# Validation and Application of Thresholds to Define Meaningful Change in Vasomotor Symptoms Frequency: Analysis of Pooled SKYLIGHT 1 and 2 Data

Antonia Morga 🔟 · Lisa Zimmermann · Udaya Valluri ·
Emad Siddiqui · Lori McLeod · Randall H. Bender

## ABSTRACT

***Introduction***: Vasomotor symptoms (VMS), the characteristic symptoms of menopausal transition, are often the primary reason women seek treatment. Current treatment options for VMS include fezolinetant, a nonhormonal, selective neurokinin 3 receptor antagonist. This study aimed to define a clinically meaningful threshold for reduction of moderate-to-severe VMS in postmenopausal women treated with fezolinetant and then apply it in a responder analysis of the pooled trial data.
***Methods***: This analysis pooled data from two identical phase 3, double-blind, placebo-controlled studies that randomized women with moderate-to-severe VMS to once-daily fezolinetant 30 mg, 45 mg, or placebo (SKYLIGHT 1 and 2). The frequency of VMS was collected daily using an electronic diary. Patients completed the Patient Global Impression of Change in VMS (PGI-C VMS) instrument, which assessed changes in hot flushes/night sweats at weeks 4 and 12 compared with baseline using a seven-point Likert scale. VMS frequency data were anchored to PGI-C VMS data; the anchor level for meaningful within-patient change in PGI-C VMS was "moderately better."
***Results***: In the pooled population ($N = 1022$), the mean (standard deviation) estimated thresholds for a meaningful within-patient change in moderate-to-severe VMS frequency were $-5.73$ (3.47) at week 4 and $-6.20$ (5.18) at week 12. Applying the thresholds for meaningful within-patient change to responder analyses ("missing as non-responder" imputation method) indicated a favorable clinical benefit: greater proportions of responders were observed in the fezolinetant 30-mg and 45-mg groups compared with placebo at week 4 (odds ratio range 2.48–2.91; $P < 0.001$) and week 12 (odds ratio range 1.908–2.68; $P < 0.001$).
***Conclusion***: PGI-C VMS is sensitive to change and correlates with VMS frequency: a reduction of approximately six VMS episodes per day from baseline to week 12 was meaningful at the individual patient level. Fezolinetant provides a meaningful clinical benefit for women with moderate-to-severe VMS associated with menopause and represents an important nonhormonal treatment option.

A. Morga (✉) · E. Siddiqui
Astellas Pharma Europe Ltd., 300 Dashwood Lang Road, Bourne Business Park, Addlestone KT15 2NX, UK
e-mail: antonia_morga@astellas.com

L. Zimmermann
Astellas Pharma GmbH Deutschland, Munich, Germany

U. Valluri
Astellas Pharma Global Development, Inc., Northbrook, IL, USA

L. McLeod · R. H. Bender
RTI Health Solutions, North Carolina, NC, USA

---

### Key Summary Points

*Why carry out this study?*

Two phase 3 clinical trials, SKYLIGHT 1 and 2, demonstrated that fezolinetant 30 mg and 45 mg once-daily were superior to placebo in improving the frequency and severity of vasomotor symptoms (VMS) at weeks 4 and 12 in participants with moderate-to-severe VMS associated with menopause.

It is important to understand what magnitude of treatment effect is clinically meaningful to patients to help guide the interpretation of these results.

Using pooled data from SKYLIGHT 1 and 2 and an anchor-based method, this study aimed to define a clinically meaningful threshold for a reduction of moderate-to-severe VMS in postmenopausal women treated with fezolinetant.

*What was learned from the study?*

The thresholds for a meaningful within-patient change in moderate-to-severe VMS frequency were estimated to be a reduction of 5.73 and 6.20 VMS episodes/day at weeks 4 and 12, respectively.

When these thresholds were applied, significantly greater proportions of responders were observed in participants treated with fezolinetant compared with placebo at both weeks 4 and 12.

This analysis indicates fezolinetant is an important nonhormonal treatment option that provides clinically meaningful improvements in symptoms for women who have VMS associated with menopause.

---

## INTRODUCTION

Vasomotor symptoms (VMS), which are caused by an imbalance of hypothalamic thermoregulation, are commonly referred to as hot flashes and/or night sweats [1–3]. VMS episodes are the characteristic symptoms of menopausal transition; they affect up to 80% of women and are considered moderate-to-severe by up to 50% of women [2, 4, 5]. VMS are bothersome for the majority of women, and they have a negative impact on health-related quality of life (HRQoL), including impacts on sleep, concentration/cognition, mood, energy, sexual activity, work, and social/leisure activities [6, 7].

VMS are reported across all races/ethnicities [5, 8], and they persist for a median duration of 7.4 years but may continue for more than 10 years in one-third of women [4, 9]. VMS are the most common primary reason for women to seek treatment for menopausal symptoms [8]. Hormone therapy remains an effective treatment [10, 11], but it is unsuitable for some women because of contraindications, and a large proportion of women are eligible but averse to taking hormone therapy (e.g., due to beliefs that menopausal symptoms will diminish without pharmacological intervention and side effects/long-term risks) [11, 12].

Fezolinetant is a nonhormonal, selective neurokinin 3 receptor antagonist that moderates neuronal activity in the thermoregulatory center in the hypothalamus [13]. It was recently approved at a once-daily dose of 45 mg by the US Food and Drug Administration (FDA) for the treatment of moderate-to-severe VMS due to menopause and by the European Medicines Agency for the treatment of moderate-to-severe VMS associated with menopause [14]. SKYLIGHT 1 and SKYLIGHT 2, two identical phase 3 trials, demonstrated that fezolinetant 30 mg and 45 mg provided a statistically significant improvement in mean daily VMS frequency and VMS severity at weeks 4 and 12 (the four co-primary endpoints) compared with placebo in participants with moderate-to-severe VMS associated with menopause [15, 16]. SKYLIGHT 4, a 52-week, randomized, phase 3

study, confirmed the safety and tolerability of fezolinetant in this patient population [17].

