Psychometric evaluation of the HFDD, PROMIS SD SF 8b and MENQOL for assessing vasomotor symptoms caused by endocrine therapy: reliability, validity, and differential item functioning

C0193

Trigg A.,[1] Seitz C.,[2,3] Gerlinger C.,[2,4] Bradley H.,[5] Barclay M.,[5] Gater A.,[5] & **Haberland, C**[2]

[1]Bayer Plc, Reading, United Kingdom, [2]Bayer AG, Berlin, Germany, [3]Institute of Clinical Pharmacology and Toxicology, Charité Universitätsmedizin Berlin, Germany, [4]University of Saarland, Homburg/Saar, Germany, [5]Adelphi Values Ltd, Bollington, United Kingdom.

OBJECTIVES

Endocrine therapy (ET) is routinely prescribed for women with breast cancer and may be offered to those at high-risk for breast cancer. Vasomotor symptoms (VMS), also known as hot flashes (HFs) or night sweats if occurring at night, are a prominent side effect and a recognized contributor to nonadherence and treatment discontinuation.[1-3] Sleep disturbances are also frequently reported in this population, with both causing substantial impacts on women's quality of life.[1,4,5]

This study evaluated the psychometric measurement properties of scores from the following patient-reported outcome (PRO) measures to confirm their suitability for use in women experiencing VMS caused by ET: the **Hot Flash Daily Diary (HFDD)** as a measure of HF frequency and severity; the **Patient-Reported Outcomes Measurement Information System Sleep Disturbance Short Form 8b (PROMIS SD SF 8b)** as a measure of sleep disturbance; and the **Menopause-Specific Quality of Life (MENQOL)** as a measure of menopause-specific quality of life (**Table 1**).

METHODS

Data from a double-blind, randomized, placebo-controlled, multicenter Phase 3 trial evaluating the efficacy and safety of elinzanetant for the treatment of moderate to severe VMS caused by ET (OASIS 4; NCT05587296) were used to assess psychometric measurement properties. Data was pooled across both treatment arms.

HFDD data collected at Baseline and throughout the first 12 weeks of the double-blind treatment period, and PROMIS SD SF 8b and MENQOL data collected at Baseline and Weeks 4, 8, and 12, were subject to psychometric analyses (**Table 1**). All data was collected using an electronic hand-held device.

Reference measures used to support evaluation of the HFDD, PROMIS SD SF 8b and MENQOL scores included the EQ-5D-5L,[6] Insomnia Severity Index (ISI),[7] and Beck Depression Inventory (BDI-II).[8] Scores from the HFDD and MENQOL were used as additional reference measures for certain analyses as appropriate.

Psychometric analyses are described in **Table 2**.

Table 1. PRO measures included in psychometric analyses

PRO	Description
HFDD *(HF frequency and HF severity scores)*	• An e-diary completed twice daily (morning and evening) to assess the number of mild, moderate and severe HFs experienced. Psychometric analyses conducted for the HFDD frequency of moderate to severe HF weekly average (HF frequency score) and the HFDD severity of HF weekly average scores (HF severity score). • HF frequency score ranges from 0 to no prespecified upper limit (as participants could record any number of episodes); HF severity score ranges from 0 to 3. Higher HF frequency and severity scores indicate a greater number of HFs and greater severity (respectively).
PROMIS SD SF 8b[9] *(PROMIS SD SF 8b T-score)*	• 8 items assessing different aspects of sleep disturbances over the past 7 days. • PROMIS SD SF 8b T-score ranges from 28.9 to 76.5, higher scores indicate greater sleep disturbance.
MENQOL[10] *(MENQOL Total score)*	• 29 items assessing the presence and bothersomeness of menopausal symptoms and impacts over the past week. • MENQOL Total score ranges from 1 to 8, higher scores indicate greater bother.

