Journal of Sex & Marital Therapy, 31:1–20, 2005

ISSN: 0092-623X print
DOI: 10.1080/00926230590475206

Brunner-Routledge Taylor & Francis healthsciences

The Female Sexual Function Index (FSFI): Cross-Validation and Development of Clinical Cutoff Scores

MARKUS WIEGEL
Department of Psychiatry, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey, Piscataway, New Jersey, USA
CINDY MESTON
Department of Psychology, University of Texas at Austin, Austin, Texas, USA
RAYMOND ROSEN
Department of Psychiatry, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey, Piscataway, New Jersey, USA

*The Female Sexual Function Index (FSFI) is a brief, multidimensional scale for assessing sexual function in women. The scale has received initial psychometric evaluation, including studies of reliability, convergent validity, and discriminant validity (Meston, 2003; Rosen et al., 2000). The present study was designed to cross-validate the FSFI in several samples of women with mixed sexual dysfunctions (*N = *568) and to develop diagnostic cut-off scores for potential classification of women's sexual dysfunction. Some of these samples were drawn from our previous validation studies (*N = *414), and some were added for purposes of the present study (*N = *154). The combined data set consisted of multiple samples of women with sexual dysfunction diagnoses (*N = *307), including female sexual arousal disorder (FSAD), hypoactive sexual desire disorder (HSDD), female sexual orgasm disorder (FSOD), dyspareunia/vaginismus (pain), and multiple sexual dysfunctions, in*

Supported by an unrestricted grant from Bayer Phamaceuticals.

In this study, women without sexual dysfunction were used as the reference group since higher FSFI scores indicate better sexual function. As a result, "sensitivity" refers to the ability of the test to correctly classify positive sexual function. Sensitivity should be interpreted as the ability of the FSFI to correctly classify women without sexual dysfunction, and specificity to the ability to correctly classify women with sexual dysfunction.

Address correspondence to Markus Wiegel, Department of Psychiatry, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey, Piscataway, NJ 08854, USA. E-mail: wiegel@bu.edu

addition to a large sample of nondysfunctional controls (n = 261*). We conducted analyses on the individual and combined samples, including replicating the original factor structure using principal components analysis with varimax rotation. We assessed Cronbach's alpha (internal reliability) and interdomain correlations and tested discriminant validity by means of a MANOVA (multivariate analysis of variance; dysfunction diagnosis x FSFI domain), with Bonferroni-corrected post hoc comparisons. We developed diagnostic cut off scores by means of standard receiver operating characteristics–curves and the CART (Classification and Regression Trees) procedure. Principal components analysis replicated the original five-factor structure, including desire/arousal, lubrication, orgasm, pain, and satisfaction. We found the internal reliability for the total FSFI and six domain scores to be good to excellent, with Cronbach alpha's >0.9 for the combined sample and above 0.8 for the sexually dysfunctional and nondysfunctional samples, independently. Discriminant validity testing confirmed the ability of both total and domain scores to differentiate between functional and nondysfunctional women. On the basis of sensitivity and specificity analyses and the CART procedure, we found an FSFI total score of 26.55 to be the optimal cut score for differentiating women with and without sexual dysfunction. On the basis of this cut-off, we found 70.7% of women with sexual dysfunction and 88.1% of the sexually functional women in the cross-validation sample to be correctly classified. Addition of the lubrication score in the model resulted in slightly improved specificity (from .707 to .772) at a slight cost of sensitivity (from .881 to .854) for identifying women without sexual dysfunction. We discuss the results in terms of potential strengths and weaknesses of the FSFI, as well in terms of further clinical and research implications.*

In the face of renewed interest in women's sexuality and sexual dysfunction in women, there is a marked need for validated, multidimensional measures of female sexual function. Availability of such measures is an essential requirement for epidemiological, diagnostic, and treatment outcome studies of sexual dysfunction in women (Meston & Derogatis, 2002; Rosen, 2002). Valid and sensitive measures may be of particular value in the clinical management of sexual problems in women, in addition to assessing the efficacy of new treatments, both pharmacological and nonpharmacological, for women's sexual dysfunction. Previous measures have lacked adequate diagnostic precision or discriminant validity testing in a broad range of diagnostic groups (Quirk et al., 2002; Rust & Golombok, 1986; Taylor, Rosen, & Leiblum, 1994). The present study was designed to further assess the discriminant validity

of the Female Sexual Function Inventory (FSFI; Meston, 2003; Rosen et al., 2000) and to develop and validate cutoff scores for diagnostic assessment.

FEMALE SEXUAL FUNCTION INDEX

The FSFI (Rosen et al., 2000) was developed as a brief, multidimensional self-report instrument for assessing key dimensions of sexual function in women. The scale consists of 19 items that assess sexual function over the past 4 weeks and yield domain scores in six areas: sexual desire, arousal, lubrication, orgasm, satisfaction, and pain. The measure was validated on an initial sample of women with female sexual arousal disorder (FSAD) and a control sample of women without sexual difficulties (Rosen et al., 2000). In a second validation study, the FSFI was shown to discriminate between women without sexual dysfunction and women who met DSM-IV-TR criteria for female sexual orgasmic disorder (FSOD) or hypoactive sexual desire disorder (HSDD; Meston, 2003). Significant discriminant validity was shown in all domains of sexual function, as well as in the total FSFI score between sexually dysfunctional and nondysfunctional samples in both studies.

In our first study (Rosen et al., 2000), we reported results of a principal components analysis that used varimax rotation in women with FSAD and controls. This analysis supported a five-factor solution, including a combined desire/arousal, lubrication, orgasm, satisfaction, and sexual pain factors (Rosen et al., 2000). On the basis of these clinical considerations, we separated the desire and arousal factors into two subscales. The resulting six subscales of the FSFI were shown to have excellent internal reliability (Cronbach's alphas $> .9$ for all subscales) and good test-retest reliability (test-retest reliability scores ranged from .79 to .88). Both Rosen et al. (2000) and Meston (2003) found the FSFI to have adequate divergent validity when compared to the Locke-Wallace, a test of marital adjustment (Locke & Wallace, 1959). Meston (2003) reported significant differences between age-matched, sexually functional women ($n = 71$), compared to women with orgasmic disorders ($n = 71$) and women with HSDD ($n = 44$).