Patient-reported outcomes (PROs) report the status of a patient's health condition directly from the patient, without interpretation by a clinician or anyone else [18]. PROs help determine the effects of treatment, directly reflecting how patients feel or function [19]. Importantly, PROs can be used to estimate improvements that are clinically meaningful from the patient's own perspective, thus providing insights for the interpretation of the estimates of the overall treatment effect of any given treatment [20].

The objective of this analysis was to define the threshold for meaningful within-patient change to guide the interpretation of results for the reduction in VMS frequency. The paper further applies this threshold, with results that support the efficacy of fezolinetant using the primary endpoints reported elsewhere [15, 16]. These analyses were pre-specified and used pooled data from SKYLIGHT 1 and SKYLIGHT 2.

# METHODS

## Study Design and Participants

The study design of SKYLIGHT 1 (NCT04003155) and SKYLIGHT 2 (NCT04003142) has been described previously [15, 16]. In brief, the two trials were phase 3, randomized, double-blind, placebo-controlled studies. Individuals aged $\geq 40$ to $\leq 65$ years who were female at birth and who had moderate-to-severe VMS (minimum average of seven VMS/day or $\geq 50$ per week) were randomized to once-daily fezolinetant 30 mg, fezolinetant 45 mg, or placebo (1:1:1) for 12 weeks followed by a 40-week extension period of active treatment (individuals initially randomized to placebo were re-randomized to fezolinetant 30 mg or 45 mg). The SKYLIGHT 1 and SKYLIGHT 2 studies were conducted in accordance with the Declaration of Helsinki, Good Clinical Practice, and International Council for Harmonisation guidelines. An independent ethics committee or institutional review board reviewed the ethical, scientific, and medical appropriateness of the study at each site before data collection.

Written informed consent was provided by all participants.

## Evaluation of Meaningful Within-Patient Change

The four co-primary endpoints of both studies were mean change in daily frequency and severity of moderate-to-severe VMS from baseline to weeks 4 and 12 [15, 16]. VMS data were collected using an electronic VMS diary that was completed by participants daily during a 24-h period; this started at screening and continued through to the follow-up visit. The VMS diary, an interactive, electronic data capture system available for data entry 24 h/day, included a reference guide with the following definitions: mild symptoms (i.e., sensation of heat without sweating), moderate symptoms (i.e., sensation of heat with sweating, able to continue activity), and severe symptoms (i.e., sensation of heat with sweating, causing cessation of activity) [21].

Meaningful within-patient change for fezolinetant treatment was assessed using the Patient Global Impression of Change in VMS (PGI-C VMS) instrument, which was a pre-specified secondary endpoint of the trials and was used as an anchor measure in this analysis. PGI-C VMS is a single-item global PRO, analogous to the Clinical Global Impression (CGI) scales [22], designed to provide a patient's assessment of change in VMS from the start of treatment. The PGI-C VMS asked the following question: "Compared to the beginning of this study, how would you rate your hot flushes/ night sweats now?" Participants rated change using a seven-point Likert scale: "much better," "moderately better," "a little better," "no change," "a little worse," "moderately worse," and "much worse." Patient responses for PGI-C VMS were collected at weeks 4 and 12. The "moderately better" response category was selected to characterize a meaningful change in PGI-C VMS in our analysis. There is evidence supporting this conservative approach from prior VMS studies that employed similar CGI scales [23, 24], in which clinical meaningfulness in VMS frequency/severity reflected changes

associated with improvements above the "minimally improved" or "a little better" category by using the two highest ranked categories in the seven-point scale. These two studies support the choice of primary anchor in our analysis, and the week 12 time point was designated as the key time point.

### Statistical Analyses

Analyses of meaningful within-patient change were performed using all randomized participants in SKYLIGHT 1 and 2 who had VMS frequency data at baseline and at least one post-baseline value at either week 4 or week 12. The primary analysis used the overall pooled population, and sensitivity analyses used subgroups of the overall pooled sample and the individual phase 3 studies (Table S1).

The number and proportion of participants for each PGI-C VMS response category were summarized at weeks 4 and 12; differences between the fezolinetant and placebo groups were estimated using the Cochran–Mantel–Haenszel test with modified ridit scores stratified by study. In addition, descriptive statistics (mean, standard deviation [SD], and median) were reported for moderate-to-severe VMS frequency at baseline, week 4, and week 12, and changes from baseline.

The appropriateness of the PGI-C VMS to serve as an anchor measure for the change in the frequency of VMS was assessed by conducting correlational analyses (polyserial and Spearman's rank) and reviewing select descriptive statistics to ensure that it was adequately related to the change in frequency of moderate-to-severe VMS at weeks 4 and 12. The subsequent meaningful within-patient change analyses would be performed only if the association between the PGI-C VMS and the change in moderate-to-severe VMS frequency was deemed to be appropriate and had a correlation value of > 0.37; this threshold was based on criteria recommended in the literature for appropriate anchor measures [25, 26].

After PGI-C VMS was deemed an appropriate anchor measure, thresholds of meaningful within-patient change in moderate-to-severe VMS frequency were estimated. The primary anchor-based estimates were summarized descriptively by mean change with SD (and median, with first and third quartiles as supportive estimates) in the frequency of moderate-to-severe VMS for the different levels of change defined by the PGI-C VMS. Estimates were determined using data for week 4 and week 12 separately. In addition, sensitivity analyses were conducted using several supportive analyses, including the distribution-based method of half-SD of baseline, receiver operating characteristic (ROC) curve analysis, and empirical cumulative distribution functions (eCDF) plots of the primary anchor measure. The ROC-based analysis applied a slightly lower level of response, defined as the change in score that differentiates the "moderately better" response and the "a little better" response on the PGI-C VMS. eCDF curves for the change in the frequency of moderate-to-severe VMS by PGI-C VMS response groups (anchor categories) at weeks 4 and 12 are presented.

