

PSYCHOMETRICS

## Psychometric Properties of the Sexual Event Diary in a Sample of Dutch Women With Female Sexual Interest/Arousal Disorder



Yvonne van Nes, MSc,<sup>1</sup> Jos Bloemers, PhD,<sup>1,2</sup> Rob Kessels, MSc,<sup>1</sup> Peter G. M. van der Heijden, PhD,<sup>3,4</sup>  
Kim van Rooij, MD, PhD,<sup>1,2</sup> Jeroen Gerritsen, MSc,<sup>1,2</sup> Leonard DeRogatis, PhD,<sup>5,6</sup> and Adriaan Tuiten, PhD<sup>1</sup>

### ABSTRACT

**Background:** Efficacy of on-demand drugs for women with hypoactive sexual desire disorder or female sexual interest/arousal disorder (FSIAD) should be assessed using a validated instrument that assesses the discrete sexual events during which the on-demand drug is taken, because this type of assessment is more proximate to an on-demand drug's efficacy compared to instruments that assess sexual function over longer periods of time.

**Aim:** The aim of this study was to assess the psychometric properties of the Dutch translation of the previously validated 11-item Sexual Event Diary (SED) for measuring sexual satisfaction and sexual functioning during discrete sexual events.

**Methods:** Psychometric assessment was performed on data of 1,840 SEDs from 139 women with hypoactive sexual desire disorder/FSIAD, collected during a randomized clinical cross-over trial conducted in the Netherlands.

**Outcomes:** Item scores of the SED at the event level, and at subject level, summarized item scores during the placebo run-in period (PRI) and active treatment period, and score changes from PRI to active treatment period.

**Results:** Reliability and convergent validity were confirmed. All item scores showed the ability to discriminate between known groups. Larger mean score changes from PRI were observed in groups with known benefit from the medication, as compared to those with no benefit. Guyatt effect sizes ranged from 0.51–1.02, thereby demonstrating ability to detect change.

**Clinical Translation:** The Dutch version of the SED is an excellent instrument for assessing female sexual functioning and sexual satisfaction during discrete sexual events and for assessing these concepts over longer periods of time.

**Conclusions:** Data were collected in a randomized, well-controlled trial. The large number of data points gave high statistical power, and the results confirmed previous findings. However, care is needed when generalizing the SED's validity to other areas of research, eg, recreational drug use and sexual risky behaviors, since the current validation study has not used such data. Consistent with the US-English version, the Dutch version of the SED is a reliable, valid, and responsive instrument, and suitable for use in evaluating effects of on-demand drugs in women with FSIAD. **van Nes Y, Bloemers J, Kessels R, et al. Psychometric Properties of the Sexual Event Diary in a Sample of Dutch Women With Female Sexual Interest/Arousal Disorder. J Sex Med 2018;15:722–731.**

Copyright © 2018, International Society for Sexual Medicine. Published by Elsevier Inc. All rights reserved.

**Key Words:** Sexual Event Diary; Dutch; Patient-Reported Outcome; Psychometric Properties; Satisfactory Sexual Event; Sexual Function; Female Sexual Interest/Arousal Disorder; Hypoactive Sexual Desire Disorder

Received September 5, 2017. Accepted March 17, 2018.

<sup>1</sup>Emotional Brain BV, Almere, the Netherlands;

<sup>2</sup>Utrecht Institute for Pharmaceutical Sciences and Rudolf Magnus Institute of Neuroscience, Utrecht University, Utrecht, the Netherlands;

<sup>3</sup>Department of Social Sciences, Methodology, and Statistics, Utrecht University, Utrecht, the Netherlands;

<sup>4</sup>University of Southampton, Southampton, United Kingdom;

<sup>5</sup>Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD, USA;

<sup>6</sup>Maryland Center for Sexual Health, Lutherville, MD, USA

Copyright © 2018, International Society for Sexual Medicine. Published by Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.jsxm.2018.03.082>

## INTRODUCTION

Low sexual desire and/or low sexual arousal can lead to sexual dissatisfaction, and in turn, lead to severe personal distress.<sup>1</sup> In women, distressing low desire is classified as hypoactive sexual desire disorder (HSDD) and distressing low arousal as female sexual arousal disorder in *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision*.<sup>1,2</sup> These 2 disorders have been merged in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* into female sexual interest/arousal disorder (FSIAD).<sup>3</sup> There are currently only few pharmacotherapeutic options for these women but the large number of off-label testosterone prescriptions for female sexual dysfunction (over 4 million annually) shows there is a clear need.<sup>4</sup> For a pharmacological therapy to receive US marketing authorization by the U.S. Food and Drug Administration (FDA), efficacy must be established through 1 or more patient-reported outcomes. In their regulatory guidance<sup>5</sup> the FDA formulates 4 possible (co) primary end points for pivotal phase III trials in HSDD/FSIAD: change from baseline in the level of sexual interest or desire, change from baseline in the level of sexual arousal, change from baseline in the level of distress, and change from baseline in the number of satisfying sexual events. These end points reflect different perspectives that can be taken when investigating a drug's efficacy in HSDD/FSIAD, focusing on the primary symptom (eg, level of desire), the distress that this symptom generates, or the quality of the sexual event itself. The preferred perspective, and thus end point, is partly based on the drug's characteristics.

