

RESEARCH

Open Access



# Psychometric validation of the 1-month recall Uterine Fibroid Symptom and Health-Related Quality of Life questionnaire (UFS-QOL)

Karin S. Coyne<sup>1\*</sup>, Amanda Harrington<sup>2</sup>, Brooke M. Currie<sup>1</sup>, Jun Chen<sup>1</sup>, Patrick Gillard<sup>2</sup> and James B. Spies<sup>3</sup>

## Abstract

**Background:** To evaluate the psychometric characteristics of the 1-month recall Uterine Fibroid Symptom and Health-Related Quality of Life questionnaire (UFS-QOL), including the Revised Activities subscale.

**Methods:** VENUS I and II were phase III, randomized, double-blind, placebo-controlled trials of ulipristal acetate in women with uterine fibroids (UF) and abnormal uterine bleeding. Women completed the 1-month recall UFS-QOL at baseline and after 12 weeks' treatment. Uterine bleeding was assessed via a daily diary (both studies); the Patient Global Impression of Improvement scale (PGI-I) was completed in VENUS II. Psychometric analyses examined internal consistency reliability and construct validity of the UFS-QOL; confirmatory factor analysis (CFA) compared model fit of the original and Revised Activities subscales. Analyses were conducted separately for VENUS I and II.

**Results:** One hundred and fifty-seven patients in VENUS I and 429 in VENUS II were included. Changes in mean Symptom Severity and health-related quality of life (HRQoL) scale scores indicated symptom burden reductions and HRQoL improvements. Cronbach's alpha coefficients were high at baseline and after 12 weeks' treatment (all  $\geq 0.76$ , meeting the  $>0.70$  threshold), demonstrating strong internal consistency reliability. Correlations between UFS-QOL scores and bleeding diary responses (range:  $-0.35$  to  $-0.63$ ), and UFS-QOL scores and PGI-I responses (range:  $-0.48$  to  $-0.70$ ), ranged from moderate to strong after 12 weeks' treatment (all  $p < 0.0001$ ). Patients with absence of bleeding or controlled bleeding after 12 weeks' treatment scored significantly better ( $p < 0.001$ ) on each UFS-QOL scale than patients not achieving those end points, supporting construct validity. CFA confirmed model fit for the Revised Activities subscale.

**Conclusions:** The 1-month recall UFS-QOL, including the Revised Activities subscale, is a valid, reliable measure to assess UF symptoms and their impact on HRQoL.

**Trial registration:** ClinicalTrials.gov, [NCT02147197](https://clinicaltrials.gov/ct2/show/study/NCT02147197). Registered May 26, 2014; retrospectively registered. ClinicalTrials.gov, [NCT02147158](https://clinicaltrials.gov/ct2/show/study/NCT02147158). Registered May 26, 2014; retrospectively registered.

**Keywords:** Health-related quality of life, Symptoms, Uterine fibroids, Patient-reported outcomes

\* Correspondence: [Karin.Coyne@evidera.com](mailto:Karin.Coyne@evidera.com)

<sup>1</sup>Evidera, 7101 Wisconsin Avenue, Suite 1400, Bethesda, MD 20814, USA  
Full list of author information is available at the end of the article

## Background

Uterine fibroids (UF) are among the most common benign neoplasms of the female pelvic region. UF incidence increases as women approach menopause and has been reported to affect up to 80% of women by the age of 50 years [1]. Up to half of women with UF experience clinical symptoms [2], including abnormal uterine bleeding (AUB) and pain [3], which can cause significant emotional and psychological distress [4]. A national survey of women in the United States aged 29–59 years with self-reported symptomatic UF revealed that 31% of respondents reported symptoms interfering with physical activities “all/most of the time”, while 22% reported symptoms interfering with daily/social activities “all/most of the time.” In addition, almost one-third of employed respondents reported missing work due to their symptoms [5].

The symptoms of UF and their negative impact on health-related quality of life (HRQoL) and activities of daily living are some of the reasons why women seek therapy for UF. Patient-reported outcome (PRO) measures are, therefore, appropriate tools to measure the impact and outcome of interventions [6]. The Uterine Fibroid Symptom and Health-Related Quality of Life questionnaire (UFS-QOL) is widely used to evaluate patient-reported UF symptoms and their impact on HRQoL, and is the only disease-specific instrument developed and validated in a population of women with UF [7]. It was developed based on qualitative input from patients with UF; the original validation demonstrated its ability to discriminate between women with and without UF and also between varying patient-reported disease severity [7]. Furthermore, the UFS-QOL has been shown to be highly responsive to change following treatment [6].

In the original version of the UFS-QOL, patients are instructed to consider their experiences with UF during the previous 3 months. The instrument has since been modified to incorporate a shorter 1-month recall period to minimize recall bias [8] and to provide a more precise assessment of treatment effect based on a monthly menstrual cycle. The availability of multiple recall versions of the UFS-QOL also provides utility for women who do not have a monthly cycle; the 3-month recall version may be more useful for women who have less frequent menstrual cycles. In addition, a Revised Activities subscale has been created to include the most relevant items pertaining to physical and social activities. This scale was developed based on recent qualitative focus groups in which participants were asked to indicate whether each item of the UFS-QOL was relevant to them. The two items that ranked lowest in terms of relevancy to patients on the Activities subscale were removed during data analysis to create the Revised Activities subscale (i.e. the items were not removed from the questionnaire itself). As a result of these changes to the UFS-QOL, further validation is warranted.

An a priori planned validation of the 1-month recall UFS-QOL, including the Revised Activities subscale, was carried out to evaluate the instrument’s psychometric properties using data from two trials of ulipristal acetate (UPA), an investigational, orally administered selective progesterone receptor modulator that reversibly blocks progesterone receptors in its target tissues (endometrium, pituitary, and UF) [9, 10]. UPA has been shown in studies to provide therapeutic effects in reducing AUB in women with UF [11–15], including two pivotal phase III trials conducted in the United States and Canada (VENUS I [UL1309; NCT02147197] and VENUS II [UL1208; NCT02147158]) [16, 17], in study populations representative of women with UF in the US general population.

