



Published in final edited form as:

Menopause. 2024 January 01; 31(1): 3–9. doi:10.1097/GME.0000000000002280.

Correlations Among COMMA-recommended VMS Outcomes in MsFLASH Trials

Janet S. Carpenter, PhD, RN, FAAN¹, Joseph C. Larson, MS², Myra S. Hunter, PhD C Psychol FBPS³, Sarah Lensen, PhD⁴, Chen X. Chen, PhD, RN¹, Katherine A. Guthrie, PhD²

¹Indiana University School of Nursing, Indianapolis, IN, 46202

²MsFLASH Data Coordinating Center, Fred Hutchinson Cancer Center, Seattle, WA 98109-1024

³Institute of Psychiatry, Psychology and Neuroscience, Kings College London, Guy's Campus, London SE1 9RT

⁴Obstetrics and Gynaecology, Royal Women's Hospital, Parkville, 3052, The University of Melbourne, Australia

Abstract

Objective: To advance understanding of vasomotor symptom (VMS) outcomes measurement using pooled data from three MsFLASH trials.

Methods: Participants self-reported VMS frequency, severity, and bother using daily diaries; completed standardized measures of VMS interference, insomnia severity, and sleep quality/disturbance; and completed four treatment satisfaction items. Analyses included descriptive statistics, Pearson's correlations (baseline pooled sample, post-treatment pooled sample, post-treatment placebo only), t-tests, and analysis of variance.

Results: Participants were mostly postmenopausal (82.9%) and a mean of 54.5 years old. VMS frequency was fairly correlated with severity, bother and interference for pooled baseline and placebo post-treatment samples (r 's=0.21 to 0.39, p values < .001) and moderately correlated with severity, bother, and interference for pooled post-treatment (r 's=0.40 to 0.44, p values < .001). VMS severity, bother, and interference were moderately correlated (r 's=0.37 to 0.48, p values < .001), with one exception. VMS severity and bother were strongly correlated (r 's=0.90 to 0.92, p values < .001). VMS interference was moderately correlated with insomnia (r 's=0.45 to 0.54, p values < .001) and fairly to moderately correlated with sleep quality/disturbance (r 's=0.31 to 0.44, p values < .001). Other VMS outcomes were weakly to fairly correlated with insomnia (r 's=0.07 to 0.33, p values < .001 to < .05) and sleep quality/disturbance (r 's=0.06 to 0.26, p values < .001 to > .05). Greater improvement in VMS and sleep over time was associated with higher treatment satisfaction (p values < .001).

Conclusions: This pooled analysis advances understanding of VMS outcomes measurement and has implications for selecting measures and creating future research.

Corresponding author: Janet S. Carpenter, PhD, RN, FAAN, IU School of Nursing, 600 Barnhill Drive NU E431, Indianapolis, IN 46260, phone 317-278-6093, fax none, carpentj@iu.edu.

Conflicts of Interest: All other authors have no disclosures.

Keywords

Menopause; perimenopause; postmenopause; vasomotor symptoms; symptom assessment

INTRODUCTION

Historically, there has been little conceptual or methodological agreement on the measurement of vasomotor symptoms (VMS). One systematic review revealed nearly 50 different conceptualizations and 16 different measurement tools across 214 randomized controlled trials where VMS were the primary outcome.¹ Even when a similar measure was used across studies, there was not always agreement on what the measure represented. For example, items originally intended to measure VMS interference have been used to measure VMS severity.² The lack of standardization and resulting inconsistencies in measurement create barriers for exchanging, pooling, and comparing data.

The Comma Core Outcomes in Menopause (COMMA) global initiative is creating standards for the conceptualization and measurement of VMS.³ In 2021, COMMA recommended a minimum core outcome set for measuring VMS. The recommendations resulted from an intense vetting process involving literature reviews, modified Delphi surveys, and consensus meetings. The literature review included attention to the United States Food and Drug Administration's draft guidance to measure VMS frequency and severity as co-primary endpoints when evaluating estrogen/progestin therapies.⁴ The resulting COMMA recommendations are to measure VMS symptom dimensions of (1) frequency, (2) severity, (3) distress, bother, or interference, (4) impact on sleep (which could be assessed within distress/bother/interference measures), (5) satisfaction with VMS treatment, and (6) side effects of VMS treatment.

As COMMA recommendations will impact those who design and participate in future clinical research, it is important to understand the degree to which VMS outcomes are potentially related. Although prior publications report associations between different outcomes, some studies are specific to breast cancer survivors,⁵⁻⁷ with fewer including women with and without cancer,⁸⁻¹⁰ or only women without a history of cancer.¹¹ Some VMS outcomes have been highly correlated, suggesting either, but not both, may need to be reported in clinical trials. In a study of 284 breast cancer survivors, VMS severity was highly correlated with VMS bother ($r=0.85$, $p < .01$).⁷ Other VMS outcomes have been shown to be more distinct and less strongly related to each another. For example, a structural equation modeling analysis of data from 395 breast cancer survivors showed VMS frequency and severity to be distinct from one another ($r=0.23$, $p < .05$), suggesting both should be reported in clinical trials.⁷

A second gap is lack of information about how the impact of VMS on sleep is related to standardized sleep scale scores, such as those measuring insomnia severity or sleep quality/disturbance. During the COMMA process, participants discussed the options of using standardized sleep scales versus using the Hot Flash Related Daily Interference Scale (HFRDIS)⁸ that contains one item which specifically measures the impact of VMS on sleep (personal communication Dr. Sarah Lensen, October 1, 2022). Understanding

correlations between VMS interference measures that include a VMS impact on sleep item and standardized sleep measures can shed light on this question.