Table 2. Overview of key psychometric analyses

Analysis		Description
ePRO compliance		• Assesses the proportion of completed target PROs out of participants still in the study, at Baseline and each available week over the first 12 weeks of the study.
Distributional properties		• Provides floor and ceiling effects for the target PRO scores (≥15% of participants with a score indicating the worst possible health or best possible health).[11] Note, only ceiling effects were relevant for HFDD HF frequency score as there was no upper bound for the number of HFs a participant could record.
Reliability	**Test-retest reliability**	• Informs about the stability of scores between Weeks 8 and 12 for HFDD scores, where participants were defined as stable based on their MENQOL VMS Domain score. Intraclass correlation coefficient (ICC) with 95% confidence interval (CI) interpreted as: 0.75 to 0.90 (good), >0.90 (excellent).[12]
	Composite reliability	• Provides test information function for the PROMIS SD SF 8b T-score, and omega coefficients for the MENQOL VMS, Psychosocial and Sexual Domains, at Week 8 (≥0.7 indicates adequate composite reliability).[13] Note, as the MENQOL Physical Domain is a composite indicator model,[14] composite reliability was inappropriate for the MENQOL Physical Domain and Total score (which comprises it).
Validity	**Inter-item correlations**	• Examines item redundancy (indicated by r≥0.90)[15] for the PROMIS SD SF 8b T-score (polychoric) and MENQOL Total score (Spearman's) at Week 8.
	Confirmatory factor analysis (CFA)	• Examines the MENQOL Total score at Week 8 using CFA; a second-order model was specified as per previous analyses.[14,16,17] Items in the Physical Domain treated as composite indicators; all others treated as reflective. Point estimates >0.4 were considered supportive.[18]
	Item response theory (IRT)	• Examines the PROMIS SD SF 8b T-score at Week 8 using a unidimensional graded response model. Discrimination parameters above 1 and absence of significant local dependence deemed acceptable.[19]
	Convergent and divergent evidence	• Compares spearman or polyserial correlations for all target PRO scores at Week 8 to *a priori* hypotheses (**Table 3**). Considered adequate if >75% of hypotheses were correct.[11,20]
	Known-groups evidence	• Compares scores between ISI subgroups considered clinically distinct (0-7, no clinically significant insomnia; 8-14, subthreshold insomnia; 15-21, moderate clinical insomnia; 22-28, severe clinical insomnia)[7] for the PROMIS SD SF 8b T-score at Week 8.
Differential item functioning (DIF)		• Examines the PROMIS SD SF 8b and MENQOL Total score at Week 8 using ordinal logistic regression models according to VMS cause - ET versus menopause - by pooling OASIS 4 and OASIS 2 data (a Phase 3 study of elinzanetant in VMS associated with menopause; n=359, NCT05099159).

RESULTS

Data from n=474 women participating in the OASIS 4 trial was used for the psychometric analyses; mean (SD) age = 51.0 (7.3) years.

ePRO compliance

Most participants in the study were compliant from the Baseline visit (≥98.3%) up to and including Week 12 (≥97.3%).

Distributional properties

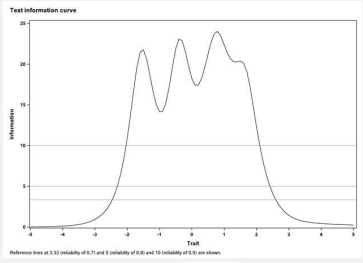
No floor or ceiling effects were observed for the scores at Baseline, Weeks 4, 8 or 12.

Reliability

Test-retest reliability. ICCs were 0.935 (95% CI 0.919, 0.948) and 0.869 (95% CI 0.837, 0.895) for the HFDD HF frequency and HF severity scores (respectively), indicating good to excellent test-retest reliability.

Composite reliability. PROMIS SD SF 8b T-score reliability exceeded 0.9 across the majority of the latent trait (i.e., -2 to +2; **Figure 1**). For the MENQOL VMS, Psychosocial and Sexual domain scores, each exhibited acceptable internal consistency with omega coefficients of 0.837 (95% CI 0.801, 0.867), 0.821 (95% CI 0.790, 0.847) and 0.780 (95% CI 0.740, 0.815) respectively.