## Methods

### Study designs and patients

This analysis included data from VENUS I and VENUS II, two phase III, multicenter, randomized, double-blind, placebo-controlled trials to assess the safety and efficacy of UPA for the treatment of AUB associated with UF. Both studies included pre-menopausal women aged 18–50 years who had: ultrasound evidence of at least one discrete UF; a history of cyclic ( $\geq 22$  and  $\leq 35$  days) AUB; and menstrual blood loss  $\geq 80$  mL. Key exclusion criteria were: a history of uterine surgery that would interfere with the study end points; known coagulation disorder; and a history of, or current, uterine, cervical, ovarian, or breast cancers [16, 17]. VENUS I included 157 patients randomized to placebo, UPA 5 mg, or UPA 10 mg for 12 weeks of treatment, followed by a 12-week drug-free follow-up period. VENUS II included 432 patients randomized to placebo followed by UPA, UPA followed by placebo, or two courses of UPA. The two 12-week treatment courses were separated by a drug-free interval of two menses. The second treatment course in VENUS II was followed by a 12-week drug-free follow-up period. This report follows recommendations described in the CONSORT PRO Extension [18].

### Questionnaires and assessments

#### UFS-QOL

The 37-item, self-administered UFS-QOL measures Symptom Severity (eight items) and HRQoL (29 items) and has been previously validated [6, 7, 19, 20]. The HRQoL Total scale consists of six subscales: Concern, Activities, Energy/Mood, Control, Self-Consciousness, and Sexual Function. Response options for Symptom Severity scale items are scored from 1 (“Not at all”) to 5 (“A very great deal”); response options for items in the HRQoL subscales range from 1 (“None of the time”) to 5 (“All of the time”). The Symptom Severity scale, HRQoL subscales, and HRQoL Total scale scores are summed and transformed into

a 0–100-point scale, with higher Symptom Severity scores indicating greater symptom severity and higher HRQoL scores indicating better HRQoL. The Symptom Severity scale is unidimensional and the HRQoL subscales can be treated as unidimensional scales of a multidimensional construct (HRQoL); the HRQoL Total scale is a sum of the HRQoL subscales.

Patients were instructed to consider their experiences with UF over a modified recall period of 1 month. The Revised Activities subscale, a shorter version of the original Activities subscale, was included in all validation analyses. The 1-month recall UFS-QOL was completed by patients at baseline (Visit 1; first on-treatment visit in both studies) and after 12 weeks of treatment (Visit 2; end of the 12-week treatment period in VENUS I and end of Treatment Course 1 in VENUS II), or on early withdrawal for patients who withdrew after Visit 1 and before Visit 2 (for both studies).

#### **Patient Global Impression of Improvement scale**

The Patient Global Impression of Improvement scale (PGI-I) is a self-administered measure used to rate patient-perceived response of a condition to therapy. The PGI-I, administered in VENUS II only, asked: “During treatment with study drug, how would you describe your menstrual/vaginal bleeding compared to before you started study drug?”. Participants responded on a 7-point Likert scale with the following options: 1, “Very much better”; 2, “Much better”; 3, “A little better”; 4, “No change”; 5, “A little worse”; 6, “Much worse”; and 7, “Very much worse”. The PGI-I was completed at Visit 2 or on early withdrawal.

#### **Bleeding diary**

Uterine bleeding was recorded in an electronic diary in both studies [16, 17]. A patient’s heaviest bleeding experienced over the preceding 24 h was captured by the following terms: “None”, no bleeding and no spotting; “Spotting”, evidence of minimal blood loss that does not require the use of sanitary protection (except for panty liners); “Bleeding”, evidence of blood loss that requires the use of sanitary pads or tampons; “Heavy bleeding”, more than normal bleeding relative to your experience, or the passage of clots. Absence of bleeding was defined as having no bleeding days during the last 35 consecutive days on treatment counting backward from the earlier of Day 84 or the last dose date in the treatment period (VENUS I) or in Treatment Course 1 (VENUS II). Controlled bleeding was defined as having 0 days of heavy bleeding and  $\leq 8$  days of bleeding within the analysis window (the last 56 days of treatment counting backward from the earlier of Day 84 or the last dose date in the treatment period [VENUS I] or in Treatment Course 1 [VENUS II]). No controlled bleeding was defined as

having  $\geq 1$  day of heavy bleeding or  $\geq 9$  days of bleeding within the analysis window. The thresholds for bleeding, absence of bleeding, and controlled bleeding were identified a priori, based on the primary end points in the VENUS I and II clinical trials.

#### **Statistical analyses**

Observed UFS-QOL and PGI-I scores were used in all analyses. Missing bleeding diary data were imputed consistent with VENUS I and II protocols. Scoring of the questionnaires was performed according to the developers’ guidelines. All statistical tests were two-sided and used a significance level of 0.05 unless otherwise noted. Baseline analyses were carried out on the intent-to-treat population using an observed cases approach, defined as patients who completed at least one item of the UFS-QOL at baseline. The per protocol population was defined as all randomized patients who completed the treatment period, in addition to completing at least one item of the UFS-QOL after 12 weeks of treatment (PRO approach). Descriptive analyses were performed on baseline patient demographic and clinical characteristics. Distributional characteristics of UFS-QOL scores were examined at baseline and after 12 weeks of treatment. Analyses were conducted separately for VENUS I and II in order to assess the reproducibility of the results; additionally, minor differences between the two trials were present, such as the PGI-I being administered only in VENUS II.

Psychometric analyses were conducted on the UFS-QOL, including the Revised Activities subscale, at baseline and after 12 weeks of treatment to examine internal consistency reliability and construct validity (convergent and known groups validity).

Internal consistency reliability explores associations between different items within a scale [21]. Cronbach’s coefficient alpha was calculated for each UFS-QOL scale at baseline and after 12 weeks of treatment; a value of  $>0.70$  was considered acceptable to demonstrate internal consistency [22]. Validity refers to the extent to which an instrument measures what it purports to measure [21]. Convergent validity is the extent to which scores from the instrument are related to scores from other related instruments or concepts [21]. Spearman’s rank correlation coefficients were used to establish convergent validity between the UFS-QOL scales and bleeding diary assessments (number of bleeding days and heavy bleeding days) at baseline and after 12 weeks of treatment, and between the UFS-QOL scales and PGI-I after 12 weeks of treatment (VENUS II only).