Pooled data from the MsFLASH trials provides a unique opportunity to address these gaps. The overall purpose of our investigation was to advance understanding of VMS outcomes measurement using data from midlife women enrolled in MsFLASH trials. Aim 1 was to examine correlations among VMS frequency, severity, bother, interference (including impact on sleep), and standardized sleep scales (insomnia severity, sleep quality/disturbance) at baseline and post-treatment. Aim 2 was to examine relationships between VMS satisfaction with treatment and baseline to post-treatment changes in VMS frequency, severity, bother, interference (including impact on sleep), and standardized sleep scales (insomnia severity, sleep quality/disturbance).

METHODS

Use of Artificial Intelligence

ChatGPT was used as a proofreading tool in writing this manuscript.

Design and Participants

This was a retrospective analysis of existing participant-level data from baseline and post-treatment follow-up from three randomized, controlled, clinical trials testing treatments for managing VMS (Menopause Strategies Finding Lasting Answers to Symptoms and Health, MsFLASH trials 01, 02, 03). Women enrolled in these trials were generally healthy and were not currently receiving treatment for VMS. Interventions included 8 weeks of escitalopram versus placebo (Trial 01), 12 weeks of omega-3 fatty acids versus placebo crossed with exercise, yoga, or usual care (Trial 02), and 12 weeks of low dose estradiol versus venlafaxine versus placebo (Trial 03). Prior publications detail the MsFLASH trial methods, including sampling criteria, procedures, and measures¹² as well as major findings.^{13–18} All trials were Institutional Review Board approved. Data from all participants who provided written informed consent and were enrolled in the trials were included in this analysis.

Data Source Variables

Sample description variables were obtained at baseline. Participants self-reported age, race, ethnicity, menopause status, and smoking history. Clinic staff measured height and weight to calculate body mass index.

Aim 1 outcomes of VMS frequency, severity, bother, and interference (including impact on sleep) were measured at baseline and post-treatment. Participants self-reported VMS frequency, severity, and bother in daily diaries by recording the number of occurrences and checking boxes to rate severity (1 mild, 2 moderate, 3 severe) and bother (1 none, 2 a little, 3 moderately, 4 a lot) each day.¹² All ratings were recorded at bedtime to reflect the day's VMS and in the morning to reflect the previous night's VMS. VMS interference was assessed using the 10-item HFRDIS which measures VMS interference with different aspects of life. Responses for each item are on an 11 point scale, from 0 (did not interfere)

to 10 (completely interfered).⁸ Scores for the 3-item Hot Flash Interference (HFI) derivative scale were also calculated using HFRDIS items.¹¹ Both the HFRDIS and HFI include one item pertaining to how much VMS interfered with sleep.¹¹ Mean total scores for the HFRDIS and HFI are interpreted as none (0), mild (1–3), moderate (4–6) and severe (7–10) interference. Average minimally clinically important difference values are 1.66 for the HFRDIS and 2.34 for the HFI.¹¹

Aim 1 sleep outcomes were assessed using standardized sleep scales administered at baseline and post-treatment. Participants self-reported insomnia symptoms on the Insomnia Severity Index (ISI).^{19, 20} ISI scores are interpreted as no (0–7), mild (8–14), moderate (15–21), and severe insomnia (22–28).²¹ Participants self-reported sleep quality/disturbance on the Pittsburgh Sleep Quality Index (PSQI) where global scores > 5.0 indicate poor sleep quality/high sleep disturbance.^{22, 23} In this dataset, correlations between insomnia and sleep quality/disturbance were strong for the pooled baseline sample ($r=0.72$, $p < .001$), pooled post-treatment sample ($r=0.75$, $p < .001$), and placebo post-treatment sample ($r=0.75$, $p < .001$).

Aim 2 satisfaction with VMS treatment was assessed at post-treatment only with four MsFLASH-designed items. All four items reflected satisfaction with pills only, as there was no separate assessment of Trial 02 behavioral treatments (exercise, yoga). First, participants indicated whether they were satisfied with the hot flash control or relief they had with the study medication or pills (yes, no). Second, participants indicated if they wanted to continue the study medication or pills (yes, no). Third, participants marked one of three statements to best describe their impression of the study benefits: (1) “The study medication/pills helped reduce my hot flashes, with minimal or no side effects”, (2) “The study medication/pills helped reduce my hot flashes, but is not worth the side effects”, and (3) “The study medication/pills did not help reduce my hot flashes”. For the fourth item, participants marked perception or guess as to which study medication/pill group they had been assigned (all trials were double-blinded). Responses were pooled across trials into three groups: active pill (Trial 01 escitalopram, Trial 02 omega-3, and Trial 03 venlafaxine, low-dose estrogen, and active study pills but not sure if venlafaxine or low-dose estrogen), inactive or placebo pill, and don’t know. MsFLASH side effects assessments were not comparable across interventions and, thus, were not included in this pooled analysis.

