


Female Sexual Function Index Short Version: A MsFLASH Item Response Analysis

Janet S. Carpenter^{1,11}  · Salene M. W. Jones² · Christina R. Studts³ · Julia R. Heiman⁴ · Susan D. Reed⁵ · Katherine M. Newton² · Katherine A. Guthrie⁶ · Joseph C. Larson⁶ · Lee S. Cohen⁷ · Ellen W. Freeman⁸ · R. Jane Lau⁹ · Lee A. Learman⁹ · Jan L. Shifren¹⁰

Received: 29 July 2015 / Revised: 15 June 2016 / Accepted: 28 June 2016 / Published online: 8 August 2016
© Springer Science+Business Media New York 2016

Abstract The Female Sexual Function Index (FSFI) is a psychometrically sound and popular 19-item self-report measure, but its length may preclude its use in studies with multiple outcome measures, especially when sexual function is not a primary endpoint. Only one attempt has been made to create a shorter scale, resulting in the Italian FSFI-6, later translated into Spanish and Korean without further psychometric analysis. Our study evaluated whether a subset of items on the 19-

item English-language FSFI would perform as well as the full-length FSFI in peri- and postmenopausal women. We used baseline data from 898 peri- and postmenopausal women recruited from multiple communities, ages 42–62 years, and enrolled in randomized controlled trials for vasomotor symptom management. Goals were to (1) create a psychometrically sound, shorter version of the FSFI for use in peri- and postmenopausal women as a continuous measure and (2) compare it to the Italian FSFI-6. Results indicated that a 9-item scale provided more information than the FSFI-6 across a spectrum of sexual functioning, was able to capture sample variability, and showed sufficient range without floor or ceiling effects. All but one of the items from the Italian 6-item version were included in the 9-item version. Most omitted FSFI items focused on frequency of events or experiences. When assessment of sexual function is a secondary endpoint and subject burden related to questionnaire length is a priority, the 9-item FSFI may provide important information about sexual function in English-speaking peri- and postmenopausal women.

✉ Janet S. Carpenter
carpentj@iu.edu

- ¹ Department of Science of Nursing Care, School of Nursing, Indiana University, Indianapolis, IN, USA
- ² Group Health Research Institute, Seattle, WA, USA
- ³ Department of Health Behavior, College of Public Health, University of Kentucky, Lexington, KY, USA
- ⁴ Psychological & Brain Sciences, The Kinsey Institute, Indiana University, Bloomington, IN, USA
- ⁵ Departments of Obstetrics/Gynecology and Epidemiology, University of Washington School of Medicine, Seattle, WA, USA
- ⁶ Data Coordinating Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
- ⁷ Center for Women's Mental Health, Massachusetts General Hospital, Boston, MA, USA
- ⁸ Departments of Obstetrics/Gynecology and Psychiatry, School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
- ⁹ Department of Obstetrics and Gynecology, School of Medicine, Indiana University, Indianapolis, IN, USA
- ¹⁰ Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA
- ¹¹ School of Nursing, Indiana University, 600 Barnhill Drive NU 340G, Indianapolis, IN 46202, USA

Keywords Menopause · Psychometrics · Female Sexual Function Index

Introduction

Sexual function is an important component of peri- and postmenopausal women's menopausal quality of life. The 19-item Female Sexual Function Index (FSFI) is one measure that has been popular worldwide. Originally developed in English (Rosen et al., 2000), the scale has been translated into multiple languages (Chang, Chang, Chen, & Lin, 2009; Fakhri, Pakpour, Burri, Morshedi, & Zeidi, 2012; Filocamo et al., 2013; Ghassamia, Asghari, Shaeiri, & Safarinejad, 2013; Giraldo et al., 2012; Kriston, Gunzler, Rohde, & Berner, 2010; Nowosielski, Wrobel, Sioma-Markowska, & Poreba, 2013; Sidi, Abdullah, Puteh,

& Midin, 2007; Sun, Li, Jin, Fan, & Wang, 2011; Takahashi, Inokuchi, Watanabe, Saito, & Kai, 2011). FSFI questions are coded from 0.0 to 5.0. Based on clinical considerations, the scale is considered to have six sexual domains (desire, arousal, lubrication, orgasm, satisfaction, pain), each contributing to the overarching construct of female sexual function (Opperman, Benson, & Milhausen, 2013; Rosen et al., 2000). The maximum score for each domain is 6.0, obtained by summing item responses and multiplying by a correction factor. The total composite sexual function score is a sum of domain scores and ranges from 2.0 (not sexually active and no desire) to 36.0.

The FSFI has demonstrated reliability and validity in a variety of populations. A total score of 26.55 was identified as the threshold for differentiating those with and without sexual dysfunction in a sample of 568 women, 66 % of whom were premenopausal (Wiegel, Meston, & Rosen, 2005). Use of the measure in postmenopausal women suggests that a lower threshold of 20 may be appropriate for identifying women with low sexual function (Reed et al., 2012, 2014). In addition, a sexual desire domain score of 5.0 or less has been suggested as a threshold for hypoactive sexual desire disorder (Gerstenberger et al., 2010).