Known groups validity is the extent to which scores from an instrument are distinguishable from groups that differ by a key indicator, often clinical in nature [20]. Known groups validity of the UFS-QOL was assessed by

number of bleeding days (categorized by  $\leq 5$ ,  $>5$  to 9, and  $>9$  days), achievement of absence of bleeding, and achievement of controlled bleeding for both VENUS I and II. Known groups validity was also assessed based on PGI-I responses after 12 weeks of treatment in VENUS II. Specifically, two separate assessments were conducted based on patients' PGI-I data: 1) by individual PGI-I score; and 2) by the collapsed PGI-I response categories of "Improved" (responses of "Very much better", "Much better", and "A little better"), "No change", and "Worsened" (responses of "A little worse", "Much worse", and "Very much worse").

The unidimensional factor structures of the Activities and Revised Activities subscales were examined using confirmatory factor analysis (CFA) using *Mplus*. Model fit was assessed by examining three fit statistics: the Comparative Fit Index (CFI) – the model was considered to have a good fit if the CFI was  $\geq 0.90$  [23]; Root Mean Square Error of Approximation (RMSEA) – the goodness of fit of the model was considered acceptable for values  $< 0.07$  [24]; and Standardized Root Mean Square Residual (SRMR) – the goodness of fit of the model was considered acceptable for values  $\leq 0.08$  [25].

All analyses were performed in SAS version 9.4 and *Mplus* version 7.4 [26, 27].

## Results

### Patients

All 157 randomized patients in VENUS I had baseline UFS-QOL data; 135 completed at least one item of the UFS-QOL after 12 weeks of treatment (or early termination) (Fig. 1). Of 432 randomized patients in VENUS II, 429 had baseline UFS-QOL data and 348 completed at least one item of the UFS-QOL after 12 weeks of treatment (or early termination) (Fig. 1). Mean (standard deviation [SD]) age was 41.1 (5.4) years and 41.0 (5.6) years in VENUS I and II, respectively. Most patients were black (68.8% in VENUS I and 66.9% in VENUS II) and mean (SD) body mass index was 31.7 (8.0)  $\text{kg}/\text{m}^2$  and 32.2 (7.9)  $\text{kg}/\text{m}^2$  in VENUS I and II, respectively (Table 1). There were no significant differences ( $p > 0.05$ ) between women who completed 12 weeks of treatment and those who discontinued in terms of age, race, ethnicity, and body mass index.

### UFS-QOL: scale analysis

Descriptive statistics for UFS-QOL scales are shown in Table 2. In both studies at baseline, the mean Symptom Severity scale score was relatively high (62.0 in VENUS I and 65.5 in VENUS II), decreasing to approximately half its baseline value after 12 weeks of treatment (30.5 for patients in VENUS I and 33.0 for patients in VENUS II), indicating a reduction in symptom burden. Mean scores on the HRQoL Total scale and six HRQoL subscales

(and the Revised Activities subscale) at baseline ranged from 22.9 in VENUS I and 21.0 in VENUS II (both Concern) to 45.0 in VENUS I and 43.3 in VENUS II (both Control). After 12 weeks of treatment, mean scores increased, ranging from 66.1 in VENUS I (Concern) and 62.9 in VENUS II (Self-Consciousness) to 77.7 in VENUS I and 73.9 in VENUS II (both Control), indicating improved HRQoL.

### Internal consistency reliability

For both VENUS I and II, each of the UFS-QOL scales showed strong internal consistency at baseline and after 12 weeks of treatment, as demonstrated by the high Cronbach's alpha values (range at baseline: 0.76 [Self-Consciousness] to 0.96 [HRQoL Total] in VENUS I and 0.79 [Self-Consciousness] to 0.96 [HRQoL Total] in VENUS II; higher alphas were reported after 12 weeks of treatment) (Additional file 1: Table S1).