The present study was designed to further investigate the psychometric properties of the FSFI. In particular, we aimed to extend the findings on discriminant validity through inclusion of additional patient groups and diagnostic categories and to develop clinically relevant cut-off scores using the Classification and Regression Trees (CART) methodology. This well-validated approach combines scores from several data sets to determine optimal cut-off points for classification of diagnostic categories (Breiman, Friedman, Olshen, & Stone, 1984). To maximize the generalizability of the resulting cut-off scores, we based our analysis on a large heterogeneous group of women representing a wide range of sexual dysfunction diagnoses; we also included nondysfunctional controls. For this purpose, we combined five different data

sets from three investigator groups. These data sets included the validation samples from the Rosen et al. (2000) study, which included 128 women diagnosed with FSAD and 131 sexually functional women, and the samples from the Meston (2003) study, which included 71 sexually functional women, 71 women diagnosed with FSOD, and 44 women with HSDD, although many of the women with sexual dysfunction in this latter study qualified for more than one sexual dysfunction diagnosis (Meston, 2003). Data from an additional 123 subjects, including women with multiple sexual dysfunction and sexual pain disorders were added to these previous data sets to provide a total sample of 568 women. This combined sample is sufficiently large and heterogenous for conducting a valid CART analysis.

Combining datasets from multiple studies embodies potential risks. For example, it is possible that selected samples may differ on essential underlying variables such as age or ethnicity. Additionally, repeating analyses on the same sample or using the same statistical procedures may convey a spurious sense of reliability or robustness of the data. There also are ethical restrictions against republication of previously reported data. We took several steps to guard against these risks in the present study. Samples for each of the studies were drawn from both clinical and nonclinical sources and included a diverse group of age ranges and sexual dysfunction diagnoses. Additionally, we compared psychometric data results from each of the original validation samples with results obtained from the new and combined samples. For example, the internal reliability and factor scores for each of the previously published samples and the new sample were calculated separately. Finally, FSFI cut-off scores have not previously been reported for any of the samples studied.

METHOD

Participants

Participants included 568 women obtained from several data sets. Data sets were included if they contained a majority of cases for which sexual dysfunction status, age, ethnic background, and menopausal status were known. Of these 568 women, 54.0% ($n = 307$) met criteria for a DSM-IV-TR diagnosis of sexual dysfunction, compared to 46.0% ($n = 261$) who did not meet criteria (see Table 1). We based sexual dysfunction diagnoses on interview assessment by a qualified sex therapist in each of the previous studies (Meston, 2003; Rosen et al., 2000). All of the original studies used clinical interviews in order to diagnose sexual dysfunction based on DSM-IV or DSM-IV-TR criteria.

The mean age of the combined sample was 36.2 years (±13.2) and ranged from 18 years to 74 years. The ethnic make-up of the sample was 73.3% (411) Caucasian, 8.7% ($n = 49$) African American, 7.8% ($n = 44$) Hispanic/ Hispanic American, 1.8% ($n = 10$) Asian/Asian American, 0.5% ($n = 3$) Native American, and 7.8% ($n = 44$) other/unknown. The majority

TABLE 1. Weighted Means (SD) of Domain and Individual Item Scores by Sexual Dysfunction Diagnosis (Combined Data)

Item	HSDD ($n = 14$)	FSAD ($n = 152$)	FSOD ($n = 27$)	Pain ($n = 31$)	Multiple Dx ($n = 60$)	Control ($n = 244$)
FSFI desire subscore	3.04 (0.90)	2.99 (1.33)	3.89 (1.27)	3.09 (1.23)	3.10 (1.08)	4.28 (1.12)
FSFI1	2.36 (0.74)	2.50 (1.16)	3.26 (1.23)	2.55 (1.12)	2.47 (0.93)	3.54(1.04)
FSFI2	2.71 (0.91)	2.48 (1.14)	3.22 (1.01)	2.60 (1.02)	2.70 (1.05)	3.59 (0.95)
FSFI arousal subscore	3.85 (1.46)	3.09 (1.46)	4.29 (1.16)	3.30 (1.91)	3.31 (1.14)	5.08 (1.11)
FSFI3	3.36 (1.22)	2.74 (1.46)	3.88 (1.11)	2.94 (1.73)	3.05 (1.12)	4.42 (1.04)
FSFI4	3.15 (1.41)	2.64 (1.26)	3.59 (0.93)	2.74 (1.61)	2.83 (0.91)	4.04 (1.01)
FSFI5	2.86 (1.29)	2.62 (1.28)	3.30 (1.14)	2.47 (1.50)	2.67 (1.13)	4.12 (1.08)
FSFI6	3.29 (1.20)	2.28 (1.33)	3.52 (1.19)	2.84 (1.77)	2.57 (1.18)	4.37 (1.03)
FSFI lubrication subscore	4.67 (1.55)	3.31 (1.60)	5.02 (1.20)	3.67 (2.07)	3.52 (1.66)	5.45 (1.14)
FSFI7	3.71 (1.59)	2.64 (1.43)	4.11 (1.25)	3.08 (1.72)	2.73 (1.54)	4.58 (1.00)
FSFI8	3.93 (1.27)	2.83 (1.43)	4.19 (1.14)	3.21 (1.80)	3.22 (1.39)	4.59 (0.97)
FSFI9	3.93 (1.38)	2.55 (1.49)	4.11 (1.09)	2.74 (1.79)	2.73 (1.59)	4.41 (1.08)
FSFI10	4.00 (1.30)	3.03 (1.52)	4.33 (0.96)	3.21 (1.87)	3.05 (1.57)	4.58 (0.98)
FSFI orgasm subscore	4.40 (1.44)	3.05 (1.66)	2.98 (1.48)	3.04 (2.11)	2.79 (1.31)	5.05 (1.30)
FSFI11	3.64 (1.34)	2.53 (1.54)	2.30 (1.35)	2.44 (1.83)	2.42 (1.34)	4.13 (1.20)
FSFI12	3.93 (1.21)	2.72 (1.50)	2.56 (1.40)	2.65 (1.85)	2.45 (1.33)	4.27 (1.12)
FSFI13	3.43 (1.22)	2.38 (1.45)	2.59 (1.28)	2.52 (1.82)	2.10 (1.17)	4.25 (1.19)
FSFI satisfaction subscore	3.63 (1.34)	3.38 (1.45)	4.16 (1.37)	3.81 (1.60)	3.83 (1.26)	5.04 (1.19)
FSFI14	3.64 (1.34)	3.40 (1.57)	3.70 (1.35)	3.15 (1.85)	3.68 (1.350)	4.23 (1.20)
FSFI15	3.00 (1.18)	2.67 (1.39)	3.41 (1.15)	3.09 (1.60)	3.20 (1.27)	4.17 (1.09)
FSFI16	2.43 (1.22)	2.39 (1.24)	3.30 (1.32)	3.12 (1.42)	2.68 (1.27)	4.12 (1.10)
FSFI pain subscore	4.49 (2.00)	4.21 (1.80)	5.45 (1.05)	2.02 (1.89)	4.99 (1.50)	5.51 (1.29)
FSFI17	3.79 (1.81)	3.37 (1.66)	4.52 (0.98)	1.53 (1.59)	4.12 (1.33)	4.55 (1.17)
FSFI18	3.64 (1.69)	3.65 (1.66)	4.59 (0.93)	1.60 (1.69)	4.28 (1.30)	4.62 (1.10)
FSFI19	3.79 (1.58)	3.49 (1.49)	4.52 (0.85)	1.92 (1.62)	4.08 (1.28)	4.62 (1.06)
FSFI total score	23.89 (6.47)	20.05 (6.74)	25.80 (4.51)	19.73 (8.73)	21.59 (5.48)	30.75 (4.80)