Figure 1. Test information curve for PROMIS SD SF 8b

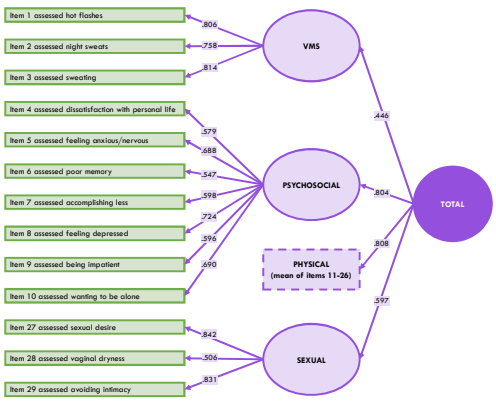


Validity

Inter-item correlations. No evidence of item redundancy within the PROMIS SD SF 8b or MENQOL, as no item pairs correlated at r≥0.90.

Confirmatory factor analysis. First-order loadings ranged from 0.506 to 0.842, supporting MENQOL items as indicators of each domain (**Figure 2**). Second-order loadings were all >0.4 (range: 0.446 to 0.808). Adequate fit was demonstrated by comparative fit index (CFI) = 0.914, root mean square error of approximation (RMSEA) = 0.077 and standardized root mean square residual (SRMR) = 0.062.

Figure 2. CFA path diagram for MENQOL



DIF. Using ordinal logistic regression, no uniform or non-uniform DIF according to VMS indication (i.e., VMS caused by ET versus VMS associated with menopause) was observed for the PROMIS SD SF 8b or MENQOL.

IRT. Item discrimination parameters for the PROMIS SD SF 8b were all >1, difficulty thresholds were ordered, and no significant local dependence was detected.

Convergent/divergent evidence. Supportive convergent and divergent evidence was obtained, with all but two correlations consistent with pre-specified hypotheses (**Table 3**).

Table 3. Convergent and divergent correlations at Week 8

Convergent / divergent score	Hypothesis	Estimate
HFDD frequency		
MENQOL VMS Domain	>0.5	0.669
MENQOL Total	>0.4	0.433
MENQOL Sexual Domain	<0.3	0.128
HFDD severity		
MENQOL VMS Domain	>0.4	0.583
MENQOL Total	>0.3	0.375
MENQOL Sexual Domain	<0.3	0.095
PROMIS SD SF 8b (T-score)		
ISI Total	>0.4	0.866
EQ-5D-5L Self-care	<0.3	0.181
HFDD Frequency of NTAs	>0.5	0.548
HFDD Sleep disturbance	>0.5	0.473*
MENQOL Total score		
HFDD Frequency	>0.4	0.433
HFDD Severity	>0.3	0.375
BDI-II Total	>0.4	0.643
EQ-5D-5L VAS	>0.3	-0.360
EQ-5D-5L self-care	<0.3	0.340*

Note: HFDD Frequency of night-time awakenings (NTAs), HFDD Sleep disturbance, MENQOL VMS domain and MENQOL Sexual domain scores were utilized as additional reference measures to evaluate convergent/divergent validity of the target PRO scores. *Hypothesis not met.

Known-groups evidence. For the PROMIS SD SF 8b, differences between clinically distinct ISI subgroups and large between-group effect sizes (ranging from 1.80 to 3.79) were observed.

CONCLUSIONS

Consistent with previous results in women experiencing VMS associated with natural or surgical menopause,[16,17,21] findings provide evidence that the HFDD, PROMIS SD SF 8b and MENQOL yield valid and reliable scores, supporting their application in assessing clinical trial endpoints in women experiencing VMS caused by ET.

Scan the QR code to view the reference list and online poster.

Disclosures & Acknowledgements: *This study was funded by Bayer. AT is a Bayer plc employee. CS, CG and CH are Bayer AG employees. MB, HB, and AG are Adelphi Values employees. Medical writing assistance was provided by MB, HB, and AG, supported by Bayer in accordance with GPP.*