Several therapies that are in late stages of clinical development<sup>6–8</sup> are on-demand therapies, ie, they are taken only when a woman with HSDD/FSIAD *wants to want* to have sex. These medications only increase sexual desire prior to and during sexual activity, instead of increasing it continuously as do drugs that are dosed daily (eg, flibanserin and transdermal testosterone applications). This different mode of action should be taken into account when testing on-demand medication in randomized clinical trials.

The efficacy of an on-demand drug for HSDD/FSIAD is best determined by assessing the quality of each discrete sexual event during which the drug was taken, in a given period. Assessing sexual functioning retrospectively over a longer period of time (eg, 4 weeks), as does the Female Sexual Function Index (FSFI),<sup>9</sup> gives a more distal estimation of an on-demand drug's efficacy. With this type of questionnaire, patients report on different aspects of their sexual functioning “on a whole,” thus encompassing significant periods of time over which there is no on-demand drug effect. Such a method of assessment is adequate for continuous dosing regimes because these regimens assume continuous effect, but it needlessly introduces noise in the assessment of on-demand dosing regimens.

The Sexual Event Diary (SED) is a patient-reported outcome instrument that has been developed for assessing sexual

satisfaction and sexual functioning during a discrete sexual event.<sup>10</sup> The SED is an 11-item questionnaire that is filled out by the subject within 24 hours of a sexual event. 3 Items assess the type and timing of sexual event and if medication was used, 2 binary items assess if the sexual event was satisfactory and if the subject reached orgasm, and six 5-point Likert scale items assess desire, mental arousal/excitement, physical arousal/excitement, presence and strength of distracting thoughts, the ability to let go, and experienced pleasure. The development and validation of the SED is described in Van Nes et al.<sup>10</sup> Content validity was established in 2 sets of cognitive debriefing interviews in women with HSDD aged between 21 and 70 years and psychometric assessment was carried out on data of nearly 11,000 SEDs. These data were collected during 3 double-blind, randomized, placebo-controlled, dose-finding phase II trials, investigating the efficacy and safety of on-demand drug therapies in over 400 women with HSDD, aged between 21 and 70 years.<sup>11</sup> Results of the psychometric assessment showed a 1-factor solution should be retained, based on exploratory factor analysis. This 1-factor solution substantiated the summing of all Likert scale items into a sexual function score. Reliability of the SED was confirmed based on Cronbach alpha coefficient, inter-item and item-rest correlations. Convergent validity was confirmed using Pearson correlation coefficients of the total score and domain scores of the validated FSFI<sup>6</sup> with SED sexual function score and the separate item scores. Construct validity was confirmed by comparing the mean SED scores between responders and non-responders based on the SED items “satisfaction” and “orgasm” and based on the binary single-item Subjective Evaluation of Gain questionnaire assessing benefit from the medication. Ability to detect change (responsiveness) was proven based on the Guyatt effect sizes and based on the comparison of the mean SED score changes from baseline to active treatment period (ATP) between those women reporting benefit from the medication as compared to those reporting no benefit.

The SED proved to be an excellent instrument for determining the effect of on-demand therapies on sexual function and satisfaction during discrete sexual events, in women with HSDD or FSIAD. Moreover, the SED was shown to be a suitable instrument for determining an on-demand drug's efficacy over a longer period of time (ie, over a per-patient variable number of multiple sexual events in a given period), using 1 of the FDA preferred primary end points for the indication HSDD/FSIAD “change in the number of satisfying sexual events from baseline.” This end point proved not only to be an excellent and comprehensive measure, but it also correlated strongly with all aspects of sexual functioning and it had an excellent ability to discriminate between drug responders and drug non-responders.<sup>10</sup> For a double-blind, randomized, placebo-controlled, cross-over phase II trial investigating the efficacy of an on-demand drug therapy for HSDD/FSIAD in a Dutch sample, the validated US-English SED version was translated to Dutch. Phase III trials necessitate the use of fully validated instruments that are used to collect the primary end point. As both the English

and Dutch versions are intended for use in phase III, the aim of this research was to assess the psychometric properties of the Dutch translation of the SED and thus establish its reliability and validity.

## METHODS

### Sexual Event Diary

#### Description

The validated US-English version of the SED contains 11 items that assess quantitative and qualitative aspects of a single sexual event.<sup>10</sup> It is filled out by the subject within 24 hours of a sexual event. The first 3 items inventory when the sexual event occurred, what type(s) of activities occurred, and whether on-demand study medication was taken prior to the sexual event. 2 binary questions (“yes”/“no” answer options) assess if the patient was satisfied with the sexual activity (item 4) and if she reached an orgasm (item 11). Six 5-point Likert scale items assess sexual desire (item 5), mental arousal/excitement (item 6), physical arousal/excitement (item 7), presence and strength of distracting thoughts (item 8), the ability to let go (item 9), and experienced pleasure (item 10).

#### Scoring

The 5-point Likert scale items are scored 0 (corresponding to, for example, “no desire” for item 5) to 4 (corresponding to “extreme desire” for item 5), so that a higher score indicates more positive sexual function. The scoring of item 8 (“distracting thoughts”) is reversed (“no distracting thoughts” is scored as 4) so that the answer categories of all items have the same direction regarding the quality of a sexual event. The SED sexual function score is the summation of all Likert scale items (items 5–10) for a discrete event. The scoring of the Dutch SED corresponds to the 11-item US version. In the validation of the US version,<sup>10</sup> a 16-item version is also described that preceded the 11-item version. In that 16-item version, the 5-point Likert scale items were scored 1–5 instead of 0–4. This change in scoring does not impact validity or reliability, but this difference must be kept in mind when comparing the responder definitions of the 16- and 11-item versions. Because of the change in scoring, scores of the Likert scale items of the 11-item SED will be 1 point lower than those of the 16-item SED.