### Convergent validity

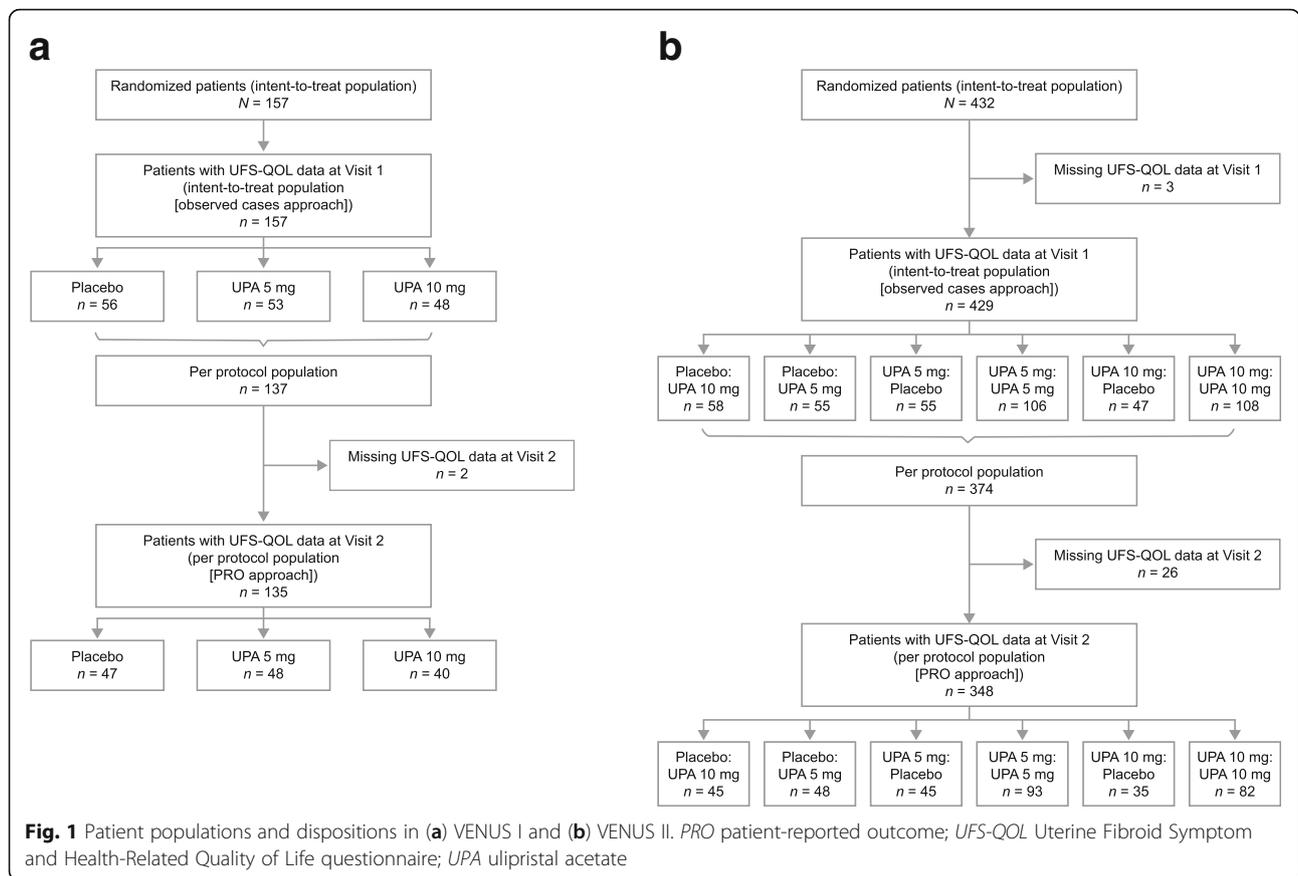
In both studies, correlations at baseline between UFS-QOL scales and the number of heavy bleeding days were significant (except for the Sexual Function subscale in VENUS I), but weak, as is to be expected for correlations between objective and subjective measures ( $r_s = 0.18$  [ $p < 0.05$ ] for the Symptom Severity scale and ranging from  $r_s = -0.15$  [ $p = \text{not significant (NS)}$ ] to  $-0.25$  [others  $p < 0.05$ ] for the HRQoL subscales in VENUS I;  $r_s = 0.24$  [ $p < 0.001$ ] for Symptom Severity and ranging from  $r_s = -0.16$  to  $-0.26$  [ $p < 0.001$ ] for the HRQoL subscales in VENUS II). In VENUS I, only Symptom Severity was significantly associated with the number of bleeding days; however, the association was weak ( $r_s = 0.16$ ;  $p < 0.05$ ). In VENUS II, most scales were weakly, but significantly, correlated with the number of bleeding days; the Energy/Mood and Self-Consciousness subscales did not have significant correlations.

After 12 weeks of treatment in both studies, there were much stronger correlations (ranging from  $-0.35$  to  $-0.63$ ; all  $p < 0.0001$ ) between all UFS-QOL scales and bleeding diary responses compared to baseline (Table 3). In addition, after 12 weeks of treatment in VENUS II, correlations with the PGI-I were moderate to strong [28], and were significant (all  $p < 0.0001$ ): 0.69 for the Symptom Severity scale and ranging from  $-0.48$  (Sexual Function) to  $-0.70$  (Concern) for the HRQoL subscales (Table 3).

### Known groups validity

#### By bleeding diary responses

At baseline in VENUS I, patients with  $\leq 5$  days of bleeding scored significantly better ( $p < 0.05$ ) on the Symptom Severity scale than those with  $>9$  days of bleeding. However, pairwise comparisons showed no significant differences



between groups on any of the HRQoL subscales, likely due to the small sample sizes in each group. In VENUS II at baseline, patients with  $\leq 5$  days of bleeding scored significantly better than those with  $>9$  days of bleeding on most of the UFS-QOL scales ( $p < 0.05$ ; except for the

**Table 1** Baseline demographics (intent-to-treat population; observed cases approach)

Characteristic	VENUS I ( $n = 157$ )	VENUS II ( $n = 429$ )
Mean (SD) age, years	41.1 (5.4)	41.0 (5.6)
Race, $n$ (%)		
White	46 (29.3)	130 (30.3)
Black	108 (68.8)	287 (66.9)
Asian	2 (1.3)	5 (1.2)
American Indian or Alaska Native	1 (0.6)	NR
Native Hawaiian or Other Pacific Islander	NR	2 (0.5)
Multiple	NR	5 (1.2)
Ethnicity, $n$ (%)		
Hispanic and Latino	14 (8.9)	59 (13.8)
Mean (SD) body mass index, $\text{kg}/\text{m}^2$	31.7 (8.0)	32.2 (7.9)

NR not reported, SD standard deviation

Energy/Mood, Self-Consciousness, and Sexual Function subscales, for which  $p = \text{NS}$ ). Both studies showed similar trends after 12 weeks of treatment. There were significant differences on all scales between patients experiencing  $\leq 5$  days of bleeding versus those experiencing  $>5$  to 9 days (all  $p < 0.01$ ; both studies) and versus those experiencing  $>9$  days ( $p < 0.05$  for VENUS I, except the Self-Consciousness and Sexual Function subscales, for which  $p = \text{NS}$ ;  $p < 0.001$  for VENUS II). In VENUS II, there was also a significant difference between patients who experienced  $>5$  to 9 days of bleeding versus  $>9$  days of bleeding on the Revised Activities and Energy/Mood subscales (both  $p < 0.05$ ) (data not shown).

**By achievement of absence of bleeding and controlled bleeding**

Patients who achieved absence of bleeding and controlled bleeding in both studies scored significantly better ( $p < 0.001$ ) on each UFS-QOL scale than patients who did not achieve those outcomes (Fig. 2). For example, in VENUS II, the mean (SD) Revised Activities subscale score for women who achieved absence of bleeding compared to those who did not was 88.7 (22.7) versus 59.9 (33.5), respectively.

