluation and Research. The Voice of the Patient: Female Sexual Dysfunction. <http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm326192.htm>. Accessed June 26, 2019.