Subsequent responder analyses of moderate-to-severe VMS frequency were performed that applied the primary threshold estimates of meaningful within-patient change associated with a PGI-C VMS "moderately better" response at week 4 and week 12 to compare fezolinetant 30 mg and fezolinetant 45 mg with placebo. A patient was classified as a responder if the change in VMS frequency from baseline to week 4 was equal to or larger than the meaningful within-patient change threshold at week 4. A similar definition was used to classify responders at week 12. Odds ratios (ORs), 95% confidence intervals (CIs), and P values were from a logistic regression model that used protocol, treatment group, and smoking status (current versus former/never) as factors and baseline frequency of VMS as a covariate.

## RESULTS

### Patient Demographics

The pooled population of SKYLIGHT 1 and 2 consisted of 1022 women who were randomized and received at least one dose of study drug

(placebo $n = 342$, fezolinetant 30 mg $n = 339$, fezolinetant 45 mg $n = 341$) (Table 1). These women had a mean (SD) age of 54.3 (5.0) years, and most were White (81.1%). Demographics were generally balanced across groups, although mean time since onset of hot flashes was slightly longer in the placebo group (81.9 months) compared with the fezolinetant 30-mg (76.7 months) and 45-mg (76.9 months) groups.

## VMS Frequency and PGI-C VMS

In the overall pooled population, the mean (SD) number of moderate-to-severe VMS episodes/day was 11.02 (5.30) at baseline and reduced to 6.22 (5.60) and 5.23 (5.40) at weeks 4 and 12, respectively, which corresponds to mean (SD) changes from baseline of − 4.86 (4.64) and − 5.82 (5.40) at weeks 4 and 12, respectively (Table 2). Similar data were reported in the pooled sub-sample

**Table 1** Key participant demographics and baseline characteristics (full analysis set)

| Parameter | Placebo ($n = 342$) | Fezolinetant 30 mg ($n = 339$) | Fezolinetant 45 mg ($n = 341$) | Total ($N = 1022$) |
|---|---|---|---|---|
| Ethnicity,[a] $n$ (%) | | | | |
|   Not Hispanic or Latina | 262 (77.1) | 263 (77.6) | 252 (73.9) | 777 (76.2) |
|   Hispanic or Latina | 78 (22.9) | 76 (22.4) | 89 (26.1) | 243 (23.8) |
| Race,[b] $n$ (%) | | | | |
|   American Indian or Alaska Native, Asian,[c] other Pacific Islander, other[d] | 7 (2.0) | 4 (1.2) | 8 (2.3) | 19 (1.9) |
|   Black or African American | 59 (17.3) | 56 (16.6) | 59 (17.3) | 174 (17.0) |
|   White | 276 (80.7) | 278 (82.2) | 274 (80.4) | 828 (81.1) |
| Age, mean (SD), years | 54.7 (4.7) | 54.0 (4.9) | 54.3 (5.3) | 54.3 (5.0) |
| Weight, mean (range), kg | 74.49 (46.2–125.0) | 75.18 (42.0–121.2) | 75.17 (45.0–110.6) | 74.95 (42.0–125.0) |
| BMI,[e] mean (range), kg/m$^2$ | 28.17 (18.6–38.0) | 28.02 (18.0–37.8) | 28.12 (18.0–37.9) | 28.10 (18.0–38.0) |
| Current smoker, $n$ (%) | 57 (16.7) | 55 (16.2) | 57 (16.7) | 169 (16.5) |
| Time since onset of VMS, mean (range), months | 81.9 (2–422) | 76.7 (3–370) | 76.9 (1–396) | 78.5 (1–422) |
| Amenorrhea, $n$ (%) | 329 (96.2) | 332 (97.9) | 334 (97.9) | 995 (97.4) |
| Hysterectomy, $n$ (%) | 102 (29.8) | 113 (33.3) | 114 (33.4) | 329 (32.2) |
| Oophorectomy, $n$ (%) | 75 (21.9) | 70 (20.6) | 76 (22.3) | 221 (21.6) |

Data shown for all participants who were randomized and received at least one dose of study intervention (full analysis set)

*BMI* body mass index, *SD* standard deviation, *VMS* vasomotor symptoms

[a]Data on ethnicity were missing for two participants in the placebo (and total) group

[b]Data on race were missing for one participant in the fezolinetant 30-mg (and total) group

[c]Chinese, Filipino, Japanese, Korean, Vietnamese, other Asian

[d]More than one race

[e]Data on BMI were missing for one participant in the fezolinetant 45-mg (and total) group

populations (Table 2) and separately in SKYLIGHT 1 and SKYLIGHT 2 (Table S2).

In the overall pooled population, greater proportions of participants in the fezolinetant 30-mg and 45-mg groups relative to placebo reported an improvement ("much better," "moderately better," or "a little better") in PGI-C VMS at weeks 4 and 12 compared with baseline (Table 3). The proportions of participants with an improvement in PGI-C VMS were 81.6% in the fezolinetant 30-mg group and 85.6% in the fezolinetant 45-mg group versus 61.4% in the placebo group at week 4, and 84.7% in the fezolinetant 30-mg group and 91.1% in the fezolinetant 45-mg group versus 66.2% in the placebo group at week 12. The association between response and treatment group (fezolinetant 30-mg and 45-mg groups versus placebo) had a $P$ value of $< 0.001$ at weeks 4 and 12. Similar data were reported in the individual phase 3 studies (Tables S3 and S4).

### Meaningful Within-Patient Change

Correlations of change scores between moderate-to-severe VMS frequency and PGI-C VMS exceeded the minimum magnitude of correlation criterion of $> 0.37$ in the overall pooled study sample; polyserial values were 0.55 at week 4 and 0.48 at week 12, and Spearman's rank values were 0.59 at week 4 and 0.52 at week 12 (Table 4). At study level, the polyserial correlations in SKYLIGHT 1 and SKYLIGHT 2 were 0.55 and 0.56 at week 4 and 0.46 and 0.49 at week 12, respectively (Table S5).