Despite being widely used, psychometrically sound, and clinically interpretable, the full FSFI may be too long for use in research utilizing long assessment batteries, especially when assessing sexual function is not a principal goal of a study. A PubMed search using the keywords “FSFI and (validation or psychometrics)” produced 61 references, none of which focused on psychometric testing to produce a shorter English-language FSFI. However, three articles did pertain to non-English-language FSFI versions. A shorter 6-item Italian-language FSFI was first developed by Isidori et al. (2010). Participants included 160 Italian women aged 21–49 who reported sexual activity in the past month and participated in two research sessions. The women were recruited from outpatient sexual and reproductive medicine clinics, and they completed a 19-item Italian FSFI along with a complete medical consultation and physical examination. Based on receiver operating curves (ROCs) generated for each individual item, a 6-item version was created using one item from each of the original six domains (desire, arousal, lubrication, orgasm, satisfaction, and pain) in the FSFI, with response options from 1 = *poor function* to 5 = *optimal function* for all questions and an additional 0 response for four questions to indicate no sexual activity in the past month. An FSFI score of ≤ 19.0 showed excellent sensitivity and specificity in identifying women with sexual dysfunction, as assessed by the full-length FSFI and medical examination, which demonstrated 100 % convergence. Cronbach’s alpha was 0.79, and test–retest reliability at 18–24 days was high ($r = .95, p < .0001$). The FSFI-6 was subsequently translated into Spanish (Chedraui et al., 2012; Perez-Lopez, Fernandez-Alonso, Trabolon-Pastor, Vara, & Chedraui, 2012) and Korean (Lee et al., 2014). These versions demonstrated strong internal consistency reliability ($\alpha = 0.91$) but their validity was not assessed.

The purpose of our analysis was to evaluate how well a subset of items from the English-language 19-item FSFI (Rosen et al., 2000) performed in peri- and postmenopausal women enrolled in treatment trials for hot flashes. We sought to develop a short English-language form using modern psychometric methods to maximize measurement information while reducing participant burden. To address the need for a shorter English-language scale and the limitations of Isidori et al.’s (2010) methods (e.g., small sample size and reliance on 19 separate analyses), we performed an item response theory analysis on baseline data from 898 peri- and postmenopausal women who participated in trials conducted within the multisite Menopause Strategies: Lasting Answers to Symptoms and Health (MsFLASH) research network. Goals of the analysis were to (1) create a psychometrically sound, shorter version of the FSFI for use with peri- and postmenopausal women that could be used as a single continuous measure for secondary outcomes and (2) compare it to the previously devised 6-item version (Isidori et al., 2010).

Method

Participants

This was a cross-sectional analysis using baseline data from 898 peri- and postmenopausal, community-dwelling women reporting hot flashes who participated in MsFLASH Trials 01, 02, and 03. The full details of these trials have been reported elsewhere (Cohen et al., 2014; Freeman et al., 2011; Joffe et al., 2014; Newton et al., 2014b; Sternfeld et al., 2014). Briefly, the trials were designed to evaluate pharmaceutical, nutraceutical, and behavioral interventions for menopausal hot flash management (Newton et al., 2014a; Sternfeld et al., 2013). Trial 01 was a multisite, randomized, placebo-controlled, double-blind trial comparing escitalopram to placebo in African American and white women. Trial 02 was a multisite, three-by-two factorial, randomized, controlled trial evaluating exercise, yoga, or usual activity and omega-three fatty acid supplements or placebo. Trial 03 was a multisite, randomized, placebo-controlled, double-blind trial of low-dose 17-beta-estradiol, venlafaxine, or placebo.

All studies were approved by institutional review boards at the Data Coordinating Center (Seattle) and the participating clinical sites. Participants were recruited mainly through mass mailings to age-eligible women using health-plan enrollment data and purchased mailing lists. All participants in all studies provided written informed consent and signed authorization to use protected health information. Common to all trials, participants were aged 40–62; peri- or postmenopausal; in good general health based on self-report, vital signs, and blood tests; not using treatments for hot flashes; and reporting no drug or alcohol abuse in the past year or a major depressive episode in

the past three months. Eligible women reported frequent weekly hot flashes (≥ 28 per week in Trial 01, ≥ 14 in Trials 02 and 03) that were bothersome or severe on four or more days or nights per week. Women were enrolled from clinical sites located in Boston, Indianapolis, Oakland, Philadelphia, and Seattle. All data analyzed here were collected during the baseline, prerandomization trial periods.

Measures

The 19-item FSFI was included in a larger questionnaire battery administered at baseline and postintervention. It was disproportionately longer compared to other scales used to measure other symptoms and experiences (Newton et al., 2014a) which resulted in questions from participants about its importance. Sexual functioning over the past four weeks was evaluated (Rosen et al., 2000). The standard formula-based scoring was used to obtain total scores ranging from 2.0 to 36.0 and domain scores ranging from 0.8 to 6.0 for satisfaction, 1.2–6.0 for desire, and 0.0–6.0 for arousal, lubrication, orgasm, and pain. Domain scores of 0.0 indicate no sexual activity during the past month. Higher domain and total scores indicate more optimal sexual functioning.

To determine how bothered or distressed women were by their levels of sexual function, we adapted a single question from the Female Sexual Distress Scale: “In the past four weeks, how often did you feel distressed or bothered about your sex life?” Scoring was: 0 = *never*, 1 = *rarely*, 2 = *occasionally*, 3 = *frequently*, and 4 = *always* (Derogatis, Rosen, Leiblum, Burnett, & Heiman, 2002).

Baseline demographic characteristics collected from all women included age, race, ethnicity, menopausal status, education, and income. Height and weight were collected in clinic by study staff to calculate body mass index.