**Table 2** Distributional characteristics of UFS-QOL scale scores at baseline and after 12 weeks of treatment

	VENUS I				VENUS II			
	Baseline (n = 157)		12 weeks (n = 135 <sup>a</sup> )		Baseline (n = 429 <sup>b</sup> )		12 weeks (n = 347)	
UFS-QOL	Mean (SD)	Median (range)	Mean (SD)	Median (range)	Mean (SD)	Median (range)	Mean (SD)	Median (range)
Symptom Severity <sup>c</sup>	62.0 (19.7)	63 (0–100)	30.5 (24.5)	25 (0–100)	65.5 (21.5)	69 (3–100)	33.0 (25.8)	28 (0–100)
Concern <sup>d</sup>	22.9 (22.6)	15 (0–100)	66.1 (35.7)	80 (0–100)	21.0 (21.7)	15 (0–100)	66.1 (35.9)	75 (0–100)
Activities <sup>d</sup>	34.3 (24.6)	32 (0–100)	75.4 (29.3)	89 (0–100)	32.0 (25.7)	29 (0–100)	71.7 (32.2)	86 (0–100)
Revised Activities <sup>d</sup>	30.7 (24.8)	30 (0–100)	74.4 (30.5)	85 (0–100)	29.9 (25.9)	25 (0–100)	71.1 (32.9)	85 (0–100)
Energy/Mood <sup>d</sup>	37.3 (24.7)	32 (0–100)	74.3 (27.5)	82 (4–100)	34.6 (25.0)	32 (0–100)	69.0 (29.6)	75 (0–100)
Control <sup>d</sup>	45.0 (27.0)	45 (0–100)	77.7 (26.9)	85 (0–100)	43.3 (28.8)	40 (0–100)	73.9 (30.4)	85 (0–100)
Self-Consciousness <sup>d</sup>	35.1 (28.9)	33 (0–100)	71.4 (29.6)	83 (0–100)	31.7 (29.4)	25 (0–100)	62.9 (34.1)	67 (0–100)
Sexual Function <sup>d</sup>	40.3 (33.8)	38 (0–100)	72.0 (33.5)	88 (0–100)	37.7 (33.7)	25 (0–100)	66.5 (35.4)	75 (0–100)
HRQoL Total <sup>d</sup>	35.4 (21.4)	35 (0–100)	73.3 (26.9)	81 (4–100)	33.0 (22.3)	29 (0–97)	69.2 (29.6)	78 (0–100)

<sup>a</sup>n = 134 for Symptom Severity, Concern, Energy/Mood, Control, Sexual Function, and HRQoL Total

<sup>b</sup>n = 428 for Activities, Revised Activities, and Self-Consciousness; n = 427 for Concern, Energy/Mood, and Control; n = 426 for Sexual Function and HRQoL Total

<sup>c</sup>Scores range from 0 to 100, higher scores indicate greater symptom severity

<sup>d</sup>Scores range from 0 to 100, higher scores indicate better HRQoL. HRQoL health-related quality of life; SD standard deviation; UFS-QOL Uterine Fibroid Symptom and Health-Related Quality of Life questionnaire

**By PGI-I score**

In VENUS II after 12 weeks of treatment, pairwise comparisons demonstrated that patients who responded “Very much better” on the PGI-I scored significantly better on almost all subscales when compared to each of the other PGI-I response categories. There were no significant differences on any subscale scores between those who responded “No change”, “A little worse”, “Much worse”, or “Very much worse” (data not shown; the latter three of these groups had very small sample sizes).

When PGI-I responses were collapsed into the categories of “Improved”, “No Change”, and “Worsened”, pairwise comparisons demonstrated that patients in the “Improved” group scored better on all UFS-QOL scales versus those in the “No Change” (all *p* < 0.001) and “Worsened”

groups (all *p* < 0.05) (Table 4). Likely due to the small sample size and greater variance in the “Worsened” category, a monotonic pattern in scores across categories was not observed for most subscales.

**CFA**

CFA confirmed the factor structure of the Revised Activities subscale in both studies: the CFIs exceeded 0.90 (0.97 and 0.99 in VENUS I; 0.96 and 0.98 in VENUS II for the original and Revised Activities subscales, respectively). Additionally, SRMR values indicated good model fit (0.05 and 0.03 in VENUS I; 0.03 and 0.07 in VENUS II for the original and Revised Activities subscales, respectively). Both models (in VENUS I and II) had RMSEA values slightly higher than the acceptable value