#### Translation

The 11-item SED was translated to Dutch and linguistically validated by certified medical translators (Corporate Translations, formerly PharmaQuest, Banbury, United Kingdom). The questionnaire was back-translated to English to ensure that the Dutch translation was accurate and cognitive debriefing interviews were held in 5 native Dutch-speaking women, aged 18–70 years, diagnosed with FSIAD. These were performed by a qualified female investigator in the Netherlands (Corporate Translations) to test the adequacy of the translated version. All 5 women provided informed consent to participate in the cognitive debriefing interviews.

### Questionnaires Used for Assessing Validity and Responsiveness of the SED

The Patient Benefit Evaluation (PBE) questionnaire, Sexual Anamnesis Questionnaire Diagnostic (SAQ-D), and Female Sexual Distress Scale—Revised (FSDS-R) questionnaire were used for assessing validity and responsiveness of the SED. The PBE is a single-item self-report questionnaire that asks if a subject perceived meaningful benefit from the study medication over the preceding 2 weeks (answer options: “yes” and “no”). The SAQ-D is a 43-item self-report questionnaire that assesses different domains of sexual functioning (sexual desire, physical and subjective arousal, inhibition, orgasm, and pain), quality of partner relationship, and the presence of comorbid factors using Likert scale items. The SAQ-D domain scores are calculated by summing related Likert scale items. The FSDS-R is a 13-item standardized quantitative scale and measures sexually related distress in women using 5-point Likert scale items. Item 13 of the FSDS-R is related specifically to desire “How often did you feel bothered by low sexual desire,” and the summation of all items gives an overall score of perceived distress concerning sexual functioning.<sup>12–14</sup>

#### Data

The data were collected during a double-blind, randomized, placebo-controlled, cross-over trial in the Netherlands. The trial was approved by the medical ethics committee Stichting Beoordeling Ethiek Biomedisch Onderzoek (Assen, the Netherlands), and registered under primary registry trial number NTR4426 (Netherlands Trial Register). The study was carried out in compliance with the Declaration of Helsinki (2008) and with the International Conference on Harmonization Good Clinical Practice guidelines for clinical research (1996). Patients were recruited from 2 clinical sites located in the Netherlands. After written informed consent was obtained, subjects were screened. If subjects met all inclusion and exclusion criteria, they were enrolled in the study.

Subjects were women between the age of 18 and 70 years diagnosed with HSDD *as well as* FSIAD. These women can be divided into 2 groups: women with sexual problems due to low sensitivity for sexual cues or women with sexual problems due to dysfunctional over-activation of sexual inhibitory mechanisms. This subdivision is based on the dual-control model of sexual response and is substantiated by cognitive,<sup>15,16</sup> psychophysiological,<sup>6,7,15–17</sup> subjective,<sup>6,7,16</sup> neuroanatomical,<sup>18,19</sup> and pharmacological<sup>6,7,15,16,18</sup> evidence. Based on these subtypes, 2 on-demand pharmacotherapies for women with FSIAD have been developed.<sup>6,7,18</sup> This current Dutch clinical trial aimed to validate the predictive power of a demarcation formula used to differentiate between these 2 subgroups. Reliability, validity, and responsiveness of the SED were assessed using SEDs with reported on-demand medication intake resulting in a total of 1840 SEDs filled out by 139 women that formed the intention-to-treat population. The study started with a 2-week placebo run-in period (PRI) and 3 consecutive 2-week ATP

separated by 2-day wash-out periods. During the ATP regimens, subjects received tablets containing 0.5 mg testosterone and 50 mg sildenafil,<sup>20</sup> tablets containing 0.5 mg testosterone and 10 mg buspirone,<sup>21</sup> or placebo tablets in randomized order. Each subject thus underwent each drug regimen. In order to validate the predictive power of a demarcation formula, the efficacy of these on-demand therapies was determined by the assessment of sexual function and sexual satisfaction over the discrete sexual events that occurred in the ATP. Because frequency of sexual events varies per individual, the number of sexual events for each subject in each 2-week period varied, and thus the number of collected SEDs for each subject.

The SED was filled out on a secure web-based system (Viedoc Me, Pharma Consulting Group, Uppsala, Sweden) that the subject could access at home via their computer or from their portable device. Instructions on how to complete the SED were given during the start-up visit and were described in the SED as well. Subjects were instructed to complete the SED within 24 hours of any sexual activity they experienced.

## Statistical Methods

The psychometric properties of the SED were assessed using conventional methods as described by Fayers and Machin.<sup>22</sup> The methods and their criteria used to assess and interpret the results of factor analysis, reliability, validity, and responsiveness are widely used. Confirmatory factor analysis (CFA) was carried out using Mplus, version 7.3.<sup>23</sup> All other statistical analyses were performed using IBM SPSS Statistics for Windows, Version 23 (IBM Corp, Armonk, NY).

### Levels of Assessment

Reliability and validity of the SED were assessed at event level and at subject level. Analyses were performed on event level to establish reliability and validity of the questionnaire's assessment of sexual function during a discrete sexual event, and on subject level for the assessment of sexual functioning of an individual over a longer period of time. Subject level analyses can thus be used to establish reliability and validity of the primary end-point change in number of SSEs from baseline to ATP. For event level analyses, SEDs filled out by the same subjects were treated as independent observations. For subject level analyses, validity and reliability of subject mean Likert scale item scores (items 5–10), the number of SSEs (item 4), and the number of orgasms (item 11) were assessed over 2-week periods. SED mean and sum scores in the PRI and those in the 3 regimens during ATP were calculated separately.