Statistical Methods

All analyses were conducted using SAS for Windows Version 9.4 (SAS Institute, Inc., Cary, NC). Descriptive statistics were used to summarize sample baseline characteristics, VMS frequency, severity, bother, interference, insomnia severity, and sleep quality/disturbance. To address limitations of prior research that examined correlations only at pre-intervention baseline,^{5, 7–9} for Aim 1, Pearson’s correlation coefficients among VMS and sleep measures were calculated pre- and post-intervention. Pearson’s correlations were calculated for baseline and post-treatment among the pooled sample ($n=899$). As the pooled sample included women who had received both active study medication/pills and placebo, Aim 1 correlations were also calculated post-treatment among the subset of women who received placebo pills only ($n=389$). Thus, this analysis yielded three sets of correlations: (1) baseline

for the pooled sample (n=899), (2) post-treatment for the pooled sample (n=899), and (3) post-treatment for placebo recipients (n=389).

To provide a consistent rubric for interpreting correlation coefficients in this paper, we combined four different criteria outlined in Overholser and Sowinski²⁴ and interpreted the strength of association based on absolute values of Pearson's correlations as follows: < 0.20 weak, 0.20 and < 0.40 fair, 0.40 to < 0.70 moderate, 0.70 strong.

For Aim 2, independent samples t-tests and analysis of variance were used to evaluate relationships between the four satisfaction with treatment items (measured at post-treatment only) and baseline to post-treatment changes in VMS frequency, severity, bother, interference, insomnia severity, and sleep quality/disturbance.

RESULTS

A description of the sample's baseline characteristics, VMS, and sleep is shown in Table 1. Participants' mean age was 54.5 and mean body mass index was 28.0. The majority of participants for this analysis self-identified as Black or White, Non-Latina, postmenopausal, and never smokers. Baseline data indicated participants reported, on average, 8 VMS per 24 hours, moderate VMS severity, moderately bothersome VMS, mild to moderate VMS interference, mild insomnia, and poor sleep quality/high sleep disturbance.

Aim 1 Relationships Among VMS Frequency, Severity, Bother, Interference, Insomnia Severity, and Sleep Quality/Disturbance at Baseline and Post-Treatment

Correlations among VMS and sleep outcomes are shown in Table 2 for the pooled sample at baseline (top shaded rows), pooled sample at post-treatment (middle unshaded rows), and placebo pills only sample post-treatment (bottom shaded rows). VMS frequency was only fairly correlated with severity, bother and interference for the pooled baseline and placebo pill post-treatment samples (r 's=0.21 to 0.39) and moderately correlated with severity, bother and interference for the pooled post-treatment sample (r 's=0.40 to 0.44). VMS severity, bother, and interference were all moderately correlated with one another (r 's=0.37 to 0.48) in all analyses, with the exception of a fair correlation between VMS severity and the HFI for the pooled baseline sample. Because the HFI is a subset of HFRDIS items, the two VMS interference measures were strongly correlated (r 's=0.90 to 0.92). Thus, correlations between HFRDIS and other outcomes were nearly identical to correlations between HFI and other outcomes. VMS severity and bother were strongly correlated (r 's=0.90 to 0.92), thus correlations between severity and other outcomes were nearly identical to correlations between bother and other outcomes. Correlations between VMS interference measures (which included VMS impact on sleep) and sleep scales were higher than correlations between other VMS measures and sleep scales. VMS interference (HFRDIS, HFI) was moderately correlated with insomnia (r 's=0.45 to 0.54) and fairly (HFRDIS) to moderately (HFI) correlated with sleep quality/disturbance (r 's=0.31 to 0.44). In comparison, other VMS outcomes were weakly to fairly correlated with insomnia (r 's=0.07 to 0.33) and sleep quality/disturbance (r 's=0.06 to 0.26).

Aim 2 Relationships Among VMS Satisfaction with Treatment and Changes from Baseline to Post-Treatment in VMS Frequency, Severity, Bother, Interference, Insomnia Severity, and Sleep Quality/Disturbance

The distribution of responses to the VMS satisfaction with treatment items for the pooled sample showed the following. Participants were evenly split between reporting they were (n=426, 47.4%) or were not satisfied (n=426, 47.4%) with the hot flash control/relief they had with the study medication/pills, with 5.2% (n=47) missing responses. More participants reported not wanting to continue the study medication or pills (n=457, 50.8%) than wanting to continue (n=385, 42.8%), with 6.3% missing responses (n=57). Slightly more participants endorsed that the study medication or pills helped reduce hot flashes with minimal or no side effects (46.6%) vs. did not help reduce hot flashes (41.9%) or helped but were not worth the side effects (5.7%), with 5.8% missing responses (n=52). Around a third of participants thought they were in an active pill group (n=331, 36.8%) versus inactive/placebo group (n=274, 30.5%) versus reporting they didn't know (n=250, 27.8%), with 4.9% missing responses (n=44).