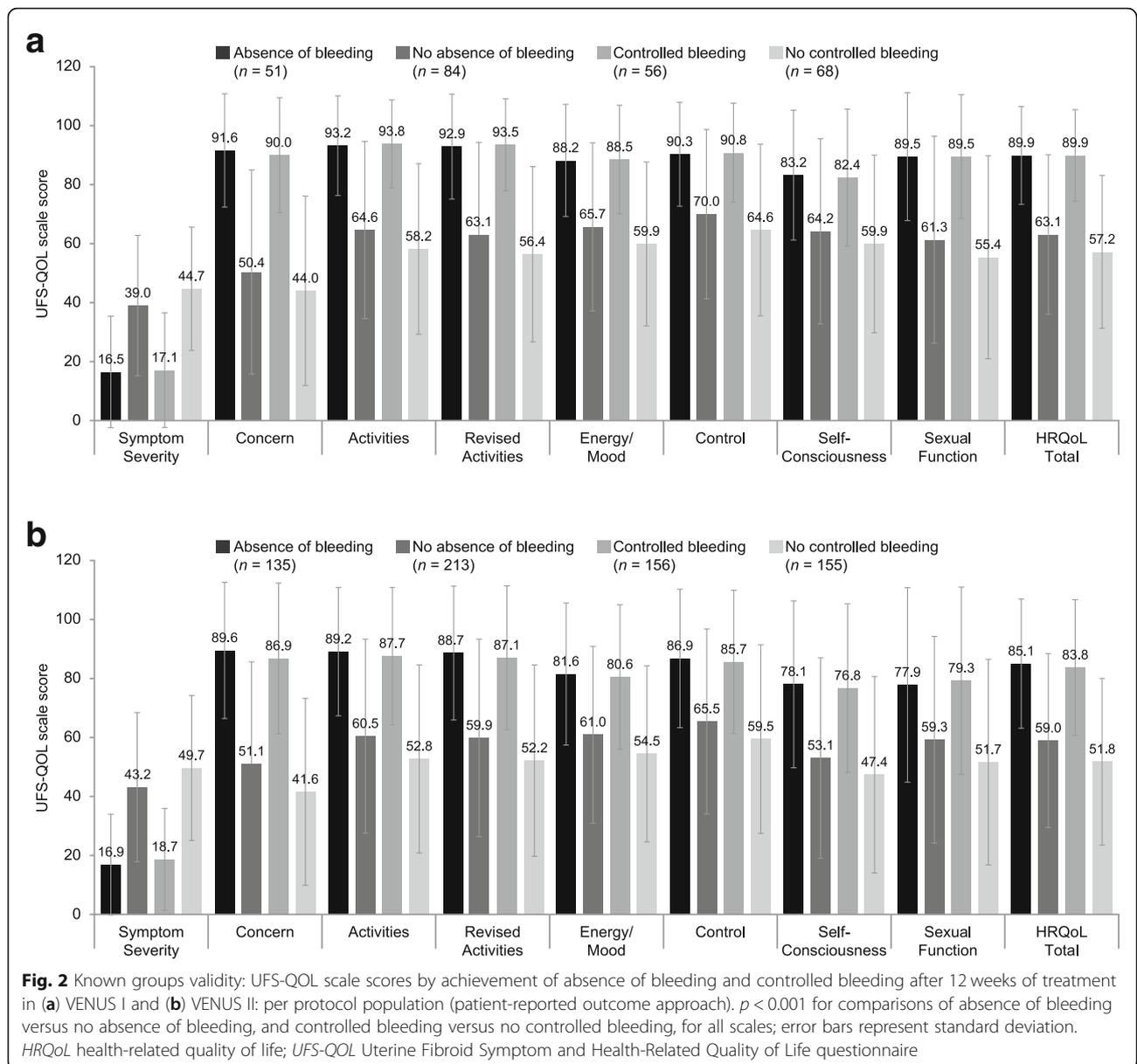
**Table 3** Relationship (Spearman’s correlations) of UFS-QOL scale scores with bleeding diary assessments in VENUS I and VENUS II and PGI-I in VENUS II after 12 weeks of treatment; per protocol population (patient-reported outcome approach)

UFS-QOL	VENUS I		VENUS II		PGI-I
	Bleeding days <sup>a</sup>	Heavy bleeding days <sup>b</sup>	Bleeding days <sup>a</sup>	Heavy bleeding days <sup>b</sup>	
Symptom Severity	0.59	0.56	0.62	0.60	0.69
Concern	−0.61	−0.58	−0.63	−0.63	−0.70
Activities	−0.63	−0.63	−0.54	−0.58	−0.67
Revised Activities	−0.63	−0.63	−0.53	−0.57	−0.66
Energy/Mood	−0.56	−0.58	−0.44	−0.51	−0.57
Control	−0.49	−0.54	−0.46	−0.52	−0.58
Self-Consciousness	−0.43	−0.41	−0.42	−0.47	−0.53
Sexual Function	−0.44	−0.47	−0.35	−0.40	−0.48
HRQoL Total	−0.61	−0.62	−0.53	−0.58	−0.66

All *p* < 0.0001

<sup>a</sup>Number of days in previous 35 days – “bleeding”: days with rating of “bleeding” and “heavy bleeding”

<sup>b</sup>“Heavy bleeding”: days with rating of “heavy bleeding”. HRQoL health-related quality of life; PGI-I Patient Global Impression of Improvement scale; UFS-QOL Uterine Fibroid Symptom and Health-Related Quality of Life questionnaire



of  $<0.07$  (0.09 and 0.07 in VENUS I, both 0.10 in VENUS II); however, the RMSEA tends not to perform well in models with small degrees of freedom, as was the case here [29].

**Discussion**

This psychometric validation study demonstrated that the 1-month recall UFS-QOL, including the Revised Activities subscale, is a valid and reliable PRO measure for the assessment of UF symptoms and their impact on HRQoL. The appropriate recall period for a PRO measure depends on what the measure captures, its intended use, and the attributes of the disease or study [8]. With a longer

recall period, there is a risk of introducing measurement error that may reduce the chances of detecting a treatment effect [8, 30]. Given that the results for the 1-month recall version of the UFS-QOL reported here are strongly consistent with those reported in the initial 3-month recall UFS-QOL validation studies [6, 7, 19], comparison with studies using either version of the instrument is feasible.

In this analysis, the UFS-QOL was shown to detect differences between known outcomes or groups. When comparing UFS-QOL scale scores for patients who achieved absence of, or controlled, bleeding after 12 weeks of treatment versus those who did not, scores were significantly better ( $p < 0.001$ ) for the groups who achieved

**Table 4** Known groups validity: UFS-QOL scale scores by PGI-I response categories after 12 weeks of treatment in VENUS II: per protocol population (patient-reported outcome approach)

	PGI-I						Pairwise comparison <i>p</i> value <sup>d</sup>
	Improved <sup>a</sup>		No change <sup>b</sup>		Worsened <sup>c</sup>		
UFS-QOL	<i>n</i>	Mean (SD)	<i>n</i>	Mean (SD)	<i>n</i>	Mean (SD)	
Symptom Severity	284	26.6 (21.7)	45	59.2 (22.5)	18	68.8 (23.2)	1***, 2***
Concern	285	74.7 (31.8)	44	25.2 (23.8)	18	30.6 (32.5)	1***, 2***
Activities	285	79.8 (27.2)	44	31.8 (26.5)	18	40.3 (27.7)	1***, 2***
Revised Activities	285	79.3 (27.8)	44	30.6 (27.4)	18	39.7 (29.4)	1***, 2***
Energy/Mood	285	75.2 (26.7)	44	39.4 (24.8)	18	43.2 (26.8)	1***, 2***
Control	285	79.9 (26.9)	44	47.8 (30.6)	18	42.2 (31.0)	1***, 2***
Self-Consciousness	285	68.9 (31.8)	44	31.8 (29.5)	18	42.6 (32.9)	1***, 2**
Sexual Function	285	71.3 (33.4)	44	42.6 (36.1)	18	48.6 (37.3)	1***, 2*
HRQoL Total	285	76.1 (26.1)	44	36.0 (22.7)	18	40.5 (25.7)	1***, 2***