For assessing known-groups validity, SED mean scores for satisfying sexual events were compared with those for unsatisfying sexual events as reported by SED item 4 ("Were you satisfied with the sexual activity?"), only on event level. For evaluating responsiveness on subject level, SED mean change scores from baseline (PRI) of subjects who reported experiencing meaningful benefit from the study medication were compared to subjects who reported experiencing no meaningful benefit from the study

medication, as assessed by the PBE. Furthermore, because all subjects started with the PRI regimen and because the 3 ATP regimens were assigned in a randomized order, the PRI was used as baseline in this study.

### Missing Data

On the subject level, missing data were observed only when there were no events reported during the PRI or 1 or more of the ATP regimens. 25 of 556 Observations (4.5%) were missing. Listwise deletion was used to handle missing data at subject level. The event level data set had no missing data.

### Sensitivity Analyses

The statistical methodology used at event level assumes that the SEDs filled out by the same participant are independent observations. In order to investigate that this assumption does not lead to bias in the results, sensitivity analyses were carried out by performing the psychometric assessments on a separate data set with 139 events, using only the last observed SED reported by each subject during the ATP.

### Confirmatory Factor Analysis

To test if the existing 1-factor solution found in the English-US SED validation study<sup>10</sup> also holds for the Dutch sample, CFA were performed at both the event and subject level. In these CFAs, the 5-point Likert scale items (items 5–10) were used as indicators for 1 factor measuring overall sexual function. Because at event level the original ordinal item scores are used, a CFA was conducted on the polychoric correlations using robust weighted least squares to estimate the parameters, which is the default setting in Mplus<sup>23</sup> for categorical indicators. Furthermore, it has been shown that this estimation method is robust under possible violations of underlying normality, which often is the case for ordinal data.<sup>24</sup> At subject level, a CFA was conducted on the Pearson correlations using maximum likelihood estimation to estimate the parameters. To obtain free factor loadings, the factor models were identified by fixing the factor variance to 1. Model fit was evaluated by looking at the  $\chi^2$  statistic, the root mean square error of approximation, comparative fit index, Tucker-Lewis index, and the standardized root mean square residual, which are considered popular fit indices in applied research. It has been argued to evaluate multiple fit indices instead of solely the  $\chi^2$  statistic as the  $\chi^2$  statistic is inflated by sample size. Non-significant  $\chi^2$  statistic, root mean square error of approximation, and standardized root mean square residual values smaller than 0.08, and comparative fit index and Tucker-Lewis index values larger than 0.95 are considered good model fit.<sup>25</sup> Factor loadings were considered sufficiently strong when exceeding 0.4.<sup>22,26</sup>

### Reliability

Internal consistency was assessed using Cronbach alpha coefficient, which provides a lower bound for reliability. A satisfactory Cronbach alpha coefficient is considered to be around 0.8.<sup>22,26</sup> Inter-item and corrected item-rest correlations were



assessed using Pearson correlation coefficients, and reliability was confirmed when values are larger than 0.3.<sup>22,26</sup> Inter-item correlations were used to assess the relationship between individual items within the SED. Item-rest correlations were used to assess the relationship between individual items and the total item sum score of the remaining Likert scale items. An unweighted SED sexual function sum score consisting of all relevant SED Likert scale items based on the results of the factor analyses was used.

### Validity

Construct validity at event level of items 5–10 was assessed by investigating their ability to discriminate between known groups. The known groups were satisfactory sexual events vs unsatisfactory sexual events (using SED item 4), and as sexual events with orgasm vs events without orgasm (using item 11). Using independent sample *t* test statistics, this type of validity was confirmed when the mean differences were statistically significant ( $P < .05$ ).<sup>22</sup>

At subject level, convergent validity was assessed using the Pearson correlation coefficients of the SED single-item scores and sexual function score with the related SAQ-D domain scores, the sum of all FSDS-R items, and related FSDS-R single-item scores (ie, SAQ-D and FSDS-R constructs that are not assessed by the SED were not used). Convergent validity was confirmed when Pearson correlation coefficients were at least 0.4 ( $P < .05$ ).<sup>22</sup> Only the SAQ-D sexual function factors sexual desire, physical arousal, subjective arousal, orgasm, and inhibition and FSDS-R items 1 “feeling distressed,” 11 “feeling dissatisfied,” and 13 “feeling bothered by low sexual desire” were used to assess convergent validity, as these factors and items coincide with the factors/items that SED items assess. The SAQ-D could only be used to assess convergent validity during the PRI as it was only administered at screening. Since the FSDS-R was reported every

**Table 1.** Baseline characteristic and demographics of study subjects

Total patients, n	139
Menopausal status, n (%)	
Pre-menopausal	114 (82.0)
Post-menopausal	25 (18.0)
Age, y, mean (range)	34.6 (18–69)
Race, n (%)	
White	127 (91.4)
Black	3 (2.2)
Asian	3 (2.2)
Other	6 (4.3)
Clinician secondary diagnosis, n (%)	
Female orgasmic disorder	34 (24.5)
Genito-pelvic pain/penetration disorder	0 (0.0)
Other	0 (0.0)
None	105 (75.5)
FSIAD caused by low sensitivity to sexual cues n (%)	67 (48.2)
FSIAD caused by dysfunctional over-activity of Sexual inhibitory mechanisms, n (%)	72 (51.8)

FSIAD = female sexual interest/arousal disorder.