Data Analysis

After sample demographics and scale scores were analyzed using descriptive statistics, item response theory (IRT) was the main analysis method. Analyses were conducted with IRTPRO 2.0 (Scientific Software International, 2013). Psychometric analyses using IRT are model based, estimating the probability of item responses as a function of the level of the underlying construct being measured (Hambleton & Swaminathan, 1985). Items are “calibrated” using IRT models, yielding parameter estimates that characterize item-level measurement performance. These parameter estimates can be generalized via linear transformation from one sample to another from the same population, unlike psychometric indices obtained via traditional classical test theory methods (e.g., summed score), which are limited to the samples investigated. With 500–1000 people sampled (Reise & Yu, 1990), the idea with IRT is that stable item parameters can be estimated, facilitating estimation of individuals’ IRT scores. The use of IRT

and IRT scores is suggested as an alternative to avoid many of the pitfalls of short-form development, including the need to evaluate in another sample, since the selected items are specifically chosen because of their accurate measurement of targeted levels of the underlying construct (Smith, McCarthy, & Anderson, 2000).

Another advantage of IRT is its ability to handle missing data (Bock & Aitkin, 1981; Lord, 1980). Because analyses focus on estimating item properties rather than participant characteristics, when participants miss or skip a particular item, their responses to other items are still preserved and used. In MsFLASH 01 ($n = 195$), one of the FSFI satisfaction questions was inadvertently missing. This single question, one of three questions in the satisfaction domain, was: “Over the past four weeks, how satisfied have you been with your overall sex life?” For examining descriptive scale scores, we used a mean imputation where item scores were imputed as an average of the answers to the two other questions in the FSFI satisfaction domain. Imputed scores were not used for the IRT analysis (Bock & Aitkin, 1981; Lord, 1980). All women were included in the IRT analyses; however, items for which they reported no sexual activity were coded as missing, not as numerical values.

The IRT models used in this study calculate two types of parameters for each item: difficulty and discrimination (Hays, Morales, & Reise, 2000). Difficulty parameters (represented by b) show what level of a trait or construct an item best measures; for example, in this study “easy” items (or those with low difficulty parameters) provide the most information in measuring lower female sexual functioning, whereas “difficult” items (or those with high difficulty parameters) provide the most information in measuring better female sexual functioning. In the case of items with multiple response options, such as those in the FSFI, several difficulty parameters are calculated, specifically, one fewer than the number of response options (Samejima, 1969). Difficulty parameter b_1 represents the level of sexual functioning required for a randomly selected participant to select response option 1 instead of 0; difficulty parameter b_2 represents the level of sexual functioning required for a randomly selected participant to select response option 2 rather than 1, and so on. Discrimination parameters (represented by a) reveal how accurately an item measures the underlying construct at its difficulty level. For example, if items X and Y have very similar difficulty parameter estimates, but item X has a higher discrimination parameter than item Y, then item X provides more discrimination among participants with sexual function near those difficulty levels than item Y.

Using these parameters, IRT analyses determine the amount of measurement information each item provides at specific levels of the underlying construct of interest (i.e., female sexual functioning). Information levels can be interpreted as the degree of measurement precision provided by an item at various levels of the underlying construct (e.g., a screening measure should provide the most information around the screening point, where-

Table 1 Item response theory difficulty and discrimination parameters for retained items

FSFI item (domain)	Version	Difficulty ^a				Discrimination <i>a</i>
		<i>b</i> ₁	<i>b</i> ₂	<i>b</i> ₃	<i>b</i> ₄	
1. How often feel sexual desire or interest (desire)		−0.53	0.58	1.24	1.94	2.47
2. Level of sexual desire or interest (desire)	9, 6	−0.56	0.46	1.42	2.06	2.61
3. How often sexually aroused (arousal)		na	na	na	na	na
4. How often satisfied with arousal (arousal)		na	na	na	na	na
5. Level of sexual arousal (arousal)	9, 6	−0.84	−0.03	0.75	1.52	2.91
6. Confidence in becoming aroused (arousal)		na	na	na	na	na
7. How often become lubricated (lubrication)	6	−0.59	0.02	0.44	0.89	2.97
8. How often maintain lubrication (lubrication)	9	−0.54	0.01	0.53	1.02	2.65
9. Difficulty becoming lubricated ^b (lubrication)	9	−0.87	−0.49	0.00	0.78	2.89
10. Difficulty maintaining lubrication ^b (lubrication)		−0.91	−0.53	−0.02	0.76	2.34
11. How often reach orgasm	9, 6	−1.24	−0.52	0.07	0.71	1.51
12. Difficulty reaching orgasm ^b (orgasm)	9	−1.30	−0.77	−0.35	0.78	2.01
13. Satisfaction with ability to orgasm (orgasm)		−1.01	−0.39	0.14	0.82	2.13
14. Satisfaction with amount of emotional closeness with partner (satisfaction)	9	−1.60	−0.83	−0.24	0.79	1.37
15. How often discomfort/pain during vaginal penetration ^b (pain)		−1.12	−0.71	−0.24	0.23	1.81
16. How often discomfort/pain following vaginal penetration ^b (pain)		−1.56	−1.19	−0.75	−0.22	1.48
17. Level of discomfort/pain during or following vaginal penetration ^b (pain)	9, 6	−1.99	−1.29	−0.49	0.23	1.62
18. Satisfaction with sexual relationship with partner (satisfaction)		−1.05	−0.30	0.35	1.05	1.90
19. Satisfaction with overall sex life (satisfaction)	9, 6	−0.88	−0.10	0.55	1.29	1.96

^a The difficulty parameter estimates ($b_1 - b_4$) indicate the levels of female sexual function at which the probability of selecting the next higher response option shifts to being higher than the probability of selecting the current response option, on a scale with a mean of 0 and a standard deviation of 1. In this case, negative numbers indicate lower levels of sexual functioning and positive numbers reflect higher levels. The discrimination parameter estimate (a) indicates how accurately the item measures female sexual functioning, with higher values indicating more accurate measurement

^b Reverse scored, 9 = selected for the 9-item short form, 6 = included in the 6-item short form developed by Isidori et al., na = item not retained in IRT analysis

as an instrument intended to measure the full range of a construct should provide high levels of information along the entire continuum). Careful consideration of estimates of item difficulty, discrimination, and information can facilitate instrument development by guiding selection of items that are most informative at specified levels of the construct of interest. Thus, IRT analyses can be used to (1) reduce respondent burden by eliminating unnecessary or redundant items (e.g., the item Y described above would be considered redundant), (2) ensure reliable measurement of the latent construct along its entire continuum by eliminating items leading to floor or ceiling effects, and (3) ensure reliable measurement at specific points at which more precision is needed.