FSFI: Female sexual function index; HSDD: Hypoactive sexual desire disorder; FSAD: Female sexual arousal disorder; FSOD: Female sexual orgasm disorder.

of women were premenopausal (65.6%, $n = 368$), 3.6% ($n = 20$) were peri- or postmenopausal or had had an ovariectomy (surgically menopausal), 10.5% ($n = 59$) were on hormone replacement therapy (HRT) with uncertain menopausal status, and 20.3% ($n = 114$) were coded as uncertain menopausal status without current HRT. Forty-eight percent ($n = 267$) of the sample was nulliparious, whereas 35.1% ($n = 197$) had at least one child; for 97 women (17.3%) this variable was unknown. For about 17% of the sample, educational level, income, and marital status were unknown. For the remaining sample ($n = 466$), 44.2% ($n = 206$) were single, 43.1% ($n = 201$) were married, and 12.7% ($n = 59$) were divorced. Almost half (48.6%) of the sample had at least a four: year college degree.

Procedures and Analyses

Prior to combining the data sets, we removed all identifying information for the participants and assigned each case a unique research identification number to be used only for this analysis. Because not all of the studies used the same coding for all of the variables, we recoded some of the demographic variables, such as ethnic background, menopause status, education, income, and marital status, to ensure a common scale for all cases.

The overall goal of the study was to establish clinical cut-off scores for the FSFI. We used combination of receiving operating characteristic–curve fitting (ROC) and CART methodology to this end. However, prior to determining clinical cut-off scores, we needed to further assess the psychometric properties of the FSFI. First, we evaluated the factor structure using principal components analyses with varimax rotation. Next we calculated the internal reliability (Cronbach's alpha) and interdomain correlations for the six FSFI domains. An important component of the study was evaluating the discriminant validity of the FSFI in this large heterogenous sample, and we accomplished this in two stages. First, we compared women with and without sexual dysfunction. Next, we compared FSFI scores for specific diagnostic categories with one other and compared scores of each diagnostic group with scores of the nondysfunctional controls. Some of these tests were conducted to replicate findings from the Rosen et al. (2000) and Meston (2003) studies.

RESULTS

Factor Analyses

In the first validation study (Rosen et al., 2000), we found a five-factor solution, including individual factors for desire/arousal, lubrication, orgasm, satisfaction, and sexual pain. To replicate this finding, we performed a principal components analysis with varimax rotation on a combined group of sexually functional and dysfunctional women in the present sample ($n = 272$), from

which all cases in the Rosen et al. (2000) data set were specifically excluded. The five-factor solution after varimax rotation was almost identical, including desire/arousal, lubrication, orgasm, pain, and satisfaction, thus replicating the factor structure in an independent sample of women with and without sexual dysfunctions. Given the robust and highly reliable factor structure observed across studies, we conducted the remaining principal components analyses with the different data sets combined. Separate principal components analyses were conducted for women with and without sexual dysfunction and for the two groups combined ($n = 527$). In the women with sexual dysfunction, the initial solution resulted in four factors with eigenvalues >1.0, and a fifth factor with eigenvalue =.98. On the basis of the four-factor solution, after varimax rotation we could identify a distinct pain factor (items 17, 18, 19), orgasm factor (items 11, 12, 13), and lubrication factor (items 7, 8, 9, 10). The fourth factor was a mixture of desire, arousal, and satisfaction items. We conducted a principal components analysis based on a 5-factor solution, including the factor with eigenvalue =.98. The resulting factor structure and item loadings are shown in Table 2. The five factors correspond to desire/arousal, lubrication, orgasm, pain, and satisfaction.

In the women without sexual dysfunction, the principal components solution yielded five factors with eigenvalues >1.0. After varimax rotation, the identifiable factors corresponded to desire/arousal (items 1, 2, 3, 4), orgasm/arousal (items 5, 6, 11, 12, 13), lubrication (items 7, 8, 9, 10), satisfaction (items 14, 15, 16), and pain (items 17, 18, 19). The resulting factor structure and item loadings are shown in Table 3. Factor analysis of the combined group (dysfunctional and nondysfunctional) yielded as essentially identical factor structure as that for the nondysfunctional group. As in Rosen et al. (2000), although the principal components analyses supported a five-factor solution, we separated the desire/arousal factor into two domains based on clinical considerations.

Interdomain Correlations

Interdomain correlations provide additional information about the associations among domains. To replicate the interdomain associations observed in the first validation study, we calculated interdomain correlations for women in our sample who had not been included in the original validation study ($n = 309$). These results are shown in Table 4 below. The Pearson correlation coefficients observed were in the same direction and of similar magnitude (range $r = .39$ to .78) as those in the original study (range $r = .37$ to .76).