In the overall sample, using the primary anchor of "moderately better" improvement in PGI-C VMS, the mean (SD) and median (first quartile, third quartile) estimated thresholds for reporting a meaningful within-patient change in moderate-to-severe VMS frequency were $-5.73$ (3.47) and $-5.79$ ($-7.44$, $-3.26$) respectively at week 4 and $-6.20$ (5.18) and $-6.28$ ($-8.29$, $-4.13$) respectively at week 12 (Table 5).

The mean and median estimates were similar at both time points, and the first and third quartiles around the median had relative symmetry, which indicates that there was no extreme skewness in the data impacting the mean estimates. Also, the anchor-based estimates were slightly larger at week 12 than week 4 (Fig. 1). Mean and median anchor-based threshold estimates were larger than the ROC estimates at weeks 4 and 12, which were larger than the half-SD estimate. At study level, mean and median estimated thresholds from SKYLIGHT 1 and SKYLIGHT 2 are reported in Tables S6 and S7, respectively; SKYLIGHT 2 had slightly higher baseline mean VMS counts and variability than SKYLIGHT 1. The eCDF curves of the primary anchor measure response categories by VMS change scores at weeks 4 and 12 show well-spaced and well-ordered patterns for the three improvement and "no change" response groups and are supportive of the anchor-based method and its estimates (Fig. S1).

**Table 2** Moderate-to-severe VMS frequency at baseline, week 4, and week 12, and change from baseline to week 4 and week 12 in the pooled population

| Study/sample | VMS frequency: mean (SD), median, $n$ | | | | |
|---|---|---|---|---|---|
| | Baseline | Week 4 | Week 12 | Change from baseline to week 4 | Change from baseline to week 12 |
| Pooled | | | | | |
| Overall | 11.02 (5.30), 9.78, 1022 | 6.22 (5.60), 5.57, 948 | 5.23 (5.40), 4.27, 834 | $-4.86$ (4.64), $-4.68$, 948 | $-5.82$ (5.40), $-5.71$, 834 |
| Sub-sample 1 | 11.09 (5.75), 9.80, 511 | 6.51 (6.00), 5.71, 472 | 5.37 (5.39), 4.71, 407 | $-4.64$ (4.36), $-4.36$, 472 | $-5.78$ (5.47), $-5.49$, 407 |
| Sub-sample 2 | 10.96 (4.82), 9.69, 511 | 5.92 (5.16), 5.29, 476 | 5.09 (5.42), 3.86, 427 | $-5.08$ (4.89), $-4.90$, 476 | $-5.87$ (5.35), $-5.93$, 427 |
| Sub-sample 3 | 11.07 (6.10), 9.60, 511 | 6.18 (6.10), 5.43, 483 | 5.17 (5.94), 3.86, 412 | $-4.91$ (4.91), $-4.90$, 483 | $-5.80$ (5.67), $-5.83$, 412 |
| Sub-sample 4 | 10.99 (4.38), 9.90, 509 | 6.25 (5.03), 5.71, 463 | 5.29 (4.83) 4.43, 421 | $-4.83$ (4.33), $-4.57$, 463 | $-5.86$ ($-5.14$), $-5.62$, 421 |

*SD* standard deviation, *VMS* vasomotor symptoms

**Table 3** PGI-C VMS categorical responses at week 4 and week 12 in the pooled population

| Time point/PGI-C VMS response | Placebo ($n$ = 342) | Fezolinetant 30 mg ($n$ = 339) | Fezolinetant 45 mg ($n$ = 341) |
|---|---|---|---|
| Week 4, $n$ (%) | $n$ = 311 | $n$ = 305 | $n$ = 319 |
| Much better | 57 (18.3) | 111 (36.4) | 140 (43.9) |
| Moderately better | 44 (14.1) | 54 (17.7) | 56 (17.6) |
| A little better | 90 (28.9) | 84 (27.5) | 77 (24.1) |
| No change | 98 (31.5) | 52 (17.0) | 44 (13.8) |
| A little worse | 7 (2.3) | 3 (1.0) | 0 |
| Moderately worse | 7 (2.3) | 1 (0.3) | 1 (0.3) |
| Much worse | 8 (2.6) | 0 | 1 (0.3) |
| $P$ value[a] | – | < 0.001 | < 0.001 |
| Week 12, $n$ (%) | $n$ = 293 | $n$ = 275 | $n$ = 303 |
| Much better | 70 (23.9) | 117 (42.5) | 144 (47.5) |
| Moderately better | 48 (16.4) | 49 (17.8) | 65 (21.5) |
| A little better | 76 (25.9) | 67 (24.4) | 67 (22.1) |
| No change | 76 (25.9) | 36 (13.1) | 19 (6.3) |
| A little worse | 12 (4.1) | 2 (0.7) | 3 (1.0) |
| Moderately worse | 6 (2.0) | 3 (1.1) | 4 (1.3) |
| Much worse | 5 (1.7) | 1 (0.4) | 1 (0.3) |
| $P$ value[a] | – | < 0.001 | < 0.001 |

*PGI-C* Patient Global Impression of Change, *VMS* vasomotor symptoms

[a]Multiplicity unadjusted $P$ values were obtained using Cochran-Mantel–Haenszel test with modified ridit scores stratified by study; the association between response and treatment group (fezolinetant 30 mg versus placebo and fezolinetant 45 mg versus placebo) at each analysis visit was tested