### Responsiveness

Responsiveness is the ability of the instrument to detect change when there is a known change in the measure of interest. Responsiveness was assessed by comparing the means of the change from PRI to the ATPs in SED item scores and SED sexual function score between responders and non-responders (using the PBE), by calculating independent sample *t* test statistics. Responsiveness was confirmed when the mean differences were statistically significant ( $P < .05$ ).

Responsiveness was also assessed by determining the effect size statistics of the ability of the SED to measure change in sexual functioning using Guyatt responsiveness index.<sup>27,28</sup>

$$\text{Guyatt index} = \frac{(\text{Change in SED scores for responders}) - (\text{Change in SED scores for non-responders})}{SD \text{ of change in SED scores for non-responders}}$$

2 weeks, convergent validity using the FSDS-R was assessed during both PRI and ATP.

Also at subject level, known-groups validity was assessed by comparing the mean SED item scores and SED sexual function score over the ATP (PRI was excluded) between responders and non-responders, using independent sample *t* test statistics. This type of validity was confirmed when the mean differences were statistically significant ( $P < .05$ ).<sup>22</sup> Responders were those subjects who indicated experiencing meaningful benefit of the study medication (PBE); non-responders were those who reported no benefit. All reported PBEs during ATP were included.

Effect sizes of about 0.20 represent small effects, those of about 0.50 represent moderate effects, and those  $\geq 0.80$  represent large effects.<sup>18</sup> A 2-sided 5% significance level was adopted for all statistical tests.

## RESULTS

### Subject Characteristics

Baseline characteristics and demographics of 139 women with FSIAD who were included in psychometric assessments are shown in Table 1.

**Table 2.** Factor loadings and fit indices of confirmatory factor analysis for the Sexual Event Diary items at event level and subject level

	Level of assessment ( $\chi^2$ )	
	Event	Subject
Fit indices		
$\chi^2$	23.263	13.116
Root mean square error of approximation	0.051	0.055
Comparative fit index	1.000	0.997
Tucker-Lewis index	0.999	0.991
Standardized root mean square residual	0.007	0.013
Factor loadings Sexual Event Diary items		
5. How would you rate your level of sexual desire during the sexual activity?	0.932	0.790
6. How mentally aroused or excited did you become during the sexual activity?	0.928	0.797
7. How physically aroused or excited did you become during the sexual activity?	0.865	0.761
8. To what extent did you have distracting thoughts?*	0.471	0.394
9. To what extent were you able to let yourself go?	0.801	0.720
10. How pleasurable was the sexual activity to you?	0.923	0.698

\*Reversed variable.

## Event Level

### Confirmatory Factor Analysis

Results of the CFA are presented in [Table 2](#). The fitted 1-factor model was extended with correlated error variances between the items to account for additional covariation that is not explained by the factor. This approach is common for items that are similarly worded.<sup>25</sup> Although the  $\chi^2$  statistic was significant (degrees of freedom = 4,  $P = .0001$ ), all other fit indices reveal the 1-factor model fits the data very well. All SED items had strong factor loadings, with a less strong, but still sufficiently strong contribution of item 8 “distracting thoughts.” Because the 1-factor model could be confirmed, the SED sexual function score was derived using Likert scale SED items 5–10.

### Reliability

Cronbach alpha coefficient was high (0.90;  $n = 1,840$ ) and the Pearson correlation coefficients, which were calculated for assessing the SED inter-item and item-rest correlations, were all larger than 0.30 ( $P < .0001$ ) ([Supplementary Table A](#)). An exception was the moderate correlation between “distracting thoughts” and “orgasm” ( $r = 0.22$ ). The average inter-item correlation was 0.55.

### Validity

All SED item scores and the SED sexual function score showed strong construct validity. The mean differences in SED scores between “yes” and “no” responders on SED items measuring

“satisfaction” and “orgasm” were highly significant ( $P < .0001$ ), and in the expected direction ([Table 3](#)).

## Subject Level

### Confirmatory Factor Analysis

Results of the CFA are presented in [Table 2](#). Also here, the fitted 1-factor model was extended with correlated error variances between the average item scores to account for additional covariation that is not explained by the factor. Although the  $\chi^2$  statistic was significant (degrees of freedom = 5,  $P = .0223$ ), all other fit indices reveal the 1-factor model fits the data very well. All SED items had strong factor loadings, with a less strong, but still moderate contribution of item 8 “distracting thoughts.” These factor loadings are all consistently lower than the ones in the event level CFA. This can be explained by the fact that at subject level, average item scores are used rather than original item scores that resulted in a loss of information. However, as results show the 1-factor model could also be confirmed at subject level, the SED sexual function score was derived using (average) Likert scale SED items 5–10.

### Reliability

Cronbach alpha coefficient was high (0.91;  $n = 531$ ) and the Pearson correlation coefficients, which were calculated for assessing the SED inter-item and item-rest correlations, were all larger than 0.30 ( $P < .0001$ ) ([Supplementary Table B](#)). An exception was the moderate correlation between “distracting thoughts” and “orgasms” ( $r = 0.18$ ,  $P < .0001$ ). The average inter-item correlation was 0.54.