We then calculated interdomain correlations separately for the samples of women with and without sexual dysfunction (all data sets combined, $n = 568$). All correlations were significant and ranged from $r = .16$ to $r = .76$ (See Table 4). In general, interdomain correlations were lower for women with sexual dysfunction than for the nondysfunctional controls. Consistent

TABLE 2. Principal Components Analyses: 5 Factor Solution with Varimax Rotation for Women with Sexual Dysfunction

Item	Factor				
	Desire/arousal	Lubrication	Orgasm	Pain	Satisfaction
1 How often did you feel sexual desire or interest?	.867	.117	.100	.094	.116
2 How would you rate your level of sexual desire or interest?	.875	.148	.106	.059	.189
3 How often did you feel sexually aroused during sexual activity?	.700	.273	.296	.072	.376
4 How would you rate your level of sexual arousal during sexual activity?	.716	.258	.328	.069	.378
5 How confident were you about becoming sexually aroused during sexual activity?	.727	.193	.329	.154	.231
6 How often have you been satisfied with your arousal during sexual activity?	.574	.287	.493	.074	.362
7 How often did you become lubricated during sexual activity?	.314	.844	.130	.095	.140
8 How difficult was it to become lubricated during sexual activity?	.251	.855	.164	.148	.055
9 How often did you maintain your lubrication until completion of sexual activity?	.145	.865	.153	.183	.148
10 How difficult was it to maintain your lubrication until completion of sexual activity?	.059	.898	.173	.228	.070
11 When you had sexual stimulation or intercourse, how often did you reach orgasm?	.225	.114	.881	.024	.114
12 When you had sexual stimulation or intercourse, how difficult was it for you to reach orgasm?	.220	.188	.882	.091	.063
13 How satisfied have you been with your ability to reach orgasm during sexual activity?	.200	.212	.827	.004	.279
14 How satisfied have you been with the amount of emotional closeness during sexual activity?	.160	.069	.095	.103	.801
15 How satisfied have you been with your sexual relationship with your partner?	.344	.122	.180	.147	.788
16 How satisfied have you been with your overall sex life?	.419	.153	.217	.114	.692
17 How often did you experience discomfort or pain during vaginal penetration?	.142	.202	.054	.912	.098
18 How often did you experience discomfort or pain following vaginal penetration?	.056	.153	.001	.937	.094
19 How would you rate the level of discomfort or pain during or following vaginal penetration?	.088	.174	.073	.933	.122
Eigenvalue	8.85	2.68	1.87	0.98	1.43

TABLE 3. Principal Components Analyses: 5 Factor Solution with Varimax Rotation for Women Without Sexual Dysfunction

Item	Factors				
	Arousal/ orgasm	Lubrication	Desire/ arousal	Pain	Satisfaction
1 How often did you feel sexual desire or interest?	.061	.127	.844	.125	.141
2 How would you rate your level of sexual desire or interest?	.105	.078	.862	.097	.244
3 How often did you feel sexually aroused during sexual activity?	.431	.340	.528	.297	.158
4 How would you rate your level of sexual arousal during sexual activity?	.411	.271	.635	.158	.249
5 How confident were you about becoming sexually aroused during sexual activity?	.533	.302	.511	.128	.153
6 How often have you been satisfied with your arousal during sexual activity?	.706	.127	.363	.070	.319
7 How often did you become lubricated during sexual activity?	.234	.863	.215	.180	.080
8 How difficult was it to become lubricated during sexual activity?	.242	.818	.174	.232	.135
9 How often did you maintain your lubrication until completion of sexual activity?	.114	.872	.137	.109	.210
10 How difficult was it to maintain your lubrication until completion of sexual activity?	.202	.860	.077	.237	.151
11 When you had sexual stimulation or intercourse, how often did you reach orgasm?	.847	.211	.116	.101	.171
12 When you had sexual stimulation or intercourse, how difficult was it for you to reach orgasm?	.795	.325	.073	.188	.166
13 How satisfied have you been with your ability to reach orgasm during sexual activity?	.820	.089	.087	.112	.286
14 How satisfied have you been with the amount of emotional closeness during sexual activity?	.242	.186	.160	.083	.787
15 How satisfied have you been with your sexual relationship with your partner?	.259	.130	.238	.108	.858
16 How satisfied have you been with your overall sex life?	.302	.205	.262	.106	.796
17 How often did you experience discomfort or pain during vaginal penetration?	.131	.199	.189	.889	.083
18 How often did you experience discomfort or pain following vaginal penetration?	.126	.167	.107	.917	.106
19 How would you rate the level of discomfort or pain during or following vaginal penetration?	.135	.239	.105	.906	.078
Eigenvalue	9.06	2.26	1.64	1.11	1.41

TABLE 4. Interdomain Correlations for Female Sexual Function Index (FSFI) Total Score and Domain Scores for Controls and Combined Sexual Dysfunctions

	Rosen et al. (2000) ($n = 259$)					
Domain	Desire	Arousal	Lubrication	Orgasm	Satisfaction	Pain
Desire	1.00					
Arousal	0.76	1.00				
Lubrication	0.56	0.75	1.00			
Orgasm	0.54	0.81	0.68	1.00		
Satisfaction	0.60	0.80	0.62	0.70	1.00	
Pain	0.37	0.47	0.64	0.41	0.53	1.00
	Cases excluding Rosen et al. (2000) ($n = 309$)					
Desire	1.00					
Arousal	0.78	1.00				
Lubrication	0.51	0.67	1.00			
Orgasm	0.55	0.76	0.57	1.00		
Satisfaction	0.63	0.69	0.43	0.52	1.00	
Pain	0.39	0.51	0.47	0.39	0.30	1.00
	Sexually functional women (Combined data set) ($n = 261$)					
Desire	1.00					
Arousal	0.65	1.00				
Lubrication	0.44	0.69	1.00			
Orgasm	0.41	0.76	0.61	1.00		
Satisfaction	0.45	0.63	0.44	0.56	1.00	
Pain	0.40	0.57	0.60	0.48	0.30	1.00
	Women with sexual dysfunction (Combined data set) ($n = 307$)					
Desire	1.00					
Arousal	0.75	1.00				
Lubrication	0.38	0.55	1.00			
Orgasm	0.39	0.65	0.42	1.00		
Satisfaction	0.54	0.69	0.35	0.44	1.00	
Pain	0.21	0.29	0.40	0.16	0.29	1.00

with the finding of a combined desire/arousal factor, the highest interdomain correlation for women with sexual dysfunction was between desire and arousal ($r = .75$). The highest correlations for the sexually functional women were between arousal and orgasm ($r = .76$), arousal and lubrication ($r = .69$), and arousal and desire ($r = .65$). Of note is the high correlation between lubrication and pain ($r = .60$) in sexually functional women, which is in contrast to a lower correlation for the dysfunctional women ($r = .40$). This finding suggests that for women without dyspareunia or vaginismus, pain during sexual activity is probably related to lack of lubrication (e.g., after prolonged penile-vaginal intercourse).