## VMS Frequency Responder Analyses

The single-responder analyses applied the primary threshold estimates based on PGI-C VMS "moderately better" from the pooled sample at week 4 (mean value − 5.73 VMS/day) and week 12 (mean value − 6.20 VMS/day), as described previously. Based on the participants with a non-missing value at the analysis visit, the proportions of responders at week 12 were greater in the fezolinetant 30-mg (50.0%; $n$ = 133/266) and 45-mg (55.1%; $n$ = 161/292) groups compared with the placebo group (31.4%; $n$ = 88/280) [27]. When a "missing as

non-responder" imputation method was used, greater proportions of responders were observed in the fezolinetant 30-mg (42.8%; $n$ = 145/339) and 45-mg (46.6%; $n$ = 159/341) groups compared with the placebo group (24.0%; $n$ = 82/342) at week 4. The corresponding ORs (95% CI) for a VMS frequency responder compared with placebo were 2.48 (1.78–3.47; $P$ < 0.001) for fezolinetant 30 mg and 2.90 (2.09–4.07; $P$ < 0.001) for fezolinetant 45 mg. Similarly, at week 12, greater proportions of participants experienced a meaningful reduction in VMS frequency in the fezolinetant 30-mg (38.9%; $n$ = 132/339) and 45-mg (47.2%; $n$ = 161/341)

**Table 4** Correlation of change scores between moderate-to-severe VMS frequency and PGI-C VMS at weeks 4 and 12 in the pooled population

| Study/sample | Polyserial $\rho$, Spearman $\rho$, n | |
|---|---|---|
| | Week 4 | Week 12 |
| Pooled | | |
| Overall | 0.55, 0.59, 909 | 0.48, 0.52, 796 |
| Sub-sample 1 | 0.57,[a] 0.61, 454 | 0.56, 0.60, 392 |
| Sub-sample 2 | 0.53, 0.57, 455 | 0.39,[b] 0.43, 404 |
| Sub-sample 3 | 0.53, 0.58, 461 | 0.50, 0.53, 399 |
| Sub-sample 4 | 0.57, 0.60, 446 | 0.46, 0.50, 396 |

*PGI-C* Patient Global Impression of Change, *VMS* vasomotor symptoms

[a]Largest correlation value

[b]Smallest correlation value

groups compared with the placebo group (25.7%; $n = 88/342$). The corresponding ORs (95% CI) compared with placebo were 1.90 (1.36–2.65; $P < 0.001$) and 2.68 (1.94–3.74; $P < 0.001$) for fezolinetant 30 mg and 45 mg, respectively.

## DISCUSSION

Understanding the patient's perspective of their condition and benefits from treatment provides value to regulatory health authorities and healthcare professionals in routine care. In this pre-specified analysis of pooled data from SKYLIGHT 1 and 2, treatment with fezolinetant led to approximately five to six fewer moderate-to-severe VMS episodes/day at weeks 4 and 12 compared with baseline. The PGI-C VMS data at weeks 4 and 12 indicated that a greater proportion of participants treated with fezolinetant experienced an improved response versus placebo. Threshold estimates based on using a "moderately better" PGI-C VMS response as the anchor for meaningful within-patient change indicate that a reduction of approximately six VMS episodes/day represents a meaningful improvement to patients (mean primary threshold values for meaningful within-patient

change were $-5.7$ at week 4 and $-6.2$ at week 12). In addition, higher odds of achieving meaningful within-patient change in the frequency of VMS were observed for both fezolinetant doses versus placebo at weeks 4 and 12. This analysis helps to highlight the timing and thresholds of improvement, which appear to be important for patients with moderate-to-severe VMS.

Our analyses show that PGI-C VMS is an appropriate anchor measure for defining meaningful within-patient change in VMS frequency. Correlations between change scores (baseline to week 4 and week 12) in PGI-C VMS and the frequency of moderate-to-severe VMS in the overall pooled sample were 0.55–0.59 at week 4 and 0.48–0.52 at week 12. These correlation estimates exceeded the minimum magnitude of correlation criterion ($> 0.37$) required to support the appropriateness of the candidate anchor measure [25, 26]. The robustness of these data is supported by consistent data estimated by both polyserial and Spearman's rank correlations and consistent data regardless of the sample (i.e., pooled population, SKYLIGHT 1 and SKYLIGHT 2, and the subsamples).

There is currently no consensus concerning the best approach to assess treatment outcomes or symptom modifications from a patient's perspective in moderate-to-severe VMS. PGI-C VMS is based on the generic PGI-C scale tool but adapted for VMS (it asks patients to rate their hot flushes/night sweats at a particular time point compared with the beginning of the study using a seven-point Likert scale of "much better" to "much worse"). PGI-C is an easy-to-use self-reported global rating measure that can help evaluate a patient's assessment of efficacy relative to anchor points across multiple conditions [28, 29]. The PGI-C VMS is consistent in construction with the global rating items that are included in nearly all clinical trials to capture patients' perceptions regarding current experience of or change from baseline in disease or symptom severity. Anchoring objective VMS frequency to subjective PGI-C VMS establishes the patient-reported meaningfulness of the change in VMS frequency. The key objective of this analysis was to define and apply a threshold

**Table 5** Meaningful within-patient change estimation statistics of change in moderate-to-severe VMS frequency by PGI-C VMS in the pooled population