The IRT analyses of the FSFI included the following: (a) fitting an appropriate IRT model (the graded response model) (Samejima, 1969) to the ordinal-level data capturing participant responses to each item; (b) calibrating the items to obtain item difficulty parameters, item discrimination parameters, and item information estimates; and (c) identifying the subset of items that simultaneously maximized the scale's measurement information along the spectrum of female sexual functioning while minimizing the number of items required in the scale. We utilized the IRT analyses to create a short form of the FSFI with the a priori

requirement that at least one item from each facet of female sexual function (desire, arousal, lubrication, orgasm, satisfaction, and pain) was included to ensure adequate coverage of the construct, as was done by a previous team (Isidori et al., 2010). To create a shorter version of the FSFI that could be used for rapid assessment, we aimed to select a set of items that would be informative at different levels of sexual functioning, both above and below the sample mean.

Following item selection, we converted participants' IRT scores obtained with IRTPRO 2.0 to summed scale scores for the set of selected items using the test characteristic curve. The test characteristic curve plots the IRT scores on the x-axis against the traditional summed scores on the y-axis. To compare sexual functioning scores to sexual distress, we first classified participants responding "frequently" or "always" on the sexual distress item as having high sexual distress. We then created three different categorizations of sexual functioning using the summed scale scores equivalent to IRT scores of -0.5 , -1.0 , and -1.5 , with participants scoring below each of these points classified as having low sexual functioning and those scoring above as having high sexual functioning. Finally, we examined the associations between sexual distress ($1 = high$, $0 = low$) and

each of the three categorizations of sexual functioning (1 = *high*, 0 = *low*) using chi-square tests.

Results

Consistent with the parent trials' inclusion criteria, the 898 women in the sample were on average 54.47 years of age (*SD* 3.83). Most were married or living with a partner (63.2 %), had completed a bachelor's degree or higher (52.7 %), and were employed full or part time (69.7 %). Across all three studies, 62.4 % were Caucasian, 33.9 % African American, 2.8 % Hispanic, 2.0 % Native American, 2.3 % Asian American, and 3.3 % another race/ethnicity. Women in the sample were 18.1 % perimenopausal and 81.9 % postmenopausal. Most (74.4 %) reported at least some sexual activity on the FSFI.

Item Response Theory Analyses

The initial IRT model using all 19 items of the FSFI resulted in significant $S-\chi^2$ values for the four arousal items (all $p < .0001$ with Bonferroni-corrected alpha of 0.003), indicating violation of the local independence assumption of IRT. Thus, only Item 5 (level of sexual arousal during sexual activity or intercourse) was retained in the model since this aspect of female sexual function was deemed important to assess. We chose this item because it showed the best ability to identify female sexual dysfunction in the study that developed the previous short form (Isidori et al., 2010). We then ran the IRT model again on the remaining 16 items, and the local independence assumption did not appear to be violated (all $p > .01$ with a Bonferroni-corrected alpha of 0.003).

As would be expected, the 16 remaining items from the FSFI had a range of difficulty and discrimination parameter estimates (Table 1). Between the two desire items (Items 1 and 2), we chose Item 2 because it had better discrimination and measured a wider range of the construct as shown by the difficulty parameters. For arousal, we used Item 5 for the same reasons. For lubrication (Items 7–10), we selected Items 8 and 9 because they provided better discrimination than Item 10 and together measured a greater range of sexual function than Item 7. For orgasm (Items 11–13), we eliminated Item 13 because of its narrow measurement range and included Items 11 and 12 because both had wider coverage of the construct than Item 13. For satisfaction (Items 14, 18, and 19), we included Items 14 and 19 because Item 14 was one of the few Items to measure very low levels of the construct and Item 19 was one of the few items to measure very high levels of the construct. For pain (Items 15–17), we included Item 17 because it was one of the few Items measuring very low levels of sexual function. This created a 9-item measure that could assess most levels of the construct without requiring all 19 items. Of note, 5 of the 6 items from the previous Italian short form (Isidori et al., 2010) were included in the 9-item measure.

Descriptive statistics for the three versions of the FSFI (full scale, 9-item, and 6-item) are shown in Table 2.

Because of the IRT assumption of local independence, the information offered by a given set of items can be determined by simply adding the information levels of the individual items comprising the set. This cumulative information is referred to as test-level information, and we examined it for each version of the scale (Fig. 1) to visually compare the amount and distribution of measurement information offered by each. As would be expected, the 16 items provided more information than either the 9-item scale or the 6-item scale (Isidori et al., 2010). However, the 9-item scale provided more information (i.e., had less error and greater precision) at all levels of sexual function than the 6-item scale. This was particularly evident from 1.5 SDs below the mean to 1.5 SDs above the mean.