Internal Consistency

Internal consistency is a measure of the relatedness of items within each factor. To replicate and extend findings from the previous two validation studies, we calculated internal consistency (Cronbach's alpha) for the FSFI

TABLE 5. Cronbach's Alphas for Female Sexual Function Index (FSFI)-Total Score and Domain Scores

Item	Rosen et al. 2000 ($n = 255$)	Meston 2003 ($n = 138$)	New Cases ($n = 134$)	Dysfunction ($n = 283$)	Controls ($n = 244$)
FSFI total score	$\alpha = .97$	$\alpha = .94$	$\alpha = .95$	$\alpha = .93$	$\alpha = .94$
Desire	$\alpha = .92$	$\alpha = .86$	$\alpha = .91$	$\alpha = .89$	$\alpha = .88$
Arousal	$\alpha = .95$	$\alpha = .93$	$\alpha = .96$	$\alpha = .94$	$\alpha = .91$
Lubrication	$\alpha = .96$	$\alpha = .94$	$\alpha = .97$	$\alpha = .94$	$\alpha = .96$
Orgasm	$\alpha = .94$	$\alpha = .94$	$\alpha = .93$	$\alpha = .91$	$\alpha = .91$
Satisfaction	$\alpha = .89$	$\alpha = .85$	$\alpha = .84$	$\alpha = .82$	$\alpha = .89$
Pain	$\alpha = .94$	$\alpha = .94$	$\alpha = .98$	$\alpha = .95$	$\alpha = .96$

total score and for each of the domain subscales separately for a sample of women not included in any previous study ($n = 134$), as well as for samples in the original Rosen et al. (2000) and Meston (2003) studies. The resulting alphas were highly similar (see Table 5). Accordingly, we combined the data sets and calculated Cronbach alphas separately for the groups of women with ($n = 283$) and without ($n = 244$) sexual dysfunction (Table 5.).

Discriminant Validity

Discriminant validity concerns the ability of a test to differentiate between cases and controls. In the present study, we conducted separate multivariate analysis of variance (MANOVA) tests to compare women with and without sexual dysfunction diagnoses in each of the two earlier studies (Meston, 2003; Rosen et al., 2000), as well as in the sample not included in the previous studies. The multivariate tests (Hotelling's Trace) were significant for all three samples: Rosen et al., 2000 ($F(6, 248) = 40.41$, $p < .001$, $\text{ETA}^2 = .49$); Meston, 2003 ($F(6, 131) = 17.50$, $p < .001$, $\text{ETA}^2 = .45$); and additional cases ($F(6, 127) = 11.63$, $p < .001$, $\text{ETA}^2 = .36$). In all three samples, further univariate tests on the total score and individual scale scores (arousal, lubrication, orgasm, satisfaction, pain) were all significant at $p < .001$. Thus, the FSFI total score and each of the domain scores discriminated significantly between women with and without sexual dysfunction. Because similar results were obtained in all three samples separately, these were combined for subsequent analyses by specific sexual dysfunction categories.

We conducted discriminant validity testing for the sexual dysfunction diagnoses by means of MANOVA (dysfunction diagnosis x FSFI domain) with Bonferroni corrected post hoc comparisons. The overall MANOVA was significant using Wilks' Lambda, $F(30, 2038) = 17.81$, $p < .001$, and $\text{ETA}^2 = .17$. Follow-up univariate analyses with between-subjects correction indicated that the six diagnostic groups (HSDD, FSAD, FSOD, sexual pain disorder, multiple sexual dysfunction, nondysfunctional controls) were significantly different on the total score ($F(5, 514) = 78.06$, $p < .001$, $\text{ETA}^2 = .43$), desire ($F(5, 514) = 29.70$, $p < .001$, $\text{ETA}^2 = .22$), arousal ($F(5, 514) = 62.24$,

$p < .001$, $ETA^2 = .38$), lubrication ($F(5, 514) = 58.53$, $p < .001$, $ETA^2 = .36$), orgasm ($F(5, 514) = 55.94$, $p < .001$, $ETA^2 = .35$), satisfaction ($F(5, 514) = 32.37$, $p < .001$, $ETA^2 = .24$), and pain domains ($F(5, 514) = 41.90$, $p < .001$, $ETA^2 = .29$).

We conducted Bonferroni corrected, pair-wise post-hoc comparison tests on the individual means. For the total score, the nondysfunctional control group differed significantly from all other groups (all $p < .01$). The HSDD group differed only from the control group. The FSAD group differed from both the control and FSOD groups. The FSOD group differed from all other groups, except the HSDD group. The sexual pain group differed from both the FSOD and control groups. The multiple dysfunction group differed from the control and FSOD groups.

The pattern of results for the total score and each domain score across diagnostic groups is shown in Table 6. The Desire domain score showed the weakest ability to discriminate between the groups of women with and without sexual dysfunction. In contrast, the Lubrication score differentiated between the FSAD and HSDD, FSAD and FSOD, and between the FSOD and control groups but did not differentiate the pain or multiple dysfunction groups. However, the Lubrication score did differentiate the FSOD group from the FSAD, pain, and multiple dysfunction groups. The Orgasm domain score was generally a poor discriminator between the various sexual dysfunction diagnostic groups. Similarly, scores on the Satisfaction domain were significantly different between nondysfunctional controls and each of the sexual dysfunction groups but not between the individual dysfunctions. In contrast, Sexual Pain scores discriminated strongly between the pain/vaginismus group and all other diagnostic groups and also between the FSAD group and most groups, not including the HSDD group.

In summary, strong evidence of discriminant validity was observed between the total score and individual domain scores between sexually dysfunctional and nondysfunctional women overall. However, a high degree of overlap was seen in the pattern of domain scores across the various diagnostic groups in the study.

Receiver Operating Characteristic Curves

As an initial step towards developing clinical cut-off scores for the FSFI, we created receiver operating characteristic (ROC) curves using the combined sexual dysfunction and nondysfunctional samples. ROC curves show the pairing of true-positive and false-positive coordinates across a range of cut-off points that distinguish positive cases from controls. For the ROC curve for the FSFI, the true-positive rate (sensitivity) represents the proportion of sexually functional women that were correctly classified by the FSFI total score or domain scores as functional. The false-positive rate (1–specificity)

TABLE 6. Female Sexual Function Index Discriminant Analyses Results: Mean Differences for Diagnostic Groups