| Time point/PGI-C VMS response | Moderate-to-severe VMS frequency | |
|---|---|---|
| | N, mean (SD) [95% CI] | Median (Q1, Q3) |
| Week 4 | | |
| Much worse | 9, − 0.27 (3.30) [− 2.80 to 2.27][a] | − 1.24 (− 2.34, 1.26)[a] |
| Moderately worse | 8, − 0.86 (3.00) [− 3.37 to 1.64][a] | − 1.44 (− 2.92, 1.91)[a] |
| A little worse | 10, − 3.18 (2.75) [− 5.15 to − 1.21][a] | − 2.70 (− 5.30, − 0.57)[a] |
| No change | 189, − 1.80 (3.08) [− 2.24 to 1.36] | − 1.37 (− 3.42, 0.11) |
| A little better | 246, − 3.68 (4.35) [− 4.22 to − 3.13] | − 3.52 (− 5.54, − 1.47) |
| Moderately better | 153, − 5.73 (3.47) [− 6.29 to − 5.18][b] | − 5.79 (− 7.44, − 3.26)[b] |
| Much better | 294, − 7.76 (4.59) [− 8.28 to − 7.23] | − 7.37 (− 9.20, − 5.89) |
| Week 12 | | |
| Much worse | 6, − 1.40 (6.00) [− 7.69 to 4.89][a] | − 0.11 (− 2.26, 1.40)[a] |
| Moderately worse | 13, − 3.40 (3.86) [− 5.74 to − 1.07][a] | − 2.26 (− 5.51, − 1.49)[a] |
| A little worse | 17, − 1.41 (5.50) [− 4.24 to 1.42][a] | − 1.30 (− 5.28, 3.14)[a] |
| No change | 128, − 2.77 (3.57) [− 3.40 to − 2.14] | − 2.99 (− 4.96, − 0.53) |
| A little better | 199, − 4.69 (5.82) [− 5.50 to − 3.87] | − 4.33 (− 6.52, − 1.89) |
| Moderately better | 150, − 6.20 (5.18) [− 7.04 to − 5.36][b] | − 6.28 (− 8.29, − 4.13)[b] |
| Much better | 283, − 8.29 (4.95) [− 8.87 to − 7.71] | − 7.76 (− 9.60, − 5.98) |
| Half-SD of VMS frequency at baseline | 2.65 | − |

*CI* confidence interval, *PGI-C* Patient Global Impression of Change, *Q1* first quartile, *Q3* third quartile, *SD* standard deviation, *VMS* vasomotor symptoms

[a]Small samples whose estimates are considered too unreliable to use

[b]Primary anchor group data

for meaningful within-patient change to support the evaluation of the relevance between clinical changes and patient-observed changes in VMS frequency with fezolinetant. In our study, the meaningful within-patient change threshold represents the smallest change in an outcome score that is considered a meaningful change at an individual level. In this context, "meaningful within-patient change" is considered equivalent to similar terms used in the literature. This approach is in line with FDA guidance [30], which states that "it is important to understand how [a clinical outcome assessment or patient reported outcome]-based

endpoint corresponds to changes relevant to patients (e.g., the type and extent of change that is meaningful to patients)". In our paper, the focus is on the derivation of thresholds that allow one to qualify a clinically meaningful change in VMS frequency at patient level (not placebo-adjusted), from baseline to weeks 4 and 12.

Over 30 years ago, the term "minimal clinically important difference" was defined as "the smallest difference in score in the domain of interest that patients perceive as beneficial and would mandate, in the absence of troublesome side effects and excessive costs, a change in the
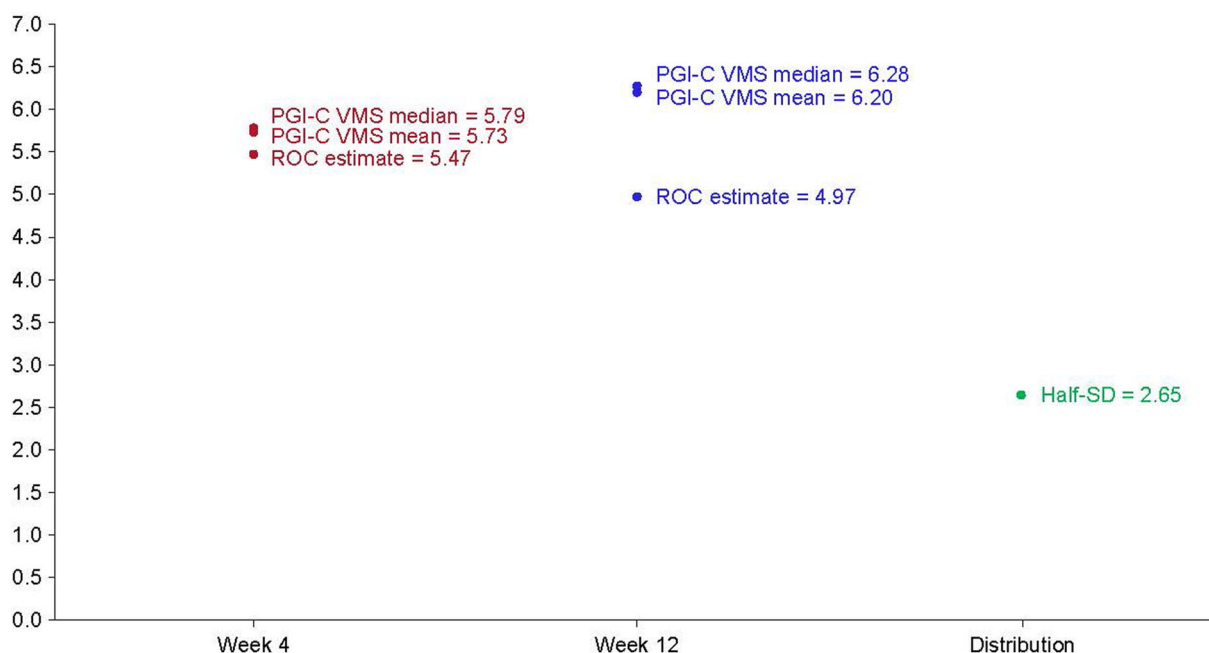
**Fig. 1** Meaningful within-patient change triangulation plot for moderate-to-severe VMS frequency in the pooled population. *PGI-C* Patient Global Impression of Change, *ROC* receiver operating characteristic, *SD* standard deviation, *VMS* vasomotor symptoms

patient's management" [31]. A number of additional terms have been introduced in the proceeding years, including "minimally important difference," "minimally important change," "clinically important difference," "minimally detectable difference," and "minimum detectable change" [32]. The precise definition applied to these alternative terms may differ across studies, but they all generally aim to quantify thresholds of change that are considered clinically relevant either for an individual or a group [32, 33]. Furthermore, anchor-based and distribution-based methods are the two most common approaches for estimating meaningful/minimal changes in PROs [32, 34, 35]. We used an anchor-based method aligned to recommendations by Revicki et al., who proposed the use of the anchor-based method to provide primary estimates of an instrument's meaningful within-patient change, and we also used the distribution-based method to provide supportive evidence when anchor-based estimates were unavailable [35].