### Validity

Convergent validity was deemed adequate. The correlation coefficients between the SAQ-D domains with their related PRI SED items ranged from 0.21–0.51 ([Supplementary Table C](#)). Correlation coefficients between the FSDS-R with their related SED items and between the FSDS-R total score and the SED sexual function score ranged from 0.44–0.58 during the PRI and ATPs ([Supplementary Table D](#)). These results provided support for adequate to strong convergence of the 11-item SED. In general, lower but adequate convergent validity based on the SAQ-D was expected. The SAQ-D was administered at screening and was compared with the SEDs over the subsequent PRI, during which placebo was taken. The FSDS-Rs were reported at the end of the PRI and following each ATP, which overlaps with the periods in which the SEDs were reported. This may explain the somewhat stronger correlation coefficients for the FSDS-R.

Known-groups validity was good, because responders, as defined by the PBE, scored significantly higher compared to non-responders on all items and the sexual function score during the ATP ( $P < .05$ ) ([Supplementary Table E](#)).

**Table 3.** Mean (SEM) in Sexual Event Diary scores separately for answering item 4 “Were you satisfied with the sexual activity?” and 11 “Did you have an orgasm?” with “yes” or “no”

SED item	4. Were you satisfied with the sexual activity?						11. Did you have an orgasm?					
	Yes		No		Test statistics		Yes		No		Test statistics	
	Mean	SEM	Mean	SEM	t Value	P value*	Mean	SEM	Mean	SEM	t Value	P value*
5. How would you rate your level of sexual desire during the sexual activity?	2.27	0.02	0.96	0.03	33.66	<.0001	2.12	0.03	1.39	0.03	16.19	<.0001
6. How mentally aroused or excited did you become during the sexual activity?	2.16	0.02	0.85	0.03	34.34	<.0001	2.01	0.03	1.29	0.03	16.21	<.0001
7. How physically aroused or excited did you become during the sexual activity?	2.42	0.02	1.18	0.03	31.04	<.0001	2.35	0.03	1.53	0.03	18.50	<.0001
8. To what extent did you have distracting thoughts? <sup>†</sup>	2.52	0.03	1.61	0.04	17.57	<.0001	2.41	0.03	1.92	0.04	9.62	<.0001
9. To what extent were you able to let yourself go?	2.47	0.03	1.11	0.04	29.40	<.0001	2.42	0.03	1.47	0.04	19.01	<.0001
10. How pleasurable was the sexual activity to you?	2.60	0.02	1.19	0.03	42.15	<.0001	2.48	0.03	1.63	0.03	20.60	<.0001
SED sexual function	14.44	0.11	6.90	0.14	41.96	<.0001	13.79	0.15	9.22	0.16	20.75	<.0001

Satisfied = “yes”: n = 1,125 and “no”: n = 715.

Orgasm = “yes”: n = 922 and “no”: n = 918.

SED = Sexual Event Diary.

\*2-sided tests were used.

†Reversed variable.

**Table 4.** Responsiveness—mean (SEM) change in Sexual Event Diary scores from placebo run-in period to active treatment periods in responders and non-responders as defined by the Patient Benefit Evaluation questionnaire

SED item	Responder (yes) n = 130*		Non-responder (no) n = 251*		t Value	P value <sup>†</sup>	Guyatt responsiveness
	Mean change	SEM	Mean change	SEM			
4. Were you satisfied with the sexual activity? <sup>‡</sup>	1.31	0.14	−0.10	0.09	8.60	<.0001	0.98
5. How would you rate your level of sexual desire during the sexual activity?	0.73	0.07	0.00	0.05	9.08	<.0001	0.98
6. How mentally aroused or excited did you become during the sexual activity?	0.64	0.07	−0.10	0.05	8.78	<.0001	0.94
7. How physically aroused or excited did you become during the sexual activity?	0.65	0.07	−0.12	0.05	8.67	<.0001	0.94
8. To what extent did you have distracting thoughts? <sup>§</sup>	0.37	0.07	−0.22	0.06	5.74	<.0001	0.59
9. To what extent were you able to let yourself go?	0.49	0.07	−0.22	0.06	7.72	<.0001	0.79
10. How pleasurable was the sexual activity to you?	0.47	0.06	−0.16	0.05	8.07	<.0001	0.85
11. Did you have an orgasm? <sup>‡</sup>	0.62	0.13	0.00	0.08	4.13	.0001	0.51
SED Sexual satisfaction	3.36	0.33	−0.82	0.26	9.73	<.0001	1.02

SED = Sexual Event Diary.

\*Each patient filled out 3 Patient Benefit Evaluation questionnaires, 1 for each of the 3 drug regimens (2 active and 1 placebo). The large number of non-responders is explained by the fact that patients typically respond to only one of the active treatments, and also not to placebo.

†2-sided tests were used.

‡Counts of “yes” answers on these items were used according to efficacy analyses.

§Reversed variable.

## Responsiveness

Responders had a significantly larger increase in change from PRI to ATP in SED item scores and the SED sexual function score compared to non-responders ( $P < .001$ ) (Table 4), showing strong responsiveness. The Guyatt effect sizes ranged from 0.51–1.02, indicating adequate to very good ability to detect changes in SED item scores (Table 4).

## Sensitivity Analyses

The sensitivity analyses gave similar reliability, validity, and responsiveness of the SED compared to the analyses performed on the event level and subject level (data not shown). The results of the sensitivity analyses supported the event and subject level analyses, confirming that our approaches were valid and did not influence the conclusions.