	Control	HSDD	FSAD	FSOD	Pain
Total score (mean difference)					
HSDD	6.86*				
FSAD	10.71*	3.84			
FSOD	4.96*	−1.90	−5.75*		
Pain	11.64*	4.16	0.32	6.06*	
Multiple	9.17*	2.30	−1.54	4.21	−1.86
Desire score (mean difference)					
HSDD	1.23*				
FSAD	1.33*	.09			
FSOD	0.45	−.79	−.88*		
Pain	1.19*	.04	−.13	.74	
Multiple	1.23*	−.00	−.09	.78	−.04
Arousal score (mean difference)					
HSDD	1.30*				
FSAD	2.05*	.76			
FSOD	.86	−.44	−1.19*		
Pain	1.63*	.33	−.42	.77	
Multiple	1.85*	.55	−.21	.99	.22
Lubrication score (mean difference)					
HSDD	.88				
FSAD	2.21*	1.33*			
FSOD	.43	−.36	−1.68*		
Pain	1.60*	.71	−.61	1.07	
Multiple	1.96*	1.08	−.24	1.44*	.37
Orgasm score (mean difference)					
HSDD	.84				
FSAD	2.06*	1.21			
FSOD	2.11*	1.26	.04		
Pain	1.87*	1.03	−.18	−.23	
Multiple	2.36*	1.51*	.30	.25	.48
Satisfaction score (mean difference)					
HSDD	1.43*				
FSAD	1.64*	.21			
FSOD	.85	−.59	−.79		
Pain	1.22*	−.21	−.42	.37	
Multiple	1.21*	−.22	−.43	.36	.01
Pain score (mean difference)					
HSDD	1.17				
FSAD	1.42*	.25			
FSOD	.17	−1.00	−1.25*		
Pain	3.51*	2.34*	2.10*	3.34*	
Multiple	.56	−.61	−.86*	.39	−2.95*

*Significantly different at $p < .001$, .01, or .05 level.
HSDD: Hypoactive sexual desire disorder; FSAD: Female sexual arousal disorder; FSOD; Female sexual arousal disorder.

represents the proportion of women with sexual dysfunction that the test incorrectly classifies as having sexual health. We ascertained the ROC curve by comparing a participant's known diagnosis with the prediction based on the potential cut-off scores for the total score and each of the individual

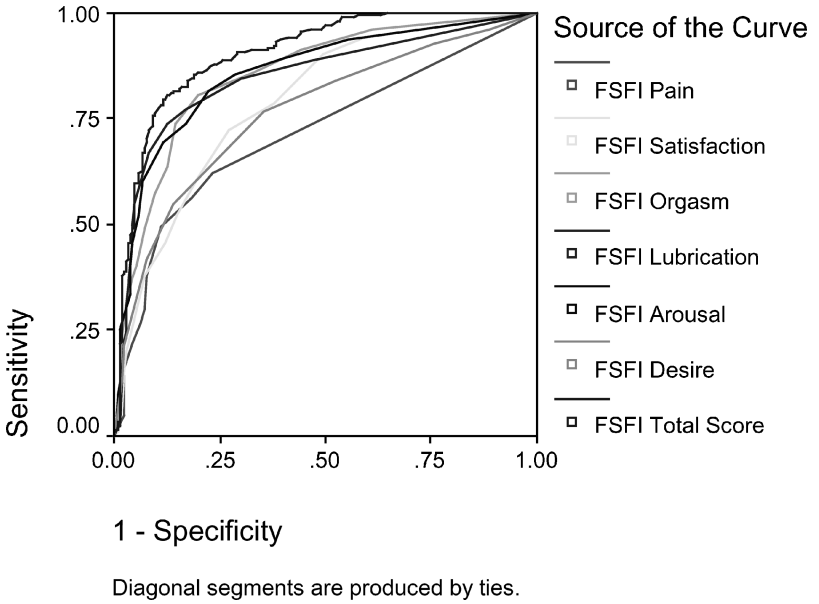

FIGURE 1. ROC curve for sexual dysfunction (Y/N).

domain scores. Figure 1 shows the ROC curves for the total score and individual domain scores. As shown, the sensitivity and specificity curves are highest for the total score, followed closely by the ROC curves for lubrication, arousal, and orgasm. The curves for desire, pain, and satisfaction were markedly lower (See Figure 1). Overall, the FSFI total score showed the best sensitivity to 1–specificity profile, as reflected by the largest area under the curve (area = .899, $p < .001$).

CART Analyses

We developed cut off scores for the FSFI using a combination of ROC-curves and CART methodology. CART is a validated statistical procedure in which classification and regression trees are constructed for predicting continuous dependent variables (regression) and categorical predictor variables (classification). In broad terms, the CART analysis is used to develop a classification algorithm based on a series of testable if-then logical (split) conditions that permits accurate prediction or classification of individual cases. The CART algorithm uses different stopping rules to determine the optimal decision tree (i.e., number of splits). The CART algorithm method was first described by Breiman, Friedman, Olshen, and Stone (1984) and has been used recently for development of clinical cut off scores in men with erectile dysfunction (Cappelleri, Siegel, Glasser, Osterloh, & Rosen, 2001; Rosen, Cappelleri, & Gendrano, 2002).

Using CART, we developed three different decision tree models in the present study. The binary variable of sexual dysfunction status (Yes/No) was defined as the dependent variable in all three models. As recommended by Breiman et al. (1984), each potential model was created on 90% (learn sample) of cases, then cross-validated on 10 subsequent subsamples consisting each of 10% of the total sample (test samples).

The first model included only one predictor variable, FSFI total score, and was used to determine the optimal clinical cut-off score based on this variable for the presence or absence of sexual dysfunction. We evaluated each potential cut-off point on the basis of a goodness-of-split criterion, as measured by the Gini diversity (impurity) index. The optimal cut-off score resulted in the lowest Gini value and the least variability in the actual status of women categorized by the CART algorithm. Because the sample included approximately equal numbers of women with and without sexual dysfunction, we assigned no a priori weights (i.e., the initial probability for the two groups was set as equal). The results indicated that a cut-off score of 26.55 on the FSFI total score resulted in the lowest rate of misclassification. On the basis of a cut-off score of 26.55 or less, we determined that 70.7% ($n = 217$) of cases in the cross validation sample were correctly classified as dysfunctional and 88.1% ($n = 230$) were correctly classified as nondysfunctional (see Table 7). This resulted in a specificity of .733 and sensitivity of .889 for the learning sample and a specificity of .707 and sensitivity of .881 for the cross-validation sample. These are within acceptable limits for diagnostic classification.