A few prior publications have assessed clinically/minimally important difference in postmenopausal women with moderate-to-severe VMS, but they involved different treatments and were anchored with different PRO measures compared with our study. In two studies of hormone therapy, weekly VMS severity [24] or weekly VMS frequency [36] were anchored to generic (not specific to VMS) CGI outcomes. Other reports of responder thresholds in moderate-to-severe VMS include VMS frequency anchored to CGI and the Menopause-Specific Quality of Life questionnaire in women treated with hormone therapy [23], VMS frequency and severity anchored to the Menopause Symptoms Treatment Satisfaction Questionnaire in women treated with desvenlafaxine [37], and VMS frequency anchored to the Hot Flash Related Daily Interference Scale/Hot Flash Interference Scale in women treated with escitalopram [38].

Further clinically important changes from a patient perspective have been observed with fezolinetant based on improvements in other PRO data reported in the SKYLIGHT 1 and 2 studies. Statistically significant improvements from baseline in Menopause-Specific Quality of Life total score and the VMS domain score were

observed at weeks 4 and 12 in both the fezolinetant 30-mg and 45-mg groups compared with the placebo group [15, 16]. In addition, both fezolinetant doses demonstrated numerical improvements in sleep scores, measured by the Patient-Reported Outcomes Information System Short Form v1.0 Sleep Disturbance 8b total scores [15, 16]. Overall, these data indicate that fezolinetant improves different aspects of HRQoL from as early as week 4 in women with moderate-to-severe VMS.

PROs are routinely collected in clinical trials and are used in conjunction with clinical objective outcomes to assess the overall impact of a medical treatment or intervention [39]. Data derived from PROs provide important evidence for the impact of treatment on patient-reported symptoms and HRQoL [39, 40]. The use of PROs is encouraged by the major international health policy and regulatory authorities (e.g., European Medicines Agency and the FDA), and findings from a validated PRO can be used to support clinical decision-making, pharmaceutical labeling claims, product reimbursement, and health technology assessment bodies/payers appraisal [18, 40]. Perhaps most importantly, PROs are valuable tools that capture the patient's perspective, allowing healthcare professionals and patients to make more informed healthcare decisions.

Strengths of this analysis include the derivation of data from a large, pooled population of approximately 1000 patients with moderate-to-severe VMS from two phase 3, randomized, double-blind, placebo-controlled studies, and the meaningful within-patient change was based on changes observed at two time points. Another strength is the evaluation of threshold-estimate plausibility using several supportive analyses. A limitation in our study is that the average weight and BMI were relatively high. While consistent with several other studies conducted in menopausal women, it is recognized that the frequency and severity of VMS can be greater in women with a higher BMI. A further limitation is that the PGI-C VMS instrument may be subject to recall bias. Additionally, data collection within a clinical trial setting may limit the generalizability of these findings to routine clinical practice.

## CONCLUSION

VMS are bothersome for most women undergoing menopausal transition, but few large clinical trials have reported patient perception of meaningful symptomatic improvement. This analysis of pooled data from the SKYLIGHT 1 and 2 trials demonstrates that PGI-C VMS is sensitive to change and correlates with VMS frequency, and a reduction of approximately six VMS per day is a meaningful improvement for patients with moderate-to-severe VMS. Application of this threshold in a responder analysis found greater odds (up to three times) of achieving meaningful within-patient reductions in VMS for fezolinetant 45 mg compared with placebo at weeks 4 and 12. These analyses may support the interpretation of data for fezolinetant with different stakeholders, including but not limited to clinicians and patients. Overall, these data support the position that fezolinetant provides a meaningful clinical benefit for women who have VMS associated with menopause and that it represents an important nonhormonal treatment option.

## ACKNOWLEDGEMENTS

### Declarations

# REFERENCES

1. Rance NE, Dacks PA, Mittelman-Smith MA, Romanovsky AA, Krajewski-Hall SJ. Modulation of body temperature and LH secretion by hypothalamic KNDy (kisspeptin, neurokinin B and dynorphin) neurons: a novel hypothesis on the mechanism of hot flushes. Front Neuroendocrinol. 2013;34(3):211–27.

2. Thurston RC, Joffe H. Vasomotor symptoms and menopause: findings from the Study of Women's Health across the Nation. Obstet Gynecol Clin N Am. 2011;38(3):489–501.

3. Deecher DC, Dorries K. Understanding the pathophysiology of vasomotor symptoms (hot flushes and night sweats) that occur in perimenopause, menopause, and postmenopause life stages. Arch Womens Ment Health. 2007;10(6):247–57.

4. Freeman EW, Sammel MD, Sanders RJ. Risk of long-term hot flashes after natural menopause: evidence

from the Penn Ovarian Aging Study cohort. Menopause. 2014;21(9):924–32.

5. Gold EB, Colvin A, Avis N, et al. Longitudinal analysis of the association between vasomotor symptoms and race/ethnicity across the menopausal transition: study of women's health across the nation. Am J Public Health. 2006;96(7):1226–35.

6. Santoro N, Roeca C, Peters BA, Neal-Perry G. The menopause transition: signs, symptoms, and management options. J Clin Endocrinol Metab. 2021;106(1):1–15.

7. Williams RE, Levine KB, Kalilani L, Lewis J, Clark RV. Menopause-specific questionnaire assessment in US population-based study shows negative impact on health-related quality of life. Maturitas. 2009;62(2):153–9.

8. Williams RE, Kalilani L, DiBenedetti DB, Zhou X, Fehnel SE, Clark RV. Healthcare seeking and treatment for menopausal symptoms in the United States. Maturitas. 2007;58(4):348–58.

9. Avis NE, Crawford SL, Greendale G, et al. Duration of menopausal vasomotor symptoms over the menopause transition. JAMA Intern Med. 2015;175(4):531–9.