Tables 3, 4, and 5 show relationships between VMS satisfaction items and changes in VMS and sleep outcomes over time. Satisfaction with treatment on all items was highly associated with greater improvements in VMS and sleep outcomes over time. Improvements in VMS interference exceeded minimally important differences for the HFRDIS and HFI. In Table 3, significantly greater reductions in VMS and sleep outcomes were seen for women who (1) were satisfied with the hot flash control/relief they had with the study medication/pills compared to women who were not ($p < 0.001$) and (2) desired to continue the study medication/pills compared to those who did not (Table 3, p values < 0.001). In Table 4, there were (1) significantly greater reductions in VMS frequency, severity, and bother between groups who thought the study medication/pills were beneficial compared to not beneficial (p values < 0.001); (2) significantly different reductions in VMS interference and sleep quality/disturbance between all three groups (p values < 0.001); and (3) significantly greater reductions in insomnia severity between those who thought the study medication/pills were beneficial with minimal side effects compared to not beneficial (p values < 0.001). In Table 5, significantly different reductions in all VMS outcomes were seen across the three groups. Post-hoc comparisons showed the greatest reductions in VMS were reported by those who thought they were on active pill, followed by the don't know group, then inactive/placebo pill. Also in Table 5, those who thought they were taking active pill reported (1) greater reductions in insomnia severity compared to the other two groups and (2) greater improvement in sleep quality/disturbance compared to those who thought they were taking placebo (p values < 0.001).

DISCUSSION

There are six important findings from this analysis. The first important finding is that VMS frequency was not strongly correlated with other VMS outcomes. Desire to measure frequency in the United States MsFLASH clinical trials partially stemmed from the only VMS measurement guidance available before COMMA, which was federal draft guidance for industry testing of estrogen and estrogen/progestin products for the treatment of VMS

issued by the Food and Drug Administration.⁴ The draft guidance statement specifies that change in frequency and severity be analyzed as co-primary endpoints and that “subjective measures (e.g., daily patient diary entries) can be used as primary efficacy endpoints”.⁴ However, it is unclear which VMS outcomes are most salient to women. In addition, researchers have discussed issues with diary entries for other non-VMS symptoms such as pain, including “hoarding” where participants who wish to appear adherent to diaries fill in multiple missing diary entries at a single sitting.²⁵ In this analysis, the low correlations between VMS frequency and other VMS outcomes may indicate diary VMS frequency is (1) capturing a concept that is distinct from VMS severity, bother, and interference, (2) prone to noise from recording errors throughout the daytime and/or nighttime, or (3) both. Further research using qualitative or mixed methods could help to untangle these potential explanations for the low correlations between VMS frequency and other VMS outcomes.

The second finding is that VMS severity and bother were strongly correlated as in prior studies of VMS⁷ and menstrual pain.²⁶ Both measures used Likert-type response options; severity with a 3-point scale (mild, moderate, severe) and bother with a 4-point scale (none, a little, moderately, a lot). The similar Likert options and their sequential appearance on the diary could have resulted in response bias, where participants rated both similarly and/or may have had difficulty differentiating severity from bother. Although some investigators have adopted an etic perspective and imposed definitions of mild, moderate, and severe VMS for participants,²⁷ doing so can be problematic in situations where there is a high degree of inter-individual variability. Anthropological and epidemiological studies have shown a high degree of inter-individual and inter-cultural variability in how VMS are experienced,^{28–31} thus pre-specified definitions that capture only a limited perspective may not be applicable to all women. For example, defining severity in terms of sweating may not be meaningful to women who sweat very little but suffer from extremely uncomfortable sensations of heat. In contrast, MsFLASH investigators adopted an emic perspective and allowed participants to self-define the response options for severity and bother. It is difficult to know whether severity and bother are highly related concepts (e.g., more severe symptoms associated with more bother), are conceptually overlapping or difficult to distinguish (e.g., severity and bother have a similar meaning to participants), or are rated similarly because of participant response bias. Further qualitative or mixed methods could help to gain women’s perspectives on their interpretation of the similarities or differences in severity and bother.

Third, correlations between interference and other VMS outcomes, as well as sleep, were very similar for the 10-item HFRDIS and 3-item HFI. While the single item of sleep interference, as captured by both the HFRDIS and HFI, assesses VMS-specific sleep interference, that single item is not intended to be used alone nor has it been validated as a stand alone item. As both the HFRDIS and HFI have been validated,¹¹ showed similar correlations with other VMS and sleep measures, and include the COMMA recommended outcome of VMS impact on sleep,³ either measure appears suitable for measuring interference while including the dimension of impact on sleep. These findings should reassure researchers who are concerned about response burden and wish to use only the 3-item HFI. Based on these results, the shorter HFI should perform as well as the longer HFRDIS in future studies.

The fourth finding highlights the importance of choosing a standardized sleep measure. At present, there is no validated questionnaire for measuring VMS-specific-related sleep disturbance in this population. Several symptom questionnaires include a single item related to sleep (e.g., Menopause Rating Scale, Menopause Quality of Life Scale) but these single items are not standardized, validated, reliable sleep measures. In this analysis, stronger correlations between VMS and insomnia severity (ISI) versus VMS and sleep quality/disturbance (PSQI) provide stronger evidence of construct validity for the ISI in the context of VMS. Although neither the ISI or PSQI standardized sleep scales are VMS-specific, the ISI items appear more relevant to VMS impact on sleep.¹⁹ The ISI has fewer items, is easier to score, and has established cutpoints.¹⁹ The PSQI contains items which may not be relevant to VMS-related sleep, such as bad dreams and snoring, and has previously been criticized for showing an unstable factor structure.²³ Thus, the ISI may be preferable as a standardized sleep measure in future studies of midlife women with VMS.