## DISCUSSION

This study examined the psychometric properties of the Dutch translation of the 11-item SED for assessing on-demand drug efficacy in women diagnosed with HSDD and FSIAD. Previous psychometric assessment and cognitive debriefing interviews with U.S. patients diagnosed with HSDD revealed the US-English version of the SED is reliable, valid, and responsive.<sup>10</sup> Based on the psychometric assessment results obtained in this study, the same conclusions concerning the reliability, validity, and responsiveness can be drawn for the Dutch translation of the 11-item SED, thereby confirming the US-English version validation results.

The reproduced 1-factor solutions and internal consistency measures indicate the Dutch SED measures the same factor (sexual function) with excellent reliability. Item 8, “distracting thoughts,” did have lower factor loadings than the other items, but the loadings were still adequate. Known-groups validity and responsiveness results revealed comparable discriminating ability between responders and non-responders based on the single-item benefit questionnaire and on the dichotomous items “satisfaction” and “orgasm.” It was also found that the primary end-point change from baseline in the number of satisfying sexual events is an excellent measure, which is in line with the US-English validation results. In this study the PRI was used as baseline, a baseline with a placebo effect already superimposed, making it a more conservative comparator. The baseline period that was used for the validation of the US-English SED was a medication-free period. Therefore, the changes from PRI in this study were, on average, smaller than the changes from baseline in the US-English validation study. The larger effects observed in this study are due to more statistical power to detect smaller changes caused by the cross-over design (all subjects were included thrice in the data). However, the instrument must be able to detect change for respondents regardless of the (baseline) comparator, making the difference between a PRI or a medication-free baseline inconsequential.

Although convergent validity was deemed adequate in this study, the correlations in this study were somewhat smaller compared to the US-English SED validation study.<sup>10</sup> This is probably caused by the questionnaires used to assess convergent validity, the SAQ-D and the FSDS-R. Both instruments are less adequate to assess convergent validity for the SED than the FSFI<sup>9</sup> that was used in the U.S. validation. The FSFI assesses sexual function, as does the SAQ-D, but the latter was only administered during screening. Therefore, the SAQ-D and SEDs taken during PRI did not match regarding the time period that they assessed. Moreover, during PRI, possible placebo effects may have occurred, thereby not reflecting a true baseline score. The use of the FSDS-R was less adequate for the intended use because it measures distress related to sexual problems, and not sexual function per se.

Another limitation of this study was that the psychometric assessments were performed predominantly in pre-menopausal women. It is not expected that a psychometric analysis performed in post-menopausal women alone will lead to different conclusions as there is no literature that suggests that there is a difference between these women with respect to which aspects of sexual function are important. This will need to be confirmed in future research.

The SED was developed and validated as a part of a drug development program for HSDD/FSIAD. Development of and modifications to the SED were based on the premise that the instrument had to be a valid and reliable tool for use in such a program. The data that were used for the validation described here were also gathered in this program. Despite this focus, however, the SED may also show merit in the assessment of sexual functioning during discrete sexual events in other areas of research, eg, recreational drug use and sexual risky behaviors.

## CONCLUSION

The Dutch translated 11-item SED has proven to be an excellent tool for assessing female sexual satisfaction and sexual functioning over a single sexual event, and is therefore suitable for use in clinical trials assessing the efficacy of on-demand drugs in the FSIAD populations.

**Corresponding Author:** Jos Bloemers, PhD, Emotional Brain BV, Louis Armstrongweg 78 Almere, 1311 RL, the Netherlands. Tel: +31(0)36-741-0120; Fax: +31(0)36-532 09 84; E-mail: [j.bloemers@emotionalbrain.nl](mailto:j.bloemers@emotionalbrain.nl)

**Conflict of Interest:** Y.vN. and R.K. are employees of Emotional Brain (EB). J.B., K.vR., and J.G. are employees of EB and own shares or share options in EB. P.G.M.vdH. is advisor to EB. L.D. is a member of the scientific advisory board for Palatin Pharmaceuticals, S1 Biopharm, Emotional Brain, Acerus, and Endoceutics. A.T. is CEO of EB and a shareholder of EB.

**Funding:** These studies were funded by Emotional Brain BV.



## STATEMENT OF AUTHORSHIP

### Category 1

#### (a) Conception and Design

Yvonne van Nes; Jos Bloemers; Peter G. M. van der Heijden; Leonard DeRogatis; Adriaan Tuiten

#### (b) Acquisition of Data

Yvonne van Nes; Jos Bloemers; Kim van Rooij; Jeroen Gerritsen

#### (c) Analysis and Interpretation of Data

Yvonne van Nes; Jos Bloemers; Rob Kessels; Peter G. M. van der Heijden

### Category 2

#### (a) Drafting the Article

Yvonne van Nes; Jos Bloemers; Rob Kessels

#### (b) Revising It for Intellectual Content

Yvonne van Nes; Jos Bloemers; Rob Kessels; Peter G. M. van der Heijden; Kim van Rooij; Jeroen Gerritsen; Leonard DeRogatis; Adriaan Tuiten

### Category 3

#### (a) Final Approval of the Completed Article

Yvonne van Nes; Jos Bloemers; Rob Kessels; Peter G. M. van der Heijden; Kim van Rooij; Jeroen Gerritsen; Leonard DeRogatis; Adriaan Tuiten