Sexual Functioning Groups

The test characteristic curve for the 9-item short form (Fig. 2) shows the corresponding summed score for each IRT score. A score of 1.5 SDs below the mean (IRT score of -1.5) corresponded to a score of 6.5 on the 9-item short form. An IRT score of -1.0 corresponded to a scale score of 10, and an IRT score of -0.5 corresponded to a score of 15.0.

We then compared the scores on the 9-item short form to the sexual distress item. Only the 599 participants from Trials 02 and 03 were included in these descriptive analyses—women from the first study were excluded because of the missing item problem described above. Women who reported low sexual function on the 9-item short form had significantly higher distress due to sexual function than women with high sexual function using scores of 15.0 ($\chi^2 = 19.69$, $p < .001$), of 10.0 ($\chi^2 = 7.41$, $p = .01$), and of 7.0 ($\chi^2 = 6.56$, $p = .01$) (Table 3).

Discussion

Using data from a sample of community-dwelling peri- and postmenopausal women experiencing hot flashes, we created a 9-item version of the FSFI. The 9-item version provided information across the entire spectrum of sexual functioning, and descriptive statistics showed that it was able to capture variability within the sample (e.g., large SDs) and had sufficient range.

The test-level information provided by the new 9-item scale proposed here points to the relative importance of the scale. Although our PubMed search revealed that the FSFI has been widely used and is psychometrically sound, we found only one psychometric analysis that was performed in an attempt to create a shorter scale. The prior work by Isidori et al. (2010) led to subsequent translations of the 6-item version without further psychometric analysis. All but one of the items from the 6-item version were included in the 9-item version. The one exception was the sexual arousal item pertaining to how often a woman

Table 2 Descriptive statistics for the three versions of the Female Sexual Function Index (FSFI)

	Possible range	Actual range	M	SD
19-item full scale	1.2–36.0	1.2–36.0	18.2	10.9
9-item short form	2.0–45.0	2.0–45.0	22.5	13.7
6-item short form ^a	2.0–30.0	2.0–30.0	14.9	8.4

^a Isidori et al. (2010); possible score range for the 19-item full scale uses a formula with a maximum of 6 points in each of the six domains, with each domain having a different number of items; possible score range for 9- and 6-item versions uses sum of items

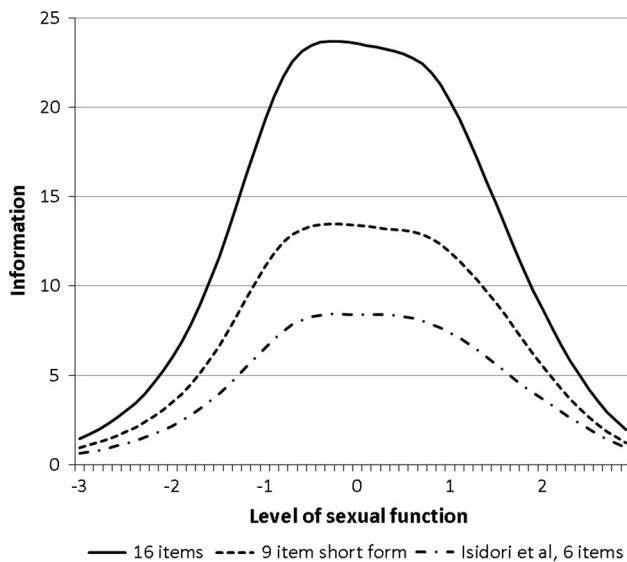


Fig. 1 Test-level information curves for each version of the Female Sexual Function Index. *Legend* For level of sexual function, 0 represents the mean and the standard deviation is 1. Higher levels of information indicate more precise measurement

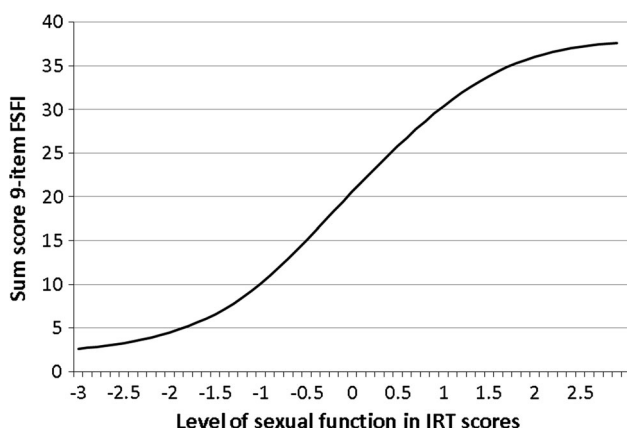


Fig. 2 Test characteristic curve for FSFI-9 item. *Legend* This curve shows the conversion of each item response theory (IRT) score (mean of 0 and standard deviation of 1) to the corresponding 9-item FSFI summed score (ranges from 2 to 45). For example, an IRT score of 0 corresponds to a summed score of about 20 and an IRT score of -0.5 corresponds to a summed score of 15

reported that she became lubricated (“wet”) during sexual activity or intercourse. Being able to maintain lubrication and not having difficulty becoming lubricated were found to be more informative items. These items may have been particularly important for our peri- and postmenopausal sample because of their unique patterns of symptoms related to hormonal changes; thus, findings need to be replicated in other age groups.