10. Neves-E-Castro M, Birkhauser M, Samsioe G, et al. EMAS position statement: the ten point guide to the integral management of menopausal health. Maturitas. 2015;81(1):88–92.

11. Advisory Panel. The 2022 hormone therapy position statement of The North American Menopause Society. Menopause. 2022;29(7):767–94.

12. Nappi RE, Kroll R, Siddiqui E, et al. Global cross-sectional survey of women with vasomotor symptoms associated with menopause: prevalence and quality of life burden. Menopause. 2021;28(8): 875–82.

13. Depypere H, Lademacher C, Siddiqui E, Fraser GL. Fezolinetant in the treatment of vasomotor symptoms associated with menopause. Expert Opin Investig Drugs. 2021;30(7):681–94.

14. Astellas Pharma US. VEOZAH™ (fezolinetant). Highlights of prescribing information. 2023. https://www.accessdata.fda.gov/drugsatfda_docs/label/2023/216578s000lbl.pdf. Accessed 12 July 2023.

15. Johnson KA, Martin N, Nappi RE, et al. Efficacy and safety of fezolinetant in moderate-to-severe vasomotor symptoms associated with menopause: a phase 3 RCT. J Clin Endocrinol Metab. 2023;108(8): 1981–97.

16. Lederman S, Ottery FD, Cano A, et al. Fezolinetant for treatment of moderate-to-severe vasomotor symptoms associated with menopause (SKYLIGHT 1): a phase 3 randomised controlled study. Lancet. 2023;401(10382):1091–102.

17. Neal-Perry G, Cano A, Lederman S, et al. Safety of fezolinetant for vasomotor symptoms associated with menopause: a randomized controlled trial. Obstet Gynecol. 2023;141(4):737–47.

18. US Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. https://www.fda.gov/media/77832/download. Accessed 12 July 2023.

19. Powers JH 3rd, Patrick DL, Walton MK, et al. Clinician-reported outcome assessments of treatment benefit: report of the ISPOR Clinical Outcome Assessment Emerging Good Practices Task Force. Value Health. 2017;20(1):2–14.

20. Johnston B, Patrick D, Devji T, et al. Chapter 18: Patient-reported outcomes. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane. 2023. https://training.cochrane.org/handbook/current/chapter-18. Accessed 12 Sept 2023.

21. US Food and Drug Administration. Estrogen and estrogen/progestin drug products to treat vasomotor symptoms and vulvar and vaginal atrophy symptoms—recommendations for clinical evaluation. 2003. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/estrogen-and-estrogenprogestin-drug-products-treat-vasomotor-symptoms-and-vulvar-and-vaginal-atrophy. Accessed 12 Sept 2023.

22. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont). 2007;4(7):28–37.

23. Constantine GD, Revicki DA, Kagan R, et al. Evaluation of clinical meaningfulness of estrogen plus progesterone oral capsule (TX-001HR) on moderate to severe vasomotor symptoms. Menopause. 2019;26(5):513–9.

24. Constantine GD, Simon JA, Kaunitz AM, et al. TX-001HR is associated with a clinically meaningful effect on severity of moderate to severe vasomotor symptoms in the REPLENISH trial. Menopause. 2020;27(11):1236–41.

25. Fayers PM, Hays RD. Don't middle your MIDs: regression to the mean shrinks estimates of minimally important differences. Qual Life Res. 2014;23(1):1–4.

26. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. COPD. 2005;2(1):63–7.

27. Nappi RE, Johnson KA, Stute P, et al. Treating moderate-to-severe menopausal vasomotor symptoms with fezolinetant: analysis of responders using pooled data from two phase 3 studies (SKYLIGHT 1 and 2). Menopause. 2024. https://doi.org/10.1097/GME.0000000000002354.

28. Rampakakis E, Ste-Marie PA, Sampalis JS, Karellis A, Shir Y, Fitzcharles MA. Real-life assessment of the validity of patient global impression of change in fibromyalgia. RMD Open. 2015;1(1):e000146.

29. Zhan S, Ye H, Li N, et al. Comparative efficacy and safety of multiple wake-promoting agents for the treatment of excessive daytime sleepiness in narcolepsy: a network meta-analysis. Nat Sci Sleep. 2023;15:217–30.

30. US Food and Drug Administration. Patient-focused drug development: incorporating clinical outcome assessments into endpoints for regulatory decision-making. Guidance for industry, Food and Drug Administration staff, and other stakeholders. 2023. https://www.fda.gov/media/166830/download. Accessed 15 Feb 2024.

31. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407–15.

32. Draak THP, de Greef BTA, Faber CG, Merkies ISJ, PeriNomS study group. The minimum clinically important difference: which direction to take. Eur J Neurol. 2019;26(6):850–5.

33. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res. 2011;11(2):171–84.

34. Amtmann D, Kim J, Chung H, Askew RL, Park R, Cook KF. Minimally important differences for patient reported outcomes measurement information system pain interference for individuals with back pain. J Pain Res. 2016;9:251–5.

35. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61(2):102–9.

36. Gerlinger C, Gude K, Hiemeyer F, Schmelter T, Schäfers M. An empirically validated responder definition for the reduction of moderate to severe hot flushes in postmenopausal women. Menopause. 2012;19(7):799–803.

37. Wyrwich KW, Spratt DI, Gass M, Yu H, Bobula JD. Identifying meaningful differences in vasomotor symptoms among menopausal women. Menopause. 2008;15(4 Pt 1):698–705.

38. Carpenter JS, Bakoyannis G, Otte JL, et al. Validity, cut-points, and minimally important differences for two hot flash-related daily interference scales. Menopause. 2017;24(8):877–85.

39. Rivera SC, Kyte DG, Aiyegbusi OL, Slade AL, McMullan C, Calvert MJ. The impact of patient-reported outcome (PRO) data from clinical trials: a systematic review and critical analysis. Health Qual Life Outcomes. 2019;17(1):156.

40. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. JAMA. 2018;319(5):483–94.