## REFERENCES

1. Davison SL, Bell RJ, LaChina M, et al. The relationship between self-reported sexual satisfaction and general well-being in women. *J Sex Med* 2009;6:2690-2697.
2. American Psychiatric Association. DSM-IV: Diagnostic and statistical manual of mental disorders. 4th edition. Text revision. Washington, DC: American Psychiatric Press; 2000.
3. American Psychiatric Association. DSM-5: Diagnostic and statistical manual of mental disorders. 5th edition. Arlington, VA: American Psychiatric Publishing; 2013.
4. Snabes M, Milling W, Simes S. Without FDA-approved testosterone to treat women with hypoactive sexual desire disorder providers rely on off-label prescribing. *J Sex Med* 2011;8:56-77.
5. US Food and Drug Administration. Low sexual interest, desire, and/or arousal in women: developing drugs for treatment. Guidance for industry. Draft guidance. Available at: <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM526362.pdf>; Accessed November 27, 2017.
6. Poels S, Bloemers J, van Rooij K, et al. Toward personalized sexual medicine (part 2): testosterone combined with a PDE5 inhibitor increases sexual satisfaction in women with HSDD and FSAD, and a low sensitive system for sexual cues. *J Sex Med* 2013;10:810-823.
7. van Rooij K, Poels S, Bloemers J, et al. Toward personalized sexual medicine (part 3): testosterone combined with a serotonin 1A receptor agonist increases sexual satisfaction in women with HSDD and FSAD, and dysfunctional activation of sexual inhibitory mechanisms. *J Sex Med* 2013;10:824-837.
8. Clayton AH, Althof SE, Kingsberg S, et al. Bremelanotide for female sexual dysfunctions in premenopausal women: a randomized, placebo-controlled dose-finding trial. *Womens Health (Lond)* 2016;12:325-337.
9. Rosen R, Brown C, Heiman J, et al. The Female Sexual Function Index (FSFI): a multidimensional self-report instrument for the assessment of female sexual function. *J Sex Marital Ther* 2000;26:191-208.
10. van Nes Y, Bloemers J, van der Heijden PGM, et al. The Sexual Event Diary (SED): development and validation of a standardized questionnaire for assessing female sexual functioning over discrete sexual events. *J Sex Med* 2017;14:1438-1450.
11. Tuiten A, van Rooij K, Bloemers J, et al. Efficacy and safety of on-demand use of 2 treatments designed for different etiologies of female sexual interest/arousal disorder: 3 randomized clinical trials. *J Sex Med* 2018;15:201-216.
12. ter Kuile MM, Brauer M, Laan E. The Female Sexual Function Index (FSFI) and the Female Sexual Distress Scale (FSDS): psychometric properties within a Dutch population. *J Sex Marital Ther* 2006;32:289-304.
13. DeRogatis LR, Clayton A, Lewis-D'Agostino D, et al. Validation of the female sexual distress scale revised for assessing distress in women with hypoactive sexual desire disorder. *J Sex Med* 2008;5:357-364.
14. Derogatis LR, Rosen R, Leiblum S, et al. The Female Sexual Distress Scale (FSDS): initial validation of a standardized scale for assessment of sexually related personal distress in women. *J Sex Marital Ther* 2002;28:317-330.
15. van der Made F, Bloemers J, van Ham D, et al. Childhood sexual abuse, selective attention for sexual cues and the effects of testosterone with or without vardenafil on physiological sexual arousal in women with sexual dysfunction: a pilot study. *J Sex Med* 2009;6:429-439.
16. van der Made F, Bloemers J, Yassem WE, et al. The influence of testosterone combined with a PDE5-inhibitor on cognitive, affective, and physiological sexual functioning in women suffering from sexual dysfunction. *J Sex Med* 2009;6:777-790.
17. Gerritsen G, van der Made F, Bloemers J, et al. The clitoral photoplethysmograph: a new way of assessing genital arousal in women. *J Sex Med* 2009;6:1678-1687.
18. Bloemers J, van Rooij K, Poels P, et al. Toward personalized sexual medicine (part 1): integrating the "dual control model" into differential drug treatments for hypoactive sexual desire disorder and female sexual arousal disorder. *J Sex Med* 2013;10:791-809.
19. Bloemers J, Scholte HS, van Rooij K, et al. Reduced gray matter volume and increased white matter fractional anisotropy in women with hypoactive sexual desire disorder. *J Sex Med* 2014;11:753-767.
20. Bloemers J, van Rooij K, de Leede L, et al. Single dose sublingual testosterone and oral sildenafil vs a dual route/dual release fixed dose combination tablet: a pharmacokinetic comparison. *Br J Clin Pharmacol* 2016;81:1091-1102.
21. Van Rooij K, de Leede L, Frijlink HW, et al. Pharmacokinetics of a prototype formulation of sublingual testosterone and a

- buspirone tablet, versus an advanced combination tablet of testosterone and buspirone in healthy premenopausal women. *Drugs R D* 2014;14:125-132.
22. Fayers PM, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. 2nd edition. Chichester, United Kingdom: John Wiley and Sons; 2007.
23. Muthén LK, Muthén BO. Mplus user's guide. 8th edition. Los Angeles, CA: Muthén and Muthén; April 2017.
24. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods* 2004; 9:466-491.
25. Brown TA. Confirmatory factor analysis for applied research. New York, NY: Guilford Press; 2014.
26. Field A. Discovering statistics using SPSS. 2nd edition. London, United Kingdom: Sage Publications; 2005.
27. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
28. Guyatt GH, Walter SD, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171-178.

## SUPPLEMENTARY DATA

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.jsxm.2018.03.082>.