An unexpected finding was that, overall, most of the items omitted from the full FSFI in developing both the 9-item and prior 6-item versions were those measuring the frequency of an event or experience. In our sample of peri- and postmenopausal women, sexual frequency over the past month may have been a relatively less accurate measure of female sexual function since it also reflects partner desire and physical capability and/or a couple’s typical sexual behavior patterns (Adams, Gold, & Burt, 1978). It is estimated that 52 % of American men aged 40–70 years are affected by some degree of erectile dysfunction (O’Donnell, Araujo, & McKinlay, 2004). Although not a preconceived study hypothesis, results of the IRT-guided selection of items may point to the relatively greater importance of severity and difficulty of experiences, rather than frequency, for assessing peri- and postmenopausal women’s sexual functioning. A frequency of no sexual activity is assigned a score of 0, which would only correctly reflect the lowest sexual functioning if the lack of sexual activity was related to the symptoms assessed by the items and not to other reasons (Baser, Li, & Carter, 2012). This finding should be explored further, since its implications may be important in clinical trials and other treatment studies that aim to use the FSFI as an outcome.

Findings from two previous studies may provide context for the arousal items showing local dependence in our analysis. The original measurement model ($n = 259$ women) yielded a 5-factor solution, with the arousal items actually loading onto the desire factor, but six factors were retained for “clinical considerations” (Rosen et al., 2000). In a subsequent article, Opperman et al. (2013) compared several models of the FSFI, including a 6-factor model as originally suggested and a 5-factor model combining the desire and arousal subscales ($n = 85$ women). Both the 5- and 6-factor models were supported. Combining desire and arousal is consistent with DSM-5 changes in definitions of female sexual dysfunctions (American Psychiatric Association, 2013). Desire and arousal disorders are now combined into a single disorder, female sexual interest/arousal disorder, since the distinction between these phases of the sexual response

Table 3 Comparison of high and low sexual function on distress

9-item FSFI total score	IRT score	% with low sexual function	% of low with distress	% with high sexual function	% of high with distress
<7	−1.5	26	31	74	20
<10	−1.0	28	30	72	20
<15	−0.5	32	33	68	18

$n = 599$, analysis included only women in Trials 02 and 03. Proportion of participants with low sexual function was determined by three scores on the 9-item FSFI (see first column), and distress was determined by participants responding “often” or “frequently” to a question about distress due to sexual function

cycle may be artificial. That desire and arousal in female sexual function are so closely related could explain the problems with the arousal items in our initial analyses, although it should be noted that the intention of this study was not to examine the latent structure of female sexual function.

Our findings suggest that, when a shorter FSFI is desired in peri- and postmenopausal samples experiencing hot flashes, the 9-item version may be advantageous for use as a single, continuous measure, particularly when participant burden is a consideration. The 9-item version demonstrated the ability to differentiate between peri- and postmenopausal women categorized by self-reported levels of sexual function and sexual function with distress, using three different potential categorizations of low versus high sexual functioning. However, our results were based only on known groups’ validity with groups defined as high versus low sexual functioning. We avoided the terms sexually functional and sexually dysfunctional, because a gold standard for assessment of sexual function such as a clinical interview by an expert in sexual function was not performed in this study. Further evaluation with a gold standard sexual function assessment will be beneficial.

Differences in items selected for our shortened FSFI scale versus Isidori et al.’s (2010) 6-item Italian version may be at least partially explained by differences in the populations studied and analytic methods. Our population was older (mean age 54.49 vs. 34.9), focused on peri- and postmenopausal women (100 vs. 4 %), and largely recruited from the community rather than during clinical visits. In addition, our IRT analysis differed from the classical test theory approach used by Isidori et al., which relied solely on sample-dependent summed score methods.

Study findings should be interpreted in view of the following study limitations. An assessment of sexual activity, partner gender, and history of physical and sexual abuse was conducted in Trial 03 only. Therefore, these data were not available for the majority of women included in this analysis. In addition, findings are generalizable to a population of symptomatic peri- and postmenopausal women, but should be interpreted cautiously or replicated in women of different ages and different medical conditions. Our population of peri- and postmenopausal women may have had particular symptoms that affected sexual functioning such as vaginal dryness and subsequent dyspareunia, but the

women were not recruited based on sexual function and vaginal dryness, which are only marginally linked to hot flashes (Carpenter et al., 2015). Our findings also reflect a population experiencing hot flashes and may not generalize to the minority 20 % of women who do not experience this cardinal menopausal symptom. The FSFI does not assess women’s bother or concern related to sexual function. This could explain why a fairly large minority of women reporting high sexual function also reported distress. Finally, we were not able to compare short-version summed scores to an external criterion such as a clinical interview, the gold standard for assessing female sexual dysfunction.

In summary, IRT analyses guided the development of a 9-item English-language version of the FSFI that was more informative when used with peri- and postmenopausal women experiencing hot flashes than a previously developed 6-item Italian version. In studies in which sexual function is the primary outcome measure, the 19-item FSFI should be used since it is the most informative. When assessment of sexual function is just one of many secondary endpoints and subject burden related to questionnaire length is a priority, this shorter version of the FSFI may allow researchers to obtain important information on sexual function in peri- and postmenopausal women experiencing hot flashes.

Acknowledgments This study was funded by the National Institutes of Health as a cooperative agreement issued by the National Institute on Aging (NIA), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), the National Center for Complementary and Alternative Medicine (NCCAM), the Office of Research on Women’s Health (ORWH), and Grants U01AG032656, U01AG032659, U01AG032669, U01AG032682, U01AG032699, and U01AG032700 from the NIA. At Indiana University, the project was funded in part with support from the Indiana Clinical and Translational Sciences Institute, Grant UL1RR02571 from the NIH, National Center for Research Resources, Clinical and Translational Sciences Award.

References

- Adams, D. B., Gold, A. R., & Burt, A. D. (1978). Rise in female-initiated sexual activity at ovulation and its suppression by oral contraceptives. *New England Journal of Medicine*, 299, 1145–1150. doi:10.1056/NEJM197811232992101.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.