The fifth important finding is the similarity of correlations at baseline and post-treatment. Despite significant effects of treatment in MsFLASH trials 01 and 03,^{14, 16} the correlations appeared relatively stable over time. There is interest in measuring multiple VMS outcomes in any given trial because, theoretically, a given intervention could differentially affect outcome measures. For example, a behavioral treatment aiming to help women manage their VMS symptoms could decrease VMS severity or interference but not VMS frequency, whereas a medication targeting a specific physiological pathway could decrease VMS frequency, severity, and interference. If that were the case, the correlations would be expected to change over time from baseline to post-treatment. However, our findings suggest that, at least within the MsFLASH trials, relationships were consistent over time. Although differential effects on outcomes remains a theoretical possibility, a systematic review of literature could help determine whether and to what extent previously tested interventions support this theoretical possibility.

The sixth finding is the consistent relationships observed across all VMS treatment satisfaction items. Despite variations in the wording of the items, all four items were correlated with changes in VMS frequency, severity, bother, interference, insomnia severity, and sleep quality/disturbance. These findings suggest that these treatment satisfaction items, or other similarly phrased satisfaction items, have potential to be further psychometrically studied and potentially used in future studies.

There were some strengths and weaknesses to this study. The dataset was large and represented Black and White women from all MsFLASH clinical sites (Boston, Indianapolis, Oakland, Philadelphia, Seattle). MsFLASH measures included all that were recommended by COMMA. However, MsFLASH considered interference to be distinct from severity and bother and this assumption was supported by the fact that we did not find strong correlations between interference and severity or bother. Thus, MsFLASH trials included a measure of interference separate from bother whereas COMMA recommended including a single measure capturing one of the concepts of bother, distress, or interference. Two standardized measures of sleep were included which extends understanding about how VMS outcomes are related to different aspects of sleep. The sample was limited to women meeting strict inclusion and exclusion criteria for the various VMS treatment

trials and therefore well represents healthy women with VMS but poorly represents women not meeting trial criteria or not seeking VMS treatment. As discussed, the differences in methodology used to capture different outcomes may have contributed to the observed correlations between some outcomes. In addition, though rating symptoms on varying point scales, in diaries, and on questionnaires have been standards in the field for several decades, none of the MsFLASH measures were subjected to consumer testing prior to their use in the trials.

CONCLUSION

By leveraging data from three large clinical trials of VMS, this analysis advances understanding of VMS outcomes measurement. Findings suggest (1) VMS frequency may be a distinct concept from severity, bother, or interference, (2) severity and bother are highly correlated and both may not need to be reported, (3) the 3-item and 10-item interference measures were similarly correlated with other outcomes suggesting the shorter scale can be substituted for the longer scale, (4) the ISI may be more relevant than the PSQI to VMS research, (5) correlations were similar over time, and (6) changes in VMS and sleep were related to all of the four treatment satisfaction items. These findings have implications for choice of measurement instruments in clinical trials.

Sources of funding:

The MsFLASH network was supported by a cooperative agreement issued by the National Institute of Aging (NIA), in collaboration with the Eunice Kennedy Shriver National Institute of Child Health and Development (NICHD), the National Center for Complementary and Alternative Medicine (NCCAM) and the Office of Research and Women's Health (ORWH), and grants U01 AG032656, U01AG032659, U01AG032669, U01AG032682, U01AG032699, U01AG032700 from the NIA and an administrative supplement for the Seattle site U01 AG032682. At the Indiana University site, the project was funded in part with support from the Indiana Clinical and Translational Sciences Institute, funded in part by grant UL1RR025761 from the National Institutes of Health, National Center for Research Resources, Clinical and Translational Sciences Award. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Financial Disclosures

Dr. Carpenter received past consulting fees from the University of Wisconsin and Simumetrix SMX Health. Dr. Hunter receives funding from TurningPoint Charity and HelloTherapeutics.

REFERENCES

1. Iliodromiti S, Wang W, Lumsden MA, et al. Variation in menopausal vasomotor symptoms outcomes in clinical trials: a systematic review. *BJOG* 2020; 127: 320–333. DOI: 10.1111/1471-0528.15990. [PubMed: 31621155]
2. Xu H, Thurston RC, Matthews KA, et al. Are hot flashes associated with sleep disturbance during midlife? Results from the STRIDE cohort study. *Maturitas* 2012; 71: 34–38. DOI: 10.1016/j.maturitas.2011.10.003. [PubMed: 22051577]
3. Lensen S, Archer D, Bell RJ, et al. A core outcome set for vasomotor symptoms associated with menopause: the COMMA (Core Outcomes in Menopause) global initiative. *Menopause* 2021; 28: 852–858. DOI: 10.1097/GME.0000000000001787. [PubMed: 33906204]
4. U.S. Department of Health and Human Services (DHHS). Estrogen and estrogen/progestin drug products to treat vasomotor symptoms and vulvar and vaginal atrophy symptoms - recommendations for clinical evaluation: Draft guidance January 2003. Rockville: Food and Drug Administration Center for Drug Evaluation and Research (CDER).