The second model again used the binary variable sexual dysfunction (Y/N) as the dependent variable; however, we included the six domain scores as predictor variables in the CART analyses. This analysis was run primarily to assess whether inclusion of the individual domain scores would result in more-accurate classification of cases. The FSFI total score was specifically excluded from the model in this analysis. The resulting CART algorithm

TABLE 7. Classification Tables for Classification and Regression Trees Models

	Learning sample		Cross validation sample			
	Correct (%)	Misclassified (%)	Correct (%)	Misclassified (%)	Specificity	Sensitivity
Model 1					.707	.881
Non-FSD*	232 (88.9)	29 (11.1)	230 (88.1)	31 (11.9)		
FSD	225 (73.3)	82 (26.7)	217 (70.7)	90 (29.3)		
Model 2					.775	.797
Non-FSD*	230 (88.1)	31 (11.9)	208 (79.7)	53 (20.3)		
FSD	247 (80.5)	60 (19.5)	238 (77.5)	69 (22.5)		
Model 3					.772	.854
Non-FSD*	224 (85.8)	37 (14.2)	223 (85.4)	38 (14.6)		
FSD	248 (80.8)	59 (19.2)	237 (77.2)	70 (22.8)		

*Used as the reference group.
FSD: Female sexual dysfunction.

showed a complicated decision tree, with 8 different end categories. Consistent with the ROC findings, the orgasm, arousal, and lubrication domains had the most-predictive value in the model. The decision tree first split the sample into those with lubrication scores above and those below 4.95. The subgroup with higher lubrication scores was further subdivided into those with orgasm domain scores above and those below 4.6. The subgroup with higher orgasm scores was one of the final categories assigned, including 192 of the 261 nondysfunctional cases. The complete algorithm, including all decision criteria, is shown in Figure 2. When we applied this model to the cross-validation sample, 77.5% ($n = 238$) of the women with sexual dysfunction were classified correctly, and 79.7% ($n = 208$) of the nondysfunctional controls were classified correctly. This resulted in a specificity of .805 and a sensitivity of .881, based on the learning sample, and a specificity of .775 and sensitivity of .797, based on the cross-validation sample. Of note, these findings are consistent with results of the discriminant validity analyses presented above, in which the arousal domain ($ETA^2 = .38$), lubrication domain ($ETA^2 = .36$), and orgasm domain ($ETA^2 = .35$) discriminated between women with and without a diagnosis of sexual dysfunction. Additionally, although the domain scores used in the model for splitting the sample should not be used in isolation as clinical cut-off scores, the decision tree represented by the model (see Figure 2) may suggest strategies to aid in differential diagnoses. For example, those individuals with low lubrication but adequate orgasm scores were further divided into those with or without sexual pain and inadequate lubrication (lubrication domain score ≤ 4.35).

A third model incorporated both the FSFI total score and six domain scores as predictor variables and the presence or absence of sexual dysfunction (Y/N) as the dependent variable. The resulting model consisted of two splits or decision points. The first decision point was based on a total score of 26.55, with scores ≤ 26.55 being classified as having sexual dysfunction. Subjects with total scores > 26.55 were split further, based on a lubrication score of 4.35 or less. On the basis this model, we determined that 77.2% ($n = 237$) of the women with sexual dysfunction were correctly classified, and 85.4% ($n = 223$) of the women without sexual dysfunction were correctly classified. Specificity and sensitivity were .81 and .86, respectively, for the learning sample and .77 and .85 for the cross-validation sample. Thus, inclusion of the lubrication domain score in the model resulted in a slight increase in specificity but a decrease in sensitivity of the diagnostic classification.

DISCUSSION

Self-report questionnaires have become the mainstay of sexual function assessment in both men and women. Despite widespread use of these measures, few questionnaires have been cross-validated in multiple samples of

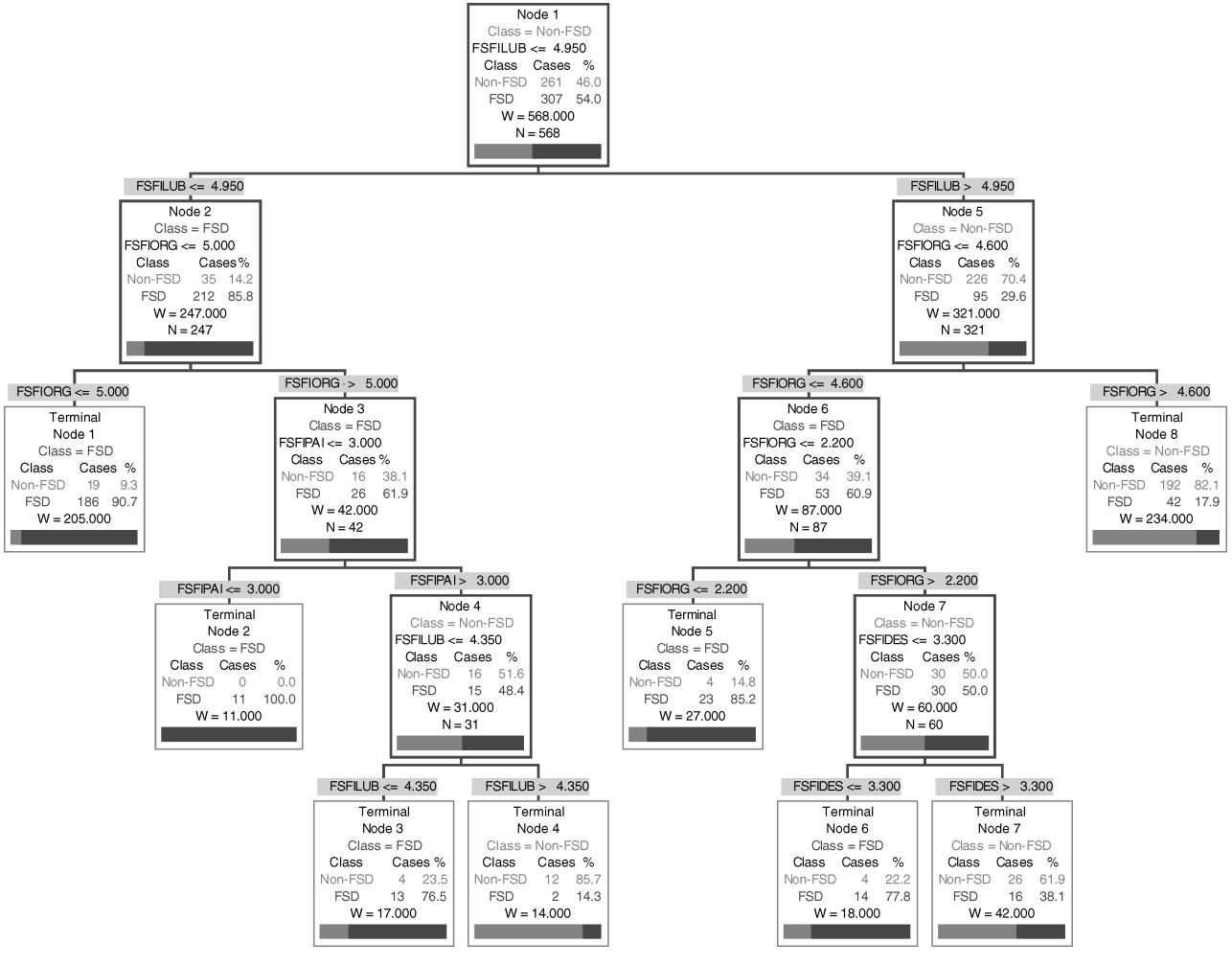

FIGURE 2. Classification and Regression Trees decision tree for model 2.