- Baser, R. E., Li, Y., & Carter, J. (2012). Psychometric validation of the Female Sexual Function Index (FSFI) in cancer survivors. *Cancer*, *118*, 4606–4618. doi:10.1002/ncr.26739.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, *46*, 443–459. doi:10.1007/BF02293801.
- Carpenter, J. S., Woods, N. F., Otte, J. L., Guthrie, K. A., Hohensee, C., Newton, K. M., & LaCroix, A. Z. (2015). MsFLASH participants' priorities for alleviating menopausal symptoms. *Climacteric*, *18*, 859–866. doi:10.3109/13697137.2015.1083003.
- Chang, S. R., Chang, T. C., Chen, K. H., & Lin, H. H. (2009). Developing and validating a Taiwan version of the Female Sexual Function Index for pregnant women. *Journal of Sexual Medicine*, *6*, 1609–1616. doi:10.1111/j.1743-6109.2009.01247.x.
- Chedraui, P., Perez-Lopez, F. R., Sanchez, H., Aguirre, W., Martinez, N., Miranda, O., & Zambrano, B. (2012). Assessment of sexual function of mid-aged Ecuadorian women with the 6-item Female Sexual Function Index. *Maturitas*, *71*, 407–412. doi:10.1016/j.maturitas.2012.01.013.
- Cohen, L. S., Joffe, H., Guthrie, K. A., Ensrud, K. E., Freeman, M., Carpenter, J. S., & Anderson, G. L. (2014). Efficacy of omega-3 for vasomotor symptoms treatment: A randomized controlled trial. *Menopause*, *21*, 347–354. doi:10.1097/GME.0b013e31829e40b8.
- Derogatis, L. R., Rosen, R., Leiblum, S., Burnett, A., & Heiman, J. (2002). The Female Sexual Distress Scale (FSDS): Initial validation of a standardized scale for assessment of sexually related personal distress in women. *Journal of Sex and Marital Therapy*, *28*, 317–330. doi:10.1080/00926230290001448.
- Fakhri, A., Pakpour, A. H., Burri, A., Morshedi, H., & Zeidi, I. M. (2012). The Female Sexual Function Index: Translation and validation of an Iranian version. *Journal of Sexual Medicine*, *9*, 514–523. doi:10.1111/j.1743-6109.2011.02553.x.
- Filocamo, M. T., Serati, M., Li Marzi, V., Costantini, E., Milanese, M., Pietropaolo, A., ... Villari, D. (2013). The Female Sexual Function Index (FSFI): Linguistic validation of the Italian version. *Journal of Sexual Medicine*, *11*, 447–453. doi:10.1111/jsm.12389.
- Freeman, E. W., Guthrie, K. A., Caan, B., Sternfeld, B., Cohen, L. S., Joffe, H., ... LaCroix, A. Z. (2011). Efficacy of escitalopram for hot flashes in healthy menopausal women: A randomized controlled trial. *Journal of the American Medical Association*, *305*, 267–274. doi:10.1001/jama.2010.2016.
- Gerstenberger, E. P., Rosen, R. C., Brewer, J. V., Meston, C. M., Brotto, L. A., Wiegel, M., & Sand, M. (2010). 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. *Journal of Sexual Medicine*, *7*, 3096–3103. doi:10.1111/j.1743-6109.2010.01871.x.
- Ghassamia, M., Asghari, A., Shaehri, M. R., & Safarinejad, M. R. (2013). Validation of psychometric properties of the Persian version of the Female Sexual Function Index. *Urology Journal*, *10*, 878–885. doi:10.3109/13651501.2014.940048.
- Giraldo, P. C., Polpetta, N. C., Juliato, C. R., Yoshida, L. P., do Amaral, R. L., & Eleuterio Junior, J. (2012). Evaluation of sexual function in Brazilian women with recurrent vulvovaginal candidiasis and localized provoked vulvodynia. *Journal of Sexual Medicine*, *9*, 805–811. doi:10.1111/j.1743-6109.2011.02584.x.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Norwell, MA: Kluwer Academic.
- Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care*, *38*, II28–II42. doi:10.1097/00005650-200009002-00007.
- Isidori, A. M., Pozza, C., Esposito, K., Giugliano, D., Morano, S., Vignozzi, L., ... Jannini, E. A. (2010). Development and validation of a 6-item version of the Female Sexual Function Index (FSFI) as a diagnostic tool for female sexual dysfunction. *Journal of Sexual Medicine*, *7*, 1139–1146. doi:10.1111/j.1743-6109.2009.01635.x.
- Joffe, H., Guthrie, K. A., LaCroix, A. Z., Reed, S. D., Ensrud, K. E., Manson, J. E., & Cohen, L. (2014). Low-dose estradiol and the serotonin-norepinephrine reuptake inhibitor venlafaxine for vasomotor symptoms: A randomized clinical trial. *JAMA Internal Medicine*, *174*, 1058–1066. doi:10.1001/jamainternmed.2014.1891.
- Kriston, L., Gunzler, C., Rohde, A., & Berner, M. M. (2010). Is one question enough to detect female sexual dysfunctions? A diagnostic accuracy study in 6,194 women. *Journal of Sexual Medicine*, *7*, 1831–1841. doi:10.1111/j.1743-6109.2010.01729.x.
- Lee, Y., Lim, M. C., Son, Y., Joo, J., Park, K., Kim, J. S., & Park, S. Y. (2014). Development and evaluation of Korean version of Quality of Sexual Function (QSF-K) in healthy Korean women. *Journal of Korean Medical Science*, *29*, 758–763. doi:10.3346/jkms.2014.29.6.758.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: L. Erlbaum Associates.
- Newton, K. M., Carpenter, J. S., Guthrie, K. A., Anderson, G. L., Caan, B., Cohen, L. S., & Lacroix, A. Z. (2014a). Methods for the design of vasomotor symptom trials: The Menopausal Strategies Finding Lasting Answers to Symptoms and Health Network. *Menopause*, *21*, 45–58. doi:10.1097/GME.0b013e31829337a4.
- Newton, K. M., Reed, S. D., Guthrie, K. A., Sherman, K. J., Booth-Laforce, C., Caan, B., & Lacroix, A. Z. (2014b). Efficacy of yoga for vasomotor symptoms: A randomized controlled trial. *Menopause*, *21*, 339–346. doi:10.1097/GME.0b013e31829e4baa.
- Nowosielski, K., Wrobel, B., Sioma-Markowska, U., & Poreba, R. (2013). Development and validation of the Polish version of the Female Sexual Function Index in the Polish population of females. *Journal of Sexual Medicine*, *10*, 386–395. doi:10.1111/jsm.12012.
- O'Donnell, A. B., Araujo, A. B., & McKinlay, J. B. (2004). The health of normally aging men: The Massachusetts Male Aging Study (1987–2004). *Experimental Gerontology*, *39*, 975–984. doi:10.1016/j.exger.2004.03.023.
- Opperman, E. A., Benson, L. E., & Milhausen, R. R. (2013). Confirmatory factor analysis of the Female Sexual Function Index. *Journal of Sex Research*, *50*, 29–36.
- Perez-Lopez, F. R., Fernandez-Alonso, A. M., Trabalon-Pastor, M., Vara, C., & Chedraui, P. (2012). Assessment of sexual function and related factors in mid-aged sexually active Spanish women with the six-item Female Sex Function Index. *Menopause*, *19*, 1224–1230. doi:10.1097/gme.0b013e3182546242.
- Reed, S. D., Guthrie, K. A., Joffe, H., Shifren, J. L., Seguin, R. A., & Freeman, E. W. (2012). Sexual function in nondepressed women using escitalopram for vasomotor symptoms: A randomized controlled trial. *Obstetrics and Gynecology*, *119*, 527–538. doi:10.1097/AOG.0b013e3182475fa4.
- Reed, S. D., Mitchell, C. M., Joffe, H., Cohen, L., Shifren, J. L., Newton, K. M., ... Guthrie, K. A. (2014). Sexual function in women on estradiol or venlafaxine for hot flashes: A randomized controlled trial. *Obstetrics & Gynecology*, *124*, 233–241. doi:10.1097/AOG.0000000000000386.
- Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model. *Journal of Educational Measurement*, *27*, 133–144.
- Rosen, R., Brown, C., Heiman, J., Leiblum, S., Meston, C., Shabsigh, R., ... D'Agostino, R., Jr. (2000). The Female Sexual Function Index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. *Journal of Sex & Marital Therapy*, *26*, 191–208.
- Samejima, F. (1969). *Calibration of latent ability using a response pattern of graded scores*. *Psychometrika Monograph No. 17*. Richmond, VA: Psychometric Society.
- Scientific Software International. (2013). *IRTPRO user guide*. Skokie, IL: Scientific Software International.