5. Carpenter JS, Rand KL. Modeling the hot flash experience in breast cancer survivors. *Menopause* 2008; 15: 469–475. DOI: 10.1097/gme.0b013e3181591db7. [PubMed: 18467951]
6. Otte JL, Flockhart D, Hayes D, et al. Comparison of subjective and objective hot flash measures over time among breast cancer survivors initiating aromatase inhibitor therapy. *Menopause* 2009; 16: 653–659. DOI: 10.1097/gme.0b013e3181a5d0d6. [PubMed: 19455068]
7. Rand KL, Otte JL, Flockhart D, et al. Modeling hot flushes and quality of life in breast cancer survivors. *Climacteric* 2011; 13: 171–180. DOI: 10.3109/13697131003717070.
8. Carpenter JS. The Hot Flash Related Daily Interference Scale: a tool for assessing the impact of hot flashes on quality of life following breast cancer. *J Pain Symptom Manage* 2001; 22: 979–989. DOI: 10.1016/s0885-3924(01)00353-0. [PubMed: 11738160]
9. Carpenter JS, Wu J, Burns DS, et al. Perceived control and hot flashes in treatment seeking breast cancer survivors and menopausal women. *Cancer Nursing* 2012; 35: 195–202. DOI: 10.1097/NCC.0b013e31822e78eb. [PubMed: 21946903]
10. Hunter MS, Nuttall J, Fenlon D. A comparison of three outcome measures of the impact of vasomotor symptoms on women's lives. *Climacteric* 2019; 22: 419–423. DOI: 10.1080/13697137.2019.1580258. [PubMed: 30905177]
11. 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: 877–885. DOI: 10.1097/GME.0000000000000871. [PubMed: 28509814]
12. Newton KM, Carpenter JS, Guthrie KA, et al. Methods for the design of vasomotor symptom trials: the Menopausal Strategies: Finding Lasting Answers to Symptoms and Health network. *Menopause* 2014; 21: 45–58. DOI: 10.1097/GME.0b013e31829337a4. [PubMed: 23760428]
13. Reed SD, LaCroix AZ, Anderson GL, et al. Lights on MsFLASH: a review of contributions. *Menopause* 2020; 27: 473–484. DOI: 10.1097/GME.0000000000001461. [PubMed: 31977667]
14. Freeman EW, Guthrie KA, Caan B, et al. Efficacy of escitalopram for hot flashes in healthy menopausal women: A randomized controlled trial. *JAMA* 2011; 305: 267–274. DOI: 10.1001/jama.2010.2016. [PubMed: 21245182]
15. Cohen LS, Joffe H, Guthrie KA, et al. Efficacy of omega-3 for vasomotor symptoms treatment: A randomized controlled trial. *Menopause* 2014; 21: 347–354. DOI: 10.1097/GME.0b013e31829e40b8. [PubMed: 23982113]
16. Joffe H, Guthrie KA, LaCroix AZ, et al. Low-dose estradiol and the serotonin-norepinephrine reuptake inhibitor venlafaxine for vasomotor symptoms: A randomized clinical trial. *JAMA internal medicine* 2014; 174: 1058–1066. DOI: 10.1001/jamainternmed.2014.1891. [PubMed: 24861828]
17. Newton KM, Reed SD, Guthrie KA, et al. Efficacy of yoga for vasomotor symptoms: A randomized controlled trial. *Menopause* 2014; 21: 339–346. DOI: 10.1097/GME.0b013e31829e4baa. [PubMed: 24045673]
18. Sternfeld B, Guthrie KA, Ensrud KE, et al. Efficacy of exercise for menopausal symptoms: A randomized controlled trial. *Menopause* 2014; 21: 330–338. DOI: 10.1097/GME.0b013e31829e4089. [PubMed: 23899828]
19. Bastien CH, Vallieres A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med* 2001; 2: 297–307. DOI: 10.1016/s1389-9457(00)00065-4. [PubMed: 11438246]
20. Morin CM, Belleville G, Belanger L, et al. The Insomnia Severity Index: Psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep* 2011; 34: 601–608. DOI: 10.1093/sleep/34.5.601. [PubMed: 21532953]
21. Smith MT, Wegener ST. Measures of sleep: The Insomnia Severity Index, Medical Outcomes Study (MOS) Sleep Scale, Pittsburgh Sleep Diary (PSD), and Pittsburgh Sleep Quality Index (PSQI). *Arthritis and Rheumatism (Arthritis Care & Research)* 2003; 49: S184–S196. DOI: 10.1002/art.11409.
22. Buysse DJ, Reynolds CF, Monk TH, et al. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Res* 1989; 28: 193–213. DOI: 10.1016/0165-1781(89)90047-4. [PubMed: 2748771]