<sup>a</sup>Includes responses of "Very much better", "Much better", and "A little better"

<sup>b</sup>Includes response of "No change"

<sup>c</sup>Includes responses of "A little worse", "Much worse", and "Very much worse"

<sup>d</sup>General linear model – pairwise comparisons between means were performed using Scheffe's test adjusting for multiple comparisons: \**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001; 1, "improved" versus "no change"; 2, "improved" versus "worsened". The comparison between "no change" versus "worsened" was not significant for each scale. HRQoL health-related quality of life, PGI-I Patient Global Impression of Improvement scale, SD standard deviation, UFS-QOL Uterine Fibroid Symptom and Health-Related Quality of Life questionnaire

those outcomes. Such discrimination of the UFS-QOL with AUB, one of the most common symptoms of UF [2], is important as it signifies that the UFS-QOL has the ability to differentiate by bleeding status.

The results of the current study corroborate the findings of an earlier validation study of the 4-week recall UFS-QOL using data from a phase IIa proof-of-concept study in 271 pre-menopausal women with heavy bleeding associated with UF [20]. Both studies provide support for the tool as a valid way to measure symptom severity and impact of UF on HRQoL. The current study is strong independently, in that it had a substantial sample size and minimal missing data. The strength of correlations between UFS-QOL scores and the number of bleeding days were weaker at baseline than after 12 weeks of treatment. These findings are comparable with those observed in the previous 4-week recall study, in which correlations between UFS-QOL scales and ratings on the Mansfield-Voda-Jorgensen Menstrual Bleeding Scale were low at baseline (<0.20), but increased after the 3-month treatment period (0.28 to 0.51; *p* < 0.0001) [20]. We would expect correlations between UFS-QOL scores and bleeding diary responses to be greater after 12 weeks of treatment compared to at baseline because the sample has changed with treatment. At the end of 12 weeks of treatment, 37.8% and 38.8% of patients were amenorrheic in VENUS I and II, respectively, with greatly improved UFS-QOL scores from baseline (Table 2); such results are reflected in the correlations between UFS-QOL scales and bleeding diary responses. In contrast, at baseline, there was much greater variability in bleeding days and UFS-QOL responses, resulting in weaker correlations.

These analyses also showed that the Revised Activities subscale performed psychometrically as well as the original Activities subscale. The Revised Activities subscale was created by removing two items ranked lowest in terms of relevancy, based on qualitative focus groups. The modified subscale showed excellent internal consistency reliability and strong model fit based on CFA results, lending support to the use of this shortened scale in future studies. While a limitation of the current study is that it was not designed to include a direct comparison of the 1-month recall UFS-QOL to the 3-month recall version of the UFS-QOL, the 1-month recall UFS-QOL demonstrated similar psychometric properties as the 3-month recall version.

## Conclusion

In conclusion, this study demonstrated that the 1-month recall UFS-QOL, including the Revised Activities subscale, is a valid and reliable PRO measure for the assessment of UF symptoms and their impact on HRQoL.

## Additional file

**Additional file 1: Table S1.** Internal consistency reliability: Cronbach's coefficient alpha values for UFS-QOL scale scores in VENUS I and VENUS II at baseline (intent-to-treat population; observed cases approach) and after 12 weeks of treatment (per protocol population; patient-reported outcome approach) (PDF 250 kb)

## Abbreviations

AUB: Abnormal uterine bleeding; CFA: Confirmatory factor analysis; CFI: Comparative Fit Index; HRQoL: Health-related quality of life; NS: Not significant; PGI-I: Patient Global Impression of Improvement scale; PRO: Patient-reported outcome; RMSEA: Root Mean Square Error of Approximation; SD: Standard deviation; SRMR: Standardized Root Mean

Square Residual; UF: Uterine fibroids; UFS-QOL: Uterine Fibroid Symptom and Health-Related Quality of Life questionnaire; UPA: Ulipristal acetate

#### Acknowledgements

Writing and editorial assistance was provided to the authors by Lindsey O'Mahony, PhD, on behalf of Complete HealthVizion, and Laura Gibbons, PhD, of Complete HealthVizion, and funded by Allergan plc, Dublin, Ireland. All authors met the ICMJE authorship criteria. Neither honoraria nor payments were made for authorship.

#### Authors' contributions

KSC, AH, BMC, PG, and JBS designed the analytic approach. JC analyzed the data. KSC, BMC, and JC reviewed and interpreted the analyses and data. All authors participated in the preparation of the manuscript, and all authors read and approved the final manuscript.

#### Funding

This study was sponsored by Allergan plc, Dublin, Ireland. Allergan (sponsor) played a role in the study design, conduct, analysis, interpretation, writing of the report, and decision to publish the study.

#### Availability of data and materials

Allergan will share de-identified patient-level data and study-level data including protocols and clinical study reports for phase II–IV trials completed after 2008 that are registered to ClinicalTrials.gov or EudraCT, have received regulatory approval in the United States and/or the European Union in a given indication, and the primary manuscript from the trial has been published. To request access to the data, the researcher must sign a data-use agreement and any shared data are to be used for non-commercial purposes. More information can be found on <http://www.allerganclinicaltrials.com/>.

#### Ethics approval and consent to participate

Each institution involved in the original studies, VENUS I and VENUS II, obtained approval from a central or local Institutional Review Board (IRB) before study initiation, and written informed consent was acquired from all study participants. The institutions involved were approved by one of the following IRBs: Shulman Associates IRB, Western IRB, Medical University of South Carolina IRB, Hamilton Integrated Research Ethics Board, Ottawa Health Science Network Research Ethics Board, Chesapeake IRB, and Virtua General IRB.

#### Consent for publication

Not applicable.

#### Competing interests

Karin S. Coyné, Brooke M. Currie, and Jun Chen are employees of Evidera, and received funding from Allergan to conduct this study. Amanda Harrington and Patrick Gillard are employees of Allergan. James B. Spies has nothing to disclose.

#### Author details

<sup>1</sup>Evidera, 7101 Wisconsin Avenue, Suite 1400, Bethesda, MD 20814, USA.