- Sidi, H., Abdullah, N., Puteh, S. E., & Midin, M. (2007). The Female Sexual Function Index (FSFI): Validation of the Malay version. *Journal of Sexual Medicine, 4*, 1642–1654. doi:[10.1111/j.1743-6109.2007.00476.x](https://doi.org/10.1111/j.1743-6109.2007.00476.x).
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102–111.
- Sternfeld, B., Guthrie, K. A., Ensrud, K. E., Lacroix, A. Z., Larson, J. C., Dunn, A. L., & Caan, B. J. (2014). Efficacy of exercise for menopausal symptoms: A randomized controlled trial. *Menopause, 21*, 330–338. doi:[10.1097/GME.0b013e31829e4089](https://doi.org/10.1097/GME.0b013e31829e4089).
- Sternfeld, B., Lacroix, A., Caan, B. J., Dunn, A. L., Newton, K. M., Reed, S. D., & Ensrud, K. E. (2013). Design and methods of a multi-site, multi-behavioral treatment trial for menopausal symptoms: The MsFLASH experience. *Contemporary Clinical Trials, 35*, 25–34. doi:[10.1016/j.cct.2013.02.009](https://doi.org/10.1016/j.cct.2013.02.009).
- Sun, X., Li, C., Jin, L., Fan, Y., & Wang, D. (2011). Development and validation of Chinese version of Female Sexual Function Index in a Chinese population—A pilot study. *Journal of Sexual Medicine, 8*, 1101–1111. doi:[10.1111/j.1743-6109.2010.02171.x](https://doi.org/10.1111/j.1743-6109.2010.02171.x).
- Takahashi, M., Inokuchi, T., Watanabe, C., Saito, T., & Kai, I. (2011). The Female Sexual Function Index (FSFI): Development of a Japanese version. *Journal of Sexual Medicine, 8*, 2246–2254. doi:[10.1111/j.1743-6109.2011.02267.x](https://doi.org/10.1111/j.1743-6109.2011.02267.x).
- Wiegel, M., Meston, C., & Rosen, R. (2005). The Female Sexual Function Index (FSFI): Cross-validation and development of clinical cutoff scores. *Journal of Sex & Marital Therapy, 31*, 1–20. doi:[10.1080/00926230590475206](https://doi.org/10.1080/00926230590475206).