23. Otte JL, Rand KL, Landis CA, et al. Confirmatory factor analysis of the Pittsburgh Sleep Quality Index in women with hot flashes. *Menopause* 2015; 22: 1190–1196. DOI: 10.1097/GME.0000000000000459. [PubMed: 25944520]
24. Overholser BR, Sowinski KM. Biostatistics primer: part 2. *Nutr Clin Pract* 2008; 23: 76–84. DOI: 10.1177/011542650802300176. [PubMed: 18203967]
25. Stone AA, Shiffman S, Schwartz JE, et al. Patient non-compliance with paper diaries. *BMJ* 2002; 324: 1193–1194. DOI: 10.1136/bmj.324.7347.1193. [PubMed: 12016186]
26. Chen CX, Ofner S, Bakoyannis G, et al. Symptoms-Based Phenotypes Among Women With Dysmenorrhea: A Latent Class Analysis. *West J Nurs Res* 2018; 40: 1452–1468. DOI: 10.1177/0193945917731778. [PubMed: 28914180]
27. Guttuso T Jr., DiGrazio WJ, Reddy SY. Review of hot flash diaries. *Maturitas* 2012; 71: 213–216. DOI: 10.1016/j.maturitas.2011.12.003. [PubMed: 22230663]
28. Sievert LL. Variation in sweating patterns: implications for studies of hot flashes through skin conductance. *Menopause* 2007; 14: 742–751. DOI: 10.1097/gme.0b013e3180577841. [PubMed: 17519802]
29. Sievert LL, Begum K, Sharmeen T, et al. Patterns of occurrence and concordance between subjective and objective hot flashes among Muslim and Hindu women in Sylhet, Bangladesh. *Am J Hum Biol* 2008; 20: 598–604. DOI: 10.1002/ajhb.20785. [PubMed: 18461600]
30. Sievert LL, Morrison L, Brown DE, et al. Vasomotor symptoms among Japanese-American and European-American women living in Hilo, Hawaii. *Menopause* 2007; 14: 261–269. DOI: 10.1097/01.gme.0000233496.13088.24. [PubMed: 17151510]
31. Sievert LL, Obermeyer CM. Symptom clusters at midlife: a four-country comparison of checklist and qualitative responses. *Menopause* 2012; 19: 133–144. DOI: 10.1097/gme.0b013e3182292af3. [PubMed: 22042326]

Table 1.

Baseline Pooled Sample Demographics, VMS, and Sleep (N=899)

Demographics	n	%	Mean	SD
Age			54.5	3.8
Race				
White	533	59.3		
African American	304	33.8		
Asian or Pacific Islander	20	2.2		
Native American	11	1.2		
Hispanic	7	0.8		
Undisclosed	24	2.7		
Menopause Status				
Postmenopausal	745	82.9		
Late transition	137	15.2		
Early transition	17	1.9		
Smoking				
Never	505	56.2		
Past	255	28.4		
Current	134	14.9		
Unknown	5	0.6		
Body Mass Index			28.0	5.9
<25	295	32.8		
25 - <30	323	35.9		
30	273	30.4		
Unknown	8	0.9		
MsFLASH Trial				
MsFLASH 01	205	22.8		
MsFLASH 02	355	39.5		
MsFLASH 03	339	37.7		
VMS and Sleep				
VMS frequency per day	899		8.3	4.9
VMS severity (1–3)	899		2.1	0.5
VMS bother (1–4)	899		3.0	0.5
VMS interference (HFRDIS)	852		3.5	2.3
VMS interference (HFI)	885		4.7	2.5
Insomnia severity (ISI)	884		11.4	5.8
Sleep quality and disturbance (PSQI)	862		7.8	3.5

HFI = Hot Flash Interference Scale 3 items, HFRDIS= Hot Flash Related Daily Interference Scale 10 items; ISI=Insomnia Severity Index; PSQI = Pittsburgh Sleep Quality Index; SD=standard deviation; VMS=vasomotor symptoms.

Table 2.

Pearson's Correlations Among VMS and Sleep Outcomes at Both Timepoints

	Outcome (Measure)	VMS Frequency	VMS Severity	VMS Bother	HFRDIS	HFI
Pooled Sample Baseline (n=899)	VMS frequency (diary)	--				
	VMS severity (diary)	0.35				
	VMS bother (diary)	0.35	0.92			
	VMS interference (HFRDIS)	0.22	0.41	0.44		
	VMS interference (HFI)	0.21	0.37	0.41	0.90	
	Insomnia severity (ISI)	0.07*	0.21	0.24	0.47	0.54
	Sleep quality/disturbance (PSQI)	0.07 ^{ns}	0.20	0.23	0.38	0.44
Pooled Sample Post-Treatment (n=899)	VMS frequency (diary)	--				
	VMS severity (diary)	0.44				
	VMS bother (diary)	0.44	0.90			
	VMS interference (HFRDIS)	0.40	0.45	0.47		
	VMS interference (HFI)	0.40	0.45	0.48	0.92	
	Insomnia severity (ISI)	0.19	0.30	0.33	0.45	0.50
	Sleep quality/disturbance (PSQI)	0.16	0.24	0.26	0.35	0.41
Placebo Pills Post-Treatment (n=389)	VMS frequency (diary)	--				
	VMS severity (diary)	0.38				
	VMS bother (diary)	0.39	0.90			
	VMS interference (HFRDIS)	0.31	0.44	0.45		
	VMS interference (HFI)	0.29	0.42	0.43	0.92	
	Insomnia severity (ISI)	0.12*	0.24	0.27	0.45	0.52
	Sleep quality/disturbance (PSQI)	0.06 ^{ns}	0.16 ^{**}	0.19	0.31	0.40

All correlations $p < .001$ except as noted.* $p < .05$,** $p < .01$,^{ns} = non-significant.