<sup>2</sup>Allergan plc, Irvine, CA, USA. <sup>3</sup>MedStar Georgetown University Hospital, Washington, DC, USA.

Received: 11 January 2019 Accepted: 6 August 2019

Published online: 23 August 2019

#### References

- Baird, D. D., Dunson, D. B., Hill, M. C., Cousins, D., & Schectman, J. M. (2003). High cumulative incidence of uterine leiomyoma in black and white women: Ultrasound evidence. *American Journal of Obstetrics and Gynecology*, *188*(1), 100–107.
- Stewart, E. A., Laughlin-Tommaso, S. K., Catherino, W. H., Lalitkumar, S., Gupta, D., & Vollenhoven, B. (2016). Uterine fibroids. *Nature Reviews. Disease Primers*, *2*, 16043.
- Stewart, E. A. (2001). Uterine fibroids. *Lancet*, *357*(9252), 293–298.
- Ghant, M. S., Sengoba, K. S., Recht, H., Cameron, K. A., Lawson, A. K., & Marsh, E. E. (2015). Beyond the physical: A qualitative assessment of the burden of symptomatic uterine fibroids on women's emotional and psychosocial health. *Journal of Psychosomatic Research*, *78*(5), 499–503.
- Borah, B. J., Nicholson, W. K., Bradley, L., & Stewart, E. A. (2013). The impact of uterine leiomyomas: a national survey of affected women. *American Journal of Obstetrics and Gynecology*, *209*(4), 319.e1–319.e20.
- Coyné, K. S., Margolis, M. K., Bradley, L. D., Guido, R., Maxwell, G. L., & Spies, J. B. (2012). Further validation of the uterine fibroid symptom and quality-of-life questionnaire. *Value in Health*, *15*(1), 135–142.
- Spies, J. B., Coyné, K., Guaou, N., Boyle, D., Skyrnarz-Murphy, K., & Gonzalves, S. M. (2002). The UFS-QOL, a new disease-specific symptom and health-related quality of life questionnaire for leiomyomata. *Obstetrics & Gynecology*, *99*(2), 290–300.
- Norquist, J. M., Girman, C., Fehnel, S., DeMuro-Mercon, C., & Santanello, N. (2012). Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. *Quality of Life Research*, *21*(6), 1013–1020.
- Horak, P., Mara, M., Dundr, P., et al. (2012). Effect of a selective progesterone receptor modulator on induction of apoptosis in uterine fibroids in vivo. *International Journal of Endocrinology*, *2012*, 436174.
- Biglia, N., Carinelli, S., Maiorana, A., D'Alonzo, M., Lo Monte, G., & Marci, R. (2014). Ulipristal acetate: a novel pharmacological approach for the treatment of uterine fibroids. *Drug Design, Development and Therapy*, *8*, 285–292.
- Donnez, J., Tomaszewski, J., Vázquez, F., et al. (2012). Ulipristal acetate versus leuprolide acetate for uterine fibroids. *New England Journal of Medicine*, *366*(5), 421–432.
- Donnez, J., Tatarчук, T. F., Bouchard, P., et al. (2012). Ulipristal acetate versus placebo for fibroid treatment before surgery. *New England Journal of Medicine*, *366*(5), 409–420.
- Donnez, J., Vázquez, F., Tomaszewski, J., et al. (2014). Long-term treatment of uterine fibroids with ulipristal acetate. *Fertility and Sterility*, *101*(6), 1565–1573 e1–e18.
- Donnez, J., Hudeček, R., Donnez, O., et al. (2015). Efficacy and safety of repeated use of ulipristal acetate in uterine fibroids. *Fertility and sterility*, *103*(2), 519–527.e3.
- Donnez, J., Donnez, O., Matule, D., et al. (2016). Long-term medical management of uterine fibroids with ulipristal acetate. *Fertility and sterility*, *105*(1), 165–173.e4.
- Liu, J. H., Soper, D., Lukes, A., et al. (2018). Ulipristal acetate for treatment of uterine leiomyomas. A randomized controlled trial. *Obstetrics & Gynecology*, *132*(5), 1241–1251.
- Simon, J. A., Catherino, W., Segars, J. H., et al. (2018). Ulipristal acetate for treatment of symptomatic uterine leiomyomas. A randomized controlled trial. *Obstetrics & Gynecology*, *131*(3), 431–439.
- Calvert, M., Blazeyby, J., Altman, D. G., et al. (2013). Reporting of patient-reported outcomes in randomized trials. The CONSORT PRO extension. *JAMA*, *309*(8), 814–822.
- Harding, G., Coyné, K. S., Thompson, C. L., & Spies, J. B. (2008). The responsiveness of the uterine fibroid symptom and health-related quality of life questionnaire (UFS-QOL). *Health and Quality of Life Outcomes*, *6*, 99.
- Coyné, K. S., Soliman, A. M., Margolis, M. K., Thompson, C. L., & Chwalisz, K. (2017). Validation of the 4 week recall version of the Uterine Fibroid Symptom and Health-Related Quality of Life (UFS-QOL) questionnaire. *Current Medical Research and Opinion*, *33*(2), 193–200.
- Frost, M. H., Reeve, B. B., Liepa, A. M., Stauffer, J. W., Hays, R. D., & the Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. (2007). What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value in Health*, *10*(Suppl 2), S94–S105.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.
- Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issues, and applications* (pp. 76–99). Thousand Oaks: Sage Publications.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park: Sage Publications.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1–55.
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus. Statistical Analysis With Latent Variables. User's Guide* (7th ed.) [https://www.statmodel.com/download/usersguide/MplusUserGuideVer\\_7.pdf](https://www.statmodel.com/download/usersguide/MplusUserGuideVer_7.pdf). Accessed 30 May 2019.
- SAS. (2013). SAS® 9.4. [https://www.sas.com/en\\_us/software/sas9.html](https://www.sas.com/en_us/software/sas9.html). Accessed 30 May 2019.

28. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.
29. Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507.
30. Stull, D. E., Leidy, N. K., Parasuraman, B., & Chassany, O. (2009). Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. *Current Medical Research and Opinion*, 25(4), 929–942.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---