sexually functional and dysfunctional women. Furthermore, cut off scores for diagnostic classification have not been developed or validated for any of the existing measures. For the present study, we combined FSFI data from two previous studies (Meston, 2003; Rosen et al., 2000) with additional data from clinical and nonclinical samples. With this combined data set of more than 500 subjects in multiple diagnostic categories, we were able to replicate our original factor analysis and interdomain correlations, to extend the discriminant validity findings on the measure, and to develop diagnostic cutoff scores based on both ROC and CART analyses. Overall, our results offer strong confirmation of the previous psychometric findings on the test, in addition to providing evidence in support of a cut-off score for diagnostic classification. Specifically, a total scale score of 26 or less is proposed for this purpose.

We conducted several factor analyses on the individual and combined data sets. The results were generally consistent in supporting a five-factor solution, as reported in the original validation study (Rosen et al., 2000). A high degree of overlap between the desire and arousal domains was noted in each of the analyses conducted to date. This overlap is reflected in both the pattern of mixed factor loadings and high interdomain correlations observed in each of the samples. This result can be viewed as evidence for a nonlinear model of sexual response in women, as proposed by Basson (2000, 2002) and others. Future models of sexual response in women should consider development of a new construct that might incorporate elements of both desire and arousal.

Discriminant validity of the FSFI was confirmed in each of the individual samples, in addition to the combined data set. Although the measure differentiated well between the various clinical samples and the nondysfunctional controls, we noted marked overlap in domain and total scores among the clinical samples. Among the different diagnostic groups, the orgasmic dysfunction group showed a significant decrease in total score and orgasm domain scores compared to controls but showed the least impact on other domains of sexual function. In contrast, the sexual pain/vaginismus group showed lower scores on all domains of sexual function compared to the controls. Not surprisingly, this group also showed the lowest pain scores, compared to each of the other diagnostic groups. The lubrication scores were significantly lower in the arousal disorder, sexual pain and multiple sexual dysfunction groups but not in the desire disorder, anorgasmia or nondysfunctional groups. A mixed pattern of domain scores was similarly noted for the other diagnostic groups. Again, these findings can be viewed as evidence for the overlap among sexual function domains in women and the multifactorial nature of sexual dysfunction problems in women.

On the basis of ROC and CART analyses, we propose a total score of $\leq$26 for diagnostic classification. Any individual woman who scores 26 or less on the total score should be considered at risk for sexual dysfunction and

be evaluated further. Additionally, increased pain scores can be viewed as evidence for a sexual pain disorder, and decreased lubrication scores may be associated with a sexual arousal disorder. A substantial association was noted between decreased lubrication and pain in our sample. In contrast, decreased desire and arousal scores were associated with most of the diagnostic groups and should not be used as the basis for diagnostic classification. However, these domains may be sensitive to treatment intervention effects and could be used as endpoints in a clinical trial. Their usefulness in this regard remains to be determined.

From a clinical perspective, these findings support the use of the FSFI as a screening tool and potential aid in diagnostic assessment. As we have previously noted (Rosen et al., 2000), however, it should not be used as the sole basis for diagnostic classification. In addition to the psychometric issues reviewed above, the assessment is limited to the past 4 weeks. No information is provided regarding the onset or duration of the problem, nor does the scale offer information on the role of etiological or maintaining factors. Psychological and relationship factors are also not adequately addressed. Additionally, the scale is not designed to assess potential interactions among domains or to differentiate primary from secondary or situational causes of sexual dysfunction in women.

Overall, these findings support further research and clinical use of the FSFI. Current studies include its use in samples of women with chronic medical illnesses (e.g., cancer, diabetes) and women seeking services for sexual dysfunction. Studies of treatment sensitivity are also underway. These latter studies are necessary to determine the validity of the measure as an endpoint in clinical trials of sexual dysfunction in women. Finally, linguistic validation studies are underway in several countries.

REFERENCES

Basson, R. (2000). The female sexual response: A different model. *Journal of Sex & Marital Therapy, 26*, 51–65.

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

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International.

Cappelleri, J. C., Siegel, R. L., Glasser, D. B., Osterloh, I. H., & Rosen, R. C. (2001). Relationship between patient self-assessment of erectile dysfunction and the sexual health inventory for men. *Clinical Therapeutics, 23*, 1707–1719.

Locke, H. J., & Wallace, K. M. (1959). Short marital adjustment and prediction tests: Their reliability and validity. *Journal of Marriage and the Family, 21*, 251–255.

Meston, C. M. (2003). Validation of the Female Sexual Function Index (FSFI) in women with female orgasmic disorder and in women with hypoactive sexual desire disorder. *Journal of Sex & Marital Therapy, 29*, 39–46.

Meston, C. M., & Derogatis, L. R. (2002). Validated instruments for assessing female sexual function. *Journal of Sex & Marital Therapy, 28*(Suppl. 1), 155–164.

Quirk, F. H., Heiman, J. R., Rosen, R. C., Laan, E., Smith, M. D., & Boolell, M. (2002). Development of a sexual function questionnaire for clinical trials of female sexual dysfunction. *Journal of Womens Health & Gender-Based Medicine, 11*, 277–289.

Rosen, R. C. (2002). Sexual function assessment and the role of vasoactive drugs in female sexual dysfunction. *Archives of Sexual Behavior, 31*, 439–443.

Rosen, R. C., Brown, C., Heiman, J., Leiblum, S., Meston, C. M., Shabsigh, R., Ferguson, D., & D'Agostino, R. (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.

Rosen, R. C., Cappelleri, J. C., & Gendrano, N. I. (2002). The International Index of Erectile Function (IIEF): A state-of-the-science review. *International Journal of Impotence Research, 14*, 226–244.

Rust, J., & Golombok, S. (1986). The GRISS: A psychometric instrument for the assessment of sexual dysfunction. *Archives of Sexual Behavior, 15*, 157–165.

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