HFI = Hot Flash Interference Scale 3 items, HFRDIS= Hot Flash Related Daily Interference Scale 10 items; ISI=Insomnia Severity Index; PSQI = Pittsburgh Sleep Quality Index; VMS=vasomotor symptoms.

Post-Treatment Minus Baseline Change in VMS and Sleep Outcomes by Satisfaction with VMS Control/Relief and Desired Continuation

Table 3.

Outcome Change	Satisfied with VMS control/relief				Desire to continue study medication/pills					
	Yes (n=426)		No (n=426)		Yes (n=385)		No (n=457)			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
VMS frequency per day	-4.9	3.8	-1.5	3.3	<0.001	-4.7	4.2	-1.9	3.2	<0.001
VMS severity (1-3)	-0.6	0.6	-0.2	0.5	<0.001	-0.6	0.5	-0.2	0.5	<0.001
VMS bother (1-4)	-0.8	0.7	-0.2	0.5	<0.001	-0.8	0.7	-0.3	0.6	<0.001
VMS interference (HFRDIS)	-2.3	2.3	-0.7	1.9	<0.001	-2.3	2.4	-0.9	1.8	<0.001
VMS interference (HFI)	-3.0	2.6	-1.1	2.2	<0.001	-3.0	2.7	-1.2	2.2	<0.001
Insomnia severity (ISI)	-5.1	5.0	-2.5	4.3	<0.001	-5.1	5.2	-2.7	4.3	<0.001
Sleep quality/disturbance (PSQI)	-2.5	3.0	-1.3	2.7	<0.001	-2.7	3.0	-1.3	2.7	<0.001

HFI = Hot Flash Interference Scale 3 items, HFRDIS= Hot Flash Related Daily Interference Scale 10 items; ISI=Insomnia Severity Index; PSQI = Pittsburgh Sleep Quality Index; SD=standard deviation; VMS=vasomotor symptoms.

Table 4. Post-Treatment Minus Baseline Change in VMS and Sleep Outcomes by Impression of Study Benefits

Outcome Change	Beneficial, minimal side effects (n=419)		Beneficial, not worth side effects (n=51)		No benefit (n=377)		p-value
	Mean	SD	Mean	SD	Mean	SD	
VMS frequency per day	-4.6 ^a	4.1	-3.6 ^a	4.2	-1.4 ^b	3.0	<0.001
VMS severity (1-3)	-0.6 ^a	0.5	-0.5 ^a	0.5	-0.1 ^b	0.5	<0.001
VMS bother (1-4)	-0.8 ^a	0.7	-0.7 ^a	0.7	-0.2 ^b	0.5	<0.001
VMS interference (HFRDIS)	-2.3 ^a	2.3	-1.6 ^b	1.9	-0.6 ^c	1.8	<0.001
VMS interference (HFI)	-3.0 ^a	2.6	-2.0 ^b	2.5	-1.0 ^c	2.2	<0.001
Insomnia severity (ISI)	-2.5 ^a	3.1	-1.8 ^{a,b}	2.8	-1.3 ^b	2.7	<0.001
Sleep quality/disturbance (PSQI)	-5.1 ^a	5.1	-2.5 ^b	4.5	-2.6 ^b	4.3	<0.001

a, b, c. Differing superscripts within a row denote significant group differences in post-hoc pairwise comparisons.

HFI = Hot Flash Interference Scale 3 items, HFRDIS= Hot Flash Related Daily Interference Scale 10 items; ISI=Insomnia Severity Index; PSQI = Pittsburgh Sleep Quality Index; SD=standard deviation; VMS=vasomotor symptoms.

Table 5. Post-Treatment Minus Baseline Change in VMS and Sleep Outcomes by Perceived Pill Group

Outcome Change	Active pill (n=331)		Inactive or placebo pill (n=274)		Don't know (n=250)		p-value
	Mean	SD	Mean	SD	Mean	SD	
VMS frequency per day	-4.4 ^a	4.0	-1.3 ^b	3.2	-3.4 ^c	3.8	<0.001
VMS severity (1-3)	-0.6 ^a	0.6	-0.2 ^b	0.5	-0.4 ^c	0.5	<0.001
VMS bother (1-4)	-0.7 ^a	0.7	-0.2 ^b	0.5	-0.5 ^c	0.6	<0.001
VMS interference (HFRDIS)	-2.0 ^a	2.2	-0.9 ^b	1.8	-1.6 ^c	2.4	<0.001
VMS interference (HFI)	-2.7 ^a	2.7	-1.2 ^b	2.2	-2.1 ^c	2.6	<0.001
Insomnia severity (ISI)	-2.4 ^a	2.9	-1.3 ^b	2.8	-1.8 ^b	3.0	<0.001
Sleep quality/disturbance (PSQI)	-4.4 ^a	5.0	-2.8 ^b	4.6	-4.0 ^a	4.8	<0.001

a, b, c. Differing superscripts within a row denote significant group differences in post-hoc pairwise comparisons.

HFI = Hot Flash Interference Scale 3 items, HFRDIS= Hot Flash Related Daily Interference Scale 10 items; ISI=Insomnia Severity Index; PSQI = Pittsburgh Sleep Quality Index; SD=standard deviation; VMS=vasomotor symptoms.