



Published in final edited form as:

Psychosom Med. 2015 ; 77(2): 167–175. doi:10.1097/PSY.000000000000143.

Placebo Improvement in Pharmacologic Treatment of Menopausal Hot Flashes: Time Course, Duration, and Predictors

Ellen W. Freeman, PhD¹, Kristine E. Ensrud, MD, MPH², Joseph C. Larson, MS³, Katherine A. Guthrie, PhD³, Janet S. Carpenter, PhD, RN⁴, Hadine Joffe, MD, MSc⁵, Katherine M. Newton, PhD⁶, Barbara Sternfeld, PhD⁷, and Andrea Z. LaCroix, PhD⁸

¹Department of Obstetrics/Gynecology and Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

²Division of Epidemiology and Community Health, Department of Medicine, Veterans Affairs Health System and Department of Medicine, University of Minnesota, Minneapolis, MN

³MsFlash Data Coordinating Center, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA

⁴School of Nursing, Indiana University, Indianapolis, IN

⁵Department of Psychiatry, Brigham and Women's Hospital and Massachusetts General Hospital, and Department of Psychosocial Oncology and Palliative Care, Dana Farber Institute, Harvard Medical School, Boston, MA

⁶Group Health Research Institute, Seattle, WA

⁷Division of Research, Kaiser Permanente of Northern California, Oakland, CA

⁸Women's Health Center, Family and Preventive Medicine, University of California San Diego, La Jolla, CA

Abstract

Objectives—This study characterized the time course, duration of improvement and clinical predictors of placebo response in treatment of menopausal hot flashes.

Methods—Data were pooled from two trials conducted in the MsFLASH network, providing a combined placebo group (N=247) and a combined active treatment group (N=297). Participants recorded hot flash frequency in diaries twice daily during treatment (week 0-8) and subsequent follow-up (week 9-11). The primary outcome variable was clinically significant improvement, defined as $\geq 50\%$ decrease in hot flash frequency from baseline and calculated for each week in the study. Subgroups were defined a priori using standard clinical definitions for significant improvement and partial improvement. Clinical and demographic characteristics of the participants were evaluated as predictors of improvement.

Results—Clinically significant improvement with placebo accrued each treatment week, with 33% significantly improved at week 8. Of placebo responders who were improved at both weeks 4

Conflicts of Interest Disclosures:
No other disclosures were reported.

and 8, 77% remained clinically improved at week 11 after treatment ended. Independent predictors of significant placebo improvement in the final multivariable model were African American race (OR 5.61, 95% CI: 2.41-13.07, $p < 0.001$); current smokers (OR 2.30, 95% CI: 1.05-5.06, $p = 0.038$); and hot flash severity in screening (OR 1.45, 95% CI: 1.00-2.10, $p = 0.047$).

Conclusions—Clinically significant improvement with placebo accrued throughout treatment with a time course similar to improvement with active drug. A meaningful number of participants in the placebo group sustained a clinically significant response after stopping placebo pills. The results suggest that non-specific effects are important components of treatment and warrant further studies to optimize their contributions in clinical care.

Keywords

placebo response; placebo improvement; hot flash treatment; menopause

Introduction

Nonspecific effects play a significant role in symptom improvement in a wide range of medical conditions, particularly those that have components of pain or psychological symptoms (1). A placebo elicits nonspecific effects and may do so by administration of an inert substance, a sham treatment or any treatment that is supported by belief that it will work (2). As discussed by Roberts (3), there is no single placebo effect and no consensus on a definition of placebo, with the result that identification and understanding of non-specific effects in treatment responses must be evaluated in the context of specific treatments.

Placebo effects are expected in hot flash treatment. A Cochrane review of 9 placebo-controlled trials of oral estrogen therapy for menopausal hot flashes clearly indicated the efficacy of hormone therapy but also showed that those with placebo treatment had a mean reduction of 58% in hot flash frequency (4). A pooled analysis of 10 clinical trials of non-hormonal pharmacologic therapies for menopausal hot flashes showed that responses to placebo ranged from 27% to 52% (5).

There is a prevailing view that a placebo response in medical treatment is characterized by rapid but transient improvement (6). While brief and partial responses to placebo treatment are clearly observed in clinical trials, there is also evidence of sustained improvement due to nonspecific effects, although a specific time course that clearly characterizes placebo responses has not been identified.

The considerable variability in the rates of placebo response is dependent on the type of disorder and numerous other factors such as the severity of symptoms, patient characteristics, particularly anxiety, mood or pain levels, treatment expectations of the clinician and patient, suggestibility, spontaneous recovery and regression to the mean, but none have exclusively or consistently characterized a placebo response. Data show that when information provided to patients was positive rather than negative or neutral, positive treatment effects significantly increased for both placebo and active drug treatment (7, 8). Other data show that placebos given after active drug treatment maintained the active drug outcome with reduced side effects (9). When drug administration was concealed to eliminate

awareness of when treatment was given (as with computer-controlled infusion), there was a significant reduction in response to placebo and also to active drug; drug doses had to be doubled to achieve the same results (10). In a trial of phytotherapy for menopausal hot flashes, predictors of responses to placebo and active treatment differed, suggesting that different mechanisms may underlie these responses (11).

The MsFlash network previously reported the efficacy of treatment for menopausal hot flashes in two randomly-assigned, double-blind trials of medications compared to matched pill placebo and had a moderate placebo response rate of approximately 31%-36% (12, 13). The objectives of this secondary analysis of data from these trials were to: 1) identify the time course of the placebo improvement and compare the time course of improvement between drug and placebo; 2) identify the extent of sustained placebo improvement; and 3) estimate associations of clinical characteristics with placebo improvement.

METHODS

Study Design

This is a secondary analysis of placebo data pooled from two clinical trials in the MsFlash (Menopausal Strategies: Finding Lasting Answers to Symptoms and Health) Network that compared active medications for menopausal hot flashes with identical-appearing, pharmacologically inert placebo pills. In the primary trials, the efficacy and tolerability of the SSRI escitalopram (10-20 mg/d) was evaluated in 205 participants (12), and the efficacy of low-dose oral 17-beta estradiol (0.5 mg/d) and the SNRI venlafaxine XR (75 mg/d) was evaluated in 339 participants (13). In both trials, participants completed daily diaries for at least 3 weeks of screening, were randomized to 8 weeks of double-blinded treatment with active medication or matching placebo pills, and recorded vasomotor symptoms in daily diaries throughout the study, as previously described (14). Treatment assignment was double blinded, using a computerized randomization algorithm (15). Participants took 1 pill per day for 8 weeks, with an increase to 2 pills/day for weeks 5-8 in the escitalopram study if symptoms were not improved at week 4. The study procedures were otherwise identical in the drug and placebo treatment arms throughout the 8 weeks of treatment. After 8 weeks of double-blind treatment, the pills were stopped for those taking 1 pill/day and tapered to 1 pill/day and then stopped for those who increased the dose in the escitalopram study. A brief telephone contact was conducted at 3 or 4 weeks after treatment endpoint to query symptom status. The daily diary reports for week 11 (3 weeks after treatment endpoint) provided the information on hot flash frequency at week 11. The trials were approved by the Institutional Review Boards at each participating site, and participants provided written informed consent.

Participant Selection

Participants were recruited for the clinical trials between July 2009 and October 2012, primarily by means of purchased mailing lists and health-plan enrollment files. In the escitalopram trial, the hot flash criteria required at least 28 hot flashes or night sweats per week as recorded in diaries twice daily for two screening weeks, ratings of bothersome or severe hot flashes/night sweats on 4 or more days or nights in each week, and no decrease

hot flash frequency in the third screening week greater than 50% from the mean frequency in the first 2 screening weeks. In the estradiol/venlafaxine trial, the hot flash criteria were the same, with exception of hot flash frequency, which required at least 14 hot flashes/night sweats per week in the first two screening weeks. Both trials included ages 40-62 years, required general good health as determined by medical history, a brief physical exam and standard blood tests, and being in the menopause transition (amenorrhea ≥ 60 days in the past year) or postmenopause (defined as ≥ 12 months since the last menstrual period, or bilateral oophorectomy, or FSH >20 mIU/mL and estradiol ≤ 50 pg/mL in the absence of a reliable menstrual marker (e.g., hysterectomy with ovarian preservation, progesterone-releasing intra-uterine device or endometrial ablation). Exclusion criteria included use of prescription, over-the-counter or herbal therapies for hot flashes in the past 30 days, hormones, hormonal contraception, selective estrogen receptor modulators (SERMS) or aromatase inhibitors in the past 2 months, current severe medical illness, major depressive episode, drug or alcohol abuse in the past year, suicide attempt in the past 3 years, lifetime diagnosis of bipolar disorder or psychosis, uncontrolled hypertension, history of endometrial or ovarian cancer, myocardial infarction, angina or cerebrovascular events, or other preexisting medical conditions.

Data Collection

After a brief telephone screen, eligible volunteers were mailed a baseline questionnaire to assess self-reported health and demographics and daily diaries to record frequency, severity and bother of hot flashes each morning and evening for 2 weeks. After clinic review of these data, women who continued to meet eligibility criteria were scheduled for 2 clinic screening visits within a 2 to 3-week interval. Participants continued to rate hot flashes twice daily and met hot flash eligibility criteria for a total of 3 screen weeks. At the second screen visit, eligible women were randomly assigned double blind to a treatment arm for 8 weeks, using a dynamic randomization algorithm (15) with stratification for clinic site in both studies and for race in the escitalopram study. Telephone contact was made one week after randomization to assess protocol adherence and adverse events. A clinic visit or telephone contact was scheduled at 4 weeks, and a clinic visit was scheduled at 8 weeks after randomization. Participants completed self-report questionnaires at the screen visits and treatment week 8 and continued to record hot flash information in diaries twice daily. Daily diaries were continued throughout treatment and the brief follow-up period.

Study Variables

The primary outcome variable was clinically significant improvement, defined as $\geq 50\%$ decrease in hot flash frequency from baseline and calculated at each treatment week. Hot flash frequency was the total number of hot flashes and night sweats in a 24-hour period, calculated from the daily diaries of each subject; the mean of daily totals was calculated for each subject for each week. Baseline hot flash frequency was the mean of daily totals reported in the first two screening weeks.

Predictor variables were baseline characteristics of hot flashes and standard clinical and demographic factors. These included the frequency, variability, severity and bother of hot flashes in the screen weeks, menopausal status (transition or postmenopause), duration of

hot flashes (years), body mass index (BMI, kg/m²), current smoking (yes, no), alcohol use (≥ 1 /day), anxiety (HSCL mean score) (16), depression (PHQ-9) (17), employed full time (yes, no), education (\leq HS, all other), adverse events (yes, no, week 4), age and self-reported race (white, African American, other). Hot flash variability in the screen period was defined as the participant's deviation around her mean hot flash frequency in the first two screen weeks. Scores for hot flash severity (rated 1 to 3, mild to severe) and bother (rated 1 to 4 (none to a lot) were calculated in the same manner as hot flash frequency, using the mean of daily ratings in the first two screen weeks; baseline scores were the means of the daily ratings for the first two screen weeks.

Statistical Analysis

Data from the two MsFlash trials were pooled to provide a combined placebo group (N=247) and a combined active treatment group (N=297). To determine if the patterns of clinically significant improvement in hot flashes over time differed between the two protocols, a logistic model was fit to test the interaction between protocol and study time separately for placebo and drug participants.

Baseline characteristics of the combined active and placebo groups that were pooled from the two studies are presented with means and standard deviations for continuous variables and frequencies and percentages for categorical variables. The statistical significance of differences between groups was calculated with t or chi-square tests as appropriate (shown in Table 1). The percent reduction in hot flash frequency from baseline was calculated for each participant in each treatment arm at each week of the trial. The percentage of participants who reported clinically significant improvement in hot flashes from baseline is shown by intervention group for each treatment week (Figure 1). Logistic regression was used to test the effect of time (weeks 1-8), intervention (active versus placebo) and their interaction. Among those who ever had $\geq 50\%$ improvement from baseline, the percentages at the first clinically significant improvement are presented for each treatment week (Table 2). Placebo participants who reported clinically significant improvement in hot flash frequency by week 4 were divided into early responders (clinical improvement at week 1 or 2) and later responders (clinical improvement at week 3 or 4), with the percentages who remained clinically improved at weeks 5, 6, 7, 8 presented in Figure 2. Differences between the percentages of early and later responders were evaluated using a repeated measures logistic regression model of clinical improvement as a function of response type (early vs later) and week (5, 6, 7, 8). Logistic regression was used to model clinically significant improvement with placebo at week 8 as a function of a series of the potential predictors listed in Table 1. Predictor variables that were associated with improvement at $p \leq 0.20$ in univariate models were included in multivariable analysis. Inclusion in the final multivariable model was guided by statistical significance at $p \leq 0.05$ (shown in Table 3). This statistics-based approach for selecting covariates was used because we had no a priori theoretical hypothesis for which variables would be confounding or effect-modifying, and at the same time we aimed to prevent over-fitted models. Intent-to-treat analysis included all available data of the study participants. All models were adjusted for protocol and clinical site.

Statistical power calculations were computed using publicly available software at <swogstat.org>. Assumptions were based on data in the sample and included number of participants, placebo improvement of 50% or more and a binomial test with 2-sided significance level of 0.05. Given 225 participants, there was 80% power to detect an odds ratio of 2.3 or higher for factors with a prevalence of 35% and an odds ratio of 2.8 or higher for factors with a prevalence of 20%. Analyses were conducted using SAS Version 9.3 (SAS Institute, Inc., Cary, NC). Statistical tests were 2-sided and considered significant at $p < 0.05$.

RESULTS

Before pooling the data, the patterns of clinical improvement in hot flashes over the trial period for the placebo and active drug participants did not differ between the two protocols (interaction for active drug $p = 0.13$, for placebo $p = 0.92$). Table 1 shows that there were no significant differences in baseline characteristics between the combined placebo group and the combined active treatment group. In the combined placebo group, the mean frequency of hot flashes at baseline was 8.5 (SD 5.0)/day, 74% were postmenopausal, and the mean age was 54.3 (SD 3.8) years.

Time course of placebo improvement

The incidence of clinically significant improvement was greater at each of the 8 treatment weeks in both the placebo and active drug groups (Figure 1). In the placebo group, 10% improved at week 1, and 33% were improved at week 8. In the active drug group, 17% improved at week 1, and 54% were improved at week 8. Improvement in the active drug group was of greater magnitude as expected ($P < 0.001$), but there was no significant interaction between time and treatment group ($P = 0.88$), indicating that the time course of improvement was similar in the active drug and placebo groups.

First clinically significant improvement

The first occurrence of clinically significant improvement in the placebo group was greatest at treatment week 1 (20%), but clinically significant improvement continued to occur throughout the 8 weeks of treatment (Table 2). In the placebo group, the cumulative proportion of first improvement increased to 68% at week 4, with the remaining 32% of the clinically improved having their first occurrence of improvement between weeks 5 and 8. Again, the pattern was similar to the active drug group, where the first clinically significant improvement at the same time points was 25%, 73% and 27%, respectively.

We further investigated whether the earliest placebo improvers (clinically significant improvement in weeks 1-2) compared to later improvers (first clinically significant improvement in weeks 3-4) remained improved in weeks 5-8. Figure 2 indicates that more early improvers were also improved in weeks 5-8, although the comparison with later improvers did not reach significance ($P = 0.076$). The proportion of early improvers who remained improved appeared to decrease slightly at week 8 (Figure 2), but these data do not indicate whether this decrease would continue with extended time.

Sustained placebo improvement

Further analysis of improvement subgroups (defined a priori as sustained improvement (clinically significant improvement in hot flash frequency of 50% or more from baseline at *both* week 4 and week 8), partial improvement (30% to 50% improvement at week 4 and/or week 8) or no improvement (<30% improvement at week 4 and week 8) indicated that 20% of the placebo group (46/231) had sustained improvement of 50% or more at *both* week 4 and week 8. Another 13% were significantly improved only at week 8, 5% were significantly improved only at week 4, and 25% had partial improvement at week 4 or week 8 or both. Thirty-eight percent had no improvement with placebo throughout the study.

At the week 11 follow-up, we investigated whether the participants who *sustained* improvement on placebo pills remained improved after stopping or tapering placebo pills at week 8. Based on daily diary reports, 77% (34/44) of the sustained improvers remained significantly improved at week 11; 6 (14%) declined to partial improvement (30%-50% improvement from baseline); and 4 (9%) were no longer improved (<30% improvement from baseline). Of 29 participants at week 11 who were significantly improved at week 8 but not week 4, 16 (55%) remained significantly improved at week 11, 8 (28%) declined to partial improvement, and 5 (17%) were no longer improved.

Predictors of placebo improvement

Table 3 shows associations of the baseline characteristics with clinically significant improvement at week 8 in the placebo group. Significant predictors of improvement in univariate analysis included being African American (OR 7.52, 95% CI: 3.31 - 17.08, $p < 0.001$), current smoking (OR 3.50, 95% CI: 1.707-23, $p < 0.001$), having greater severity of hot flashes in the screen period (OR 1.79, 95% CI: 1.27-2.52, $p < 0.001$), and younger age at screening (OR 0.62, 95% CI: 0.42-0.93, $p = 0.021$). In the final multivariable model, the significant independent predictors of placebo improvement after adjustment for the presence of all other variables in the model were African American race, current smoking and greater severity of hot flashes in the screen period (Table 3). The odds of improvement with placebo in adjusted analysis were more than 5 1/2 times higher in African American compared to white women, over 2 times higher for smokers compared to non-smokers, and approximately 1 1/2 times higher for each one point increase in the average severity of hot flashes in the screen period.

The same model was repeated to estimate associations of the baseline characteristics for the outcome of sustained improvement (defined as clinically significant improvement at both weeks 4 and 8). Both the univariate associations and the final reduced model were similar to the results shown in Table 3, which were for clinically significant improvement at week 8.

DISCUSSION

This study showed that clinically significant improvement with placebo gradually increased through 8 weeks of double-blind treatment, with 33% of the placebo group improved at treatment endpoint. While the magnitude of improvement with placebo was significantly less than improvement with active drug, as was previously reported in the primary clinical

trials (12, 13), the time course of improvement with placebo pills was markedly similar to that of active drug, with a clinically meaningful response accruing throughout the treatment period. Moreover, it appeared that early improvers (clinically improved in weeks 1 or 2) largely sustained significant improvement throughout the study interval and did not swiftly return to unimproved. Of the participants who were significantly improved at both week 4 and week 8, 77% remained clinically improved after stopping the study pills. This time course of improvement challenges the belief that placebo response is typically characterized by rapid and transient improvement that is quickly lost (6) inasmuch as a meaningful proportion of participants appeared to sustain clinical improvement during and after treatment with placebo pills.

Rates of placebo improvement are known to vary widely and are greatly influenced by context and a variety of other sources that limit comparisons across studies. Patient characteristics and settings, treatment expectations, suggestibility, psychological states, natural improvement with time and regression to the mean have all been associated with placebo response rates (3, 18). The rate of placebo improvement in the present trials ranged from 31% to 36%, which was well within the bounds of moderate placebo improvement. In 10 trials of pharmacologic treatments for hot flashes, placebo improvement ranged from 21%-52%. In a pooled analysis of these trials, 37% of placebo patients had at least a 50% reduction in hot flash scores (5). A Cochrane review of estrogen therapy trials indicated that the mean reduction in hot flash frequency was 58% with placebo treatment (4).

In this study, the independent contributors to significant placebo improvement in the final multivariable model were African American race, current smoking and greater severity of hot flashes in the screen period. Data indicate that African American women have the highest rate of hot flash reporting (19-22), but we know of no other studies that evaluated race as a predictor of placebo improvement. In a study of hot flashes and race/ethnicity from the Study of Women's Health Across the Nation, African Americans had greater sensitivity to somatic symptoms (19), but it is not known whether greater symptom sensitivity extends to greater placebo improvement.

Current smoking is associated with greater frequency of hot flashes (23-26), but we know of no studies that evaluated smoking as a predictor of placebo response. Other data suggest that the association of smoking with hot flashes may involve alterations in neurotransmitter metabolism in the dopaminergic system, which has also been considered a candidate neurotransmitter both for the development of hot flashes (23) and for high placebo responses (27, 28). Thus we speculate that the association between smoking and placebo improvement may involve the dopaminergic system, but studies to identify these associations are needed.

Symptom severity before treatment has been frequently identified as a predictor of placebo response, usually indicating that those with less severe symptoms were more likely to respond to placebo (18, 29, 30). However, the opposite association was observed in the present study. We speculate that this was due in part to the narrow range of the symptom severity scale and to the exclusion of women with inconsistent or low ratings of hot flash severity in the screen period. The exclusion of participants whose hot flashes fluctuated widely or improved during the screen period may also be the reason that our hypothesis that

placebo improvement could be predicted by greater within-subject variability in hot flash frequency during the screen period was not supported.

Whether mechanisms underlying responses to placebo and active drug differ is not well understood. One study of menopausal hot flashes observed differences between the predictors of response to placebo and active drug, which suggested that distinct mechanisms may be operating in the two treatment conditions (11). However, other data suggest that responses to placebo and active drug have shared biological mechanisms (8), particularly if the placebo effect is considered a neurobiological phenomenon (28). Further studies to identify neurobiological mechanisms that may underlie responses to both placebo and drug are needed to elucidate the complexity of placebo improvement.

Although the study cannot differentiate between placebo response, spontaneous improvement and regression to the mean, it is important to consider the possibilities of spontaneous improvement and regression to the mean as components of placebo improvement. In further studies of placebo effects, a second control group that is not given either active medication or placebo could better address the question of regression to the mean. Although the mean natural duration of menopausal hot flashes is considerable, ranging from about 4 years to more than 10 years (31), and participants were selected for stable ratings of hot flashes in the screen period, it is possible that variations in hot flash frequency could account for placebo improvement. It should also be noted that one protocol required a relatively high frequency of hot flashes in the screen period, suggesting the possibility that subsequent decreases were a return to more usual levels of hot flashes or that hot flashes improved with time for reasons other than the direct effects of study pills. The sustained improvement observed for some participants after discontinuing the placebo pills could also be interpreted as a possible regression to the mean, inasmuch as placebo effects after discontinuing pills are believed to be relatively rare.

Other limitations of the study include the 8-week treatment duration, which is a common duration for acute treatment trials to identify drug response relative to placebo but limits evaluation of long-term improvement. However, the present findings were supported in a controlled trial of botanical supplements to reduce hot flashes, where a similar time course of placebo improvement was sustained for 12 months (32). The exclusion of participants who improved or had large fluctuations in symptom levels during the screen period may have resulted in underestimating the incidence of placebo improvement in this study. Possible predictors of placebo improvement were limited to the clinical and demographic characteristics of these generally healthy mid-life women, who volunteered for treatment of menopausal hot flashes; other factors that were not evaluated in this study may be important contributors to placebo response. These findings may not be generalizable to women with other demographic characteristics or treatment conditions, and further studies, particularly those that include neurobiological factors, are needed.

Recognizing placebo phenomena is important for clinical practice, where the power of non-specific effects far exceeds what is commonly accepted (3). The effectiveness of most treatments may be substantially increased by implementing procedures that promote nonspecific effects, among which positive information and expectations about treatments of

both clinicians and patients are the most well-known. This study showed that clinically significant improvement with placebo accrued throughout treatment with a time course similar to improvement with active drug, and the majority of clinically significant improvers on placebo sustained improvement in the follow-up period after stopping placebo pills. Although it is beyond the objectives of this study, the findings of sustained responses to placebo raise the controversial issue of the ethical and practical roles of utilizing placebos in clinical practice, particularly when potential benefits of placebo treatments may outweigh adverse events of active drugs. The results indicate that non-specific effects of treatment are important components of clinical care and warrant further studies to optimize their contributions to clinical improvement.

Acknowledgments

Source of funding: All authors received funding for this study from the National Institutes of Health, under a cooperative agreement issued by the National Institute of Aging to each participating site in the MsFlash network: grant #: UO1AG032656, UO1AG032659, UO1AG032669, UO1AG032682, UO1AG 032699, and UO1AG032700.

Dr. Freeman reports receiving research support from the National Institutes of Health, Forest Laboratories, Inc. and Bionovo. Dr. Ensrud reports receiving research support from the National Institutes of Health and financial compensation for serving as a consultant on the data monitoring committee for Merck Sharpe & Dohme. Dr. Joffe reports receiving research support for the National Institutes of Health and Teva/Cephalon and serving as a consultant for Noven. Dr. LaCroix reports receiving research support from the National Institute of Aging and the National Institutes of Health. Dr. Carpenter, Dr. Guthrie, Dr. Newton and Dr. Sternfeld report receiving research support from the National Institute of Aging.

References

1. Khan A, Bhat A. Is the problem of a high placebo response unique to antidepressant trials? *J Clin Psychiatry*. 2008; 69:1979–80. [PubMed: 19203480]
2. Schweizer E, Rickels K. Placebo response in generalized anxiety: its effect on the outcome of clinical trials. *J Clin Psychiatry*. 1997; 58(suppl 11):30–8. [PubMed: 9363046]
3. Roberts A. The powerful placebo revisited: implications for headache treatment and management. *Headache Quarterly – Current Treatment and Research*. 1994; 5:208–13.
4. MacLennan AH, Broadbent JL, Lester S, Moore V. Oral estrogen and combined oestrogen/progestogen therapy versus placebo for hot flushes. *Cochrane Database of Systematic Reviews*. 2004; (4) CD002978.
5. Loprinzi CL, Sloan J, Stearns V, Slack R, Iyengar M, Diekmann B, Kimmick G, Lovato J, Gordon P, Pandya K, Guttuso T Jr, Barton D, Novotny P. Newer antidepressants and gabapentin for hot flashes: an individual patient pooled analysis. *J Clin Oncology*. 2009; 27:2831–7.
6. Quitkin FM, Stewart JW, McGrath PJ, Nunes E, Ocepek-Welikson K, Tricamo E, Rabkin JG, Klein DF. Further evidence that a placebo response to antidepressants can be identified. *Am J Psychiatry*. 1993; 150:566–70.
7. Colloca L, Miller FG. Harnessing the placebo effect: the need for translational research. *Phil Trans R Soc B*. 2011; 366:1922–30. [PubMed: 21576150]
8. Rief W, Bingel U, Schedlowski M, Enck P. Mechanisms involved in placebo and nocebo responses and implications for drug trials. *Clin Pharmacol Ther*. 2011; 90:722–6. [PubMed: 21975346]
9. Kam-Hansen S, Jakubowski M, Kelley JM, Kirch I, Hoaglin DC, Kaptchuk TJ, Burstein R. Altered placebo and drug labeling changes the outcome of episodic migraine attacks. *SCi Transl Med*. 2014; 6(218):218ra5.
10. Colloca L, Lopiano L, Lanotte M, Benedetti F. Overt versus covert treatment for pain, anxiety and Parkinson's disease. *The Lancet Neurology*. 2004; 3:679–84. [PubMed: 15488461]

11. Van Die MD, Bone KM, Burger HG, Teede HJ. Are we drawing the right conclusions from randomized placebo-controlled trials? A post-hoc analysis of data from a randomized controlled trial. *BMC Medical Research Methodology*. 2009; 9:41. [PubMed: 19549306]
12. Freeman EW, Guthrie KA, Caan B, Sternfeld B, Cohen LS, Joffe H, Carpenter JS, Anderson GL, Larson JC, Ensrud KE, Reed SD, Newton KM, Sherman S, Sammel MD, LaCroix AZ. Efficacy of escitalopram for hot flashes in healthy menopausal women. *JAMA*. 2011; 305:267–74. [PubMed: 21245182]
13. Joffe H, Guthrie KA, LaCroix AZ, Reed SD, Ensrud KE, Manson JE, Newtown KM, Freeman EW, Anderson GL, Larson JC, Hunt J, Shifren J, Rexrode KM, Caan B, Sternfeld B, Carpenter JS, Cohen L. Low-dose estradiol and the serotonin-norepinephrine reuptake inhibitor venlafaxine for vasomotor symptoms a randomized clinical trial. *JAMA Intern Med*. 2014 Epub May 26.
14. Newton KM, Carpenter JS, Guthrie KA, Anderson GL, Caan B, Cohen LS, Ensrud KE, Freeman EW, Joffe H, Sternfeld B, Reed SD, Sherman S, Sammel MD, Kroenke K, Larson JC, LaCroix AZ. Methods for the design of vasomotor symptom trials: the Menopausal Strategies: Finding Lasting Answers to Symptoms and Health network. *Menopause*. 2014; 21:45–58. [PubMed: 23760428]
15. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975; 31:103–115. [PubMed: 1100130]
16. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci*. 1974; 19:1–15. [PubMed: 4808738]
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001; 16:606–13. [PubMed: 11556941]
18. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry*. 2013; 170:723–33. [PubMed: 23318413]
19. Gold EB, Colvin A, Avis N, Bromberger J, Greendale GA, Powell L, Sternfeld B, Matthews K. 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 Pub Health*. 2006; 96:1226–35. [PubMed: 16735636]
20. Miller SR, Gallicchio LM, Lewis LM, Babus JK, Langenberg P, Zacur HA, Flaws JA. Association between race and hot flashes in midlife women. *Maturitas*. 2006; 54:260–9. [PubMed: 16423474]
21. Simpkins JW, Brown K, Bae S, Ratka A. Role of ethnicity in the expression of features of hot flashes. *Maturitas*. 2009; 63:341–6. [PubMed: 19592184]
22. Freeman EW, Sammel MD, Grisso JA, Battistini M, Garcia-Espagna B, Hollander L. Hot flashes in the late reproductive years: risk factors for African American and Caucasian women. *J Women’s Health & Gender-Based Medicine*. 2001; 10:67–76.
23. Butts SF, Freeman EW, Sammel MD, Queen K, Lin H, Rebbeck TR. Joint effects of smoking and gene variants involved in sex steroid metabolism on hot flashes in late reproductive-age women. *J Clin Endocrinol Metab*. 2012; 97:E1032–42. [PubMed: 22466345]
24. Gallicchio L, Miller SR, Visvanathan K, Lewis LM, Babus J, Zacur H, Flaws JA. *Maturitas*. 2006; 53:133–43. [PubMed: 16368467]
25. Whiteman MK, Staropoli CA, Langenberg PW, McCarter RJ, Kjerulff KH, Flaws JA. Smoking, body mass and hot flashes in midlife women. *Obstet Gynecol*. 2003; 101:264–72. [PubMed: 12576249]
26. Zhu BT, Conney AH. Functional role of estrogen metabolism in target cells: review and perspective. *Carcinogenesis*. 1998; 19:1–27. [PubMed: 9472688]
27. Scott DJ, Stohler CS, Egnatuk CM, Wang H, Koeppel RA, Zubieta JK. Placebo and nocebo effects are defined by opposite opioid and dopaminergic responses. *Arch Gen Psychiatry*. 2008; 65:220–31. [PubMed: 18250260]
28. Meissner K, Bingel U, Colloca L, Wager TD, Watson A, Flaten MA. The placebo effect: advances from different methodological approaches. *J of Neuroscience*. 2011; 31:16117–24.
29. Fournier JC, DeRubeis RJ, Hollon SD, Dimijian S, Amsterdam JD, Shelton RC, Fawcett J. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010; 303:47–53. [PubMed: 20051569]

30. Brown WA, Johnson MF, Chen MG. Clinical features of depressed patients who do and do not improve with placebo. *Psychiatry Res.* 1992; 41:203–14. [PubMed: 1594707]
31. Freeman EW, Sammel MD, Sanders RJ. Risk of long-term hot flashes after natural menopause: evidence from the Penn Ovarian Aging cohort. *Menopause.* 2014 Epub Jan 27.
32. Geller SE, Shulman LP, van Breemen RB, Banuvar S, Zhou Y, Epstein G, Hedayat S, Nikolic D, Krause EC, Pierson CE, Bolton JL, Pauli GF, Farnsworth NR. Safety and efficacy of black cohosh and red clover for the management of vasomotor symptoms: a randomized controlled trial. *Menopause.* 2009; 16:1156–66. [PubMed: 19609225]

Abbreviations

MsFlash	Menopausal Strategies: Finding Lasting Answers to Symptoms and Health (an NIH research network)
HF	Hot flashes
OR	Odds ratio
CI	Confidence interval
SSRI	selective serotonin reuptake inhibitor
SNRI	serotonin norepinephrine reuptake inhibitor
XR	extended release
Mg/d	milligrams/day (24 hours)
FSH	follicle stimulating hormone
mIU/mL	milliInternational units per milliliter
pg/mL	picograms per milliliter
SERM	selective estrogen receptor modulator
BMI	body mass index
Kg/m²	kilogram per meter squared
HSCL	Hopkins Symptom Checklist
PHQ-9	Patient Health Questionnaire – 9 items
HS	high school
SD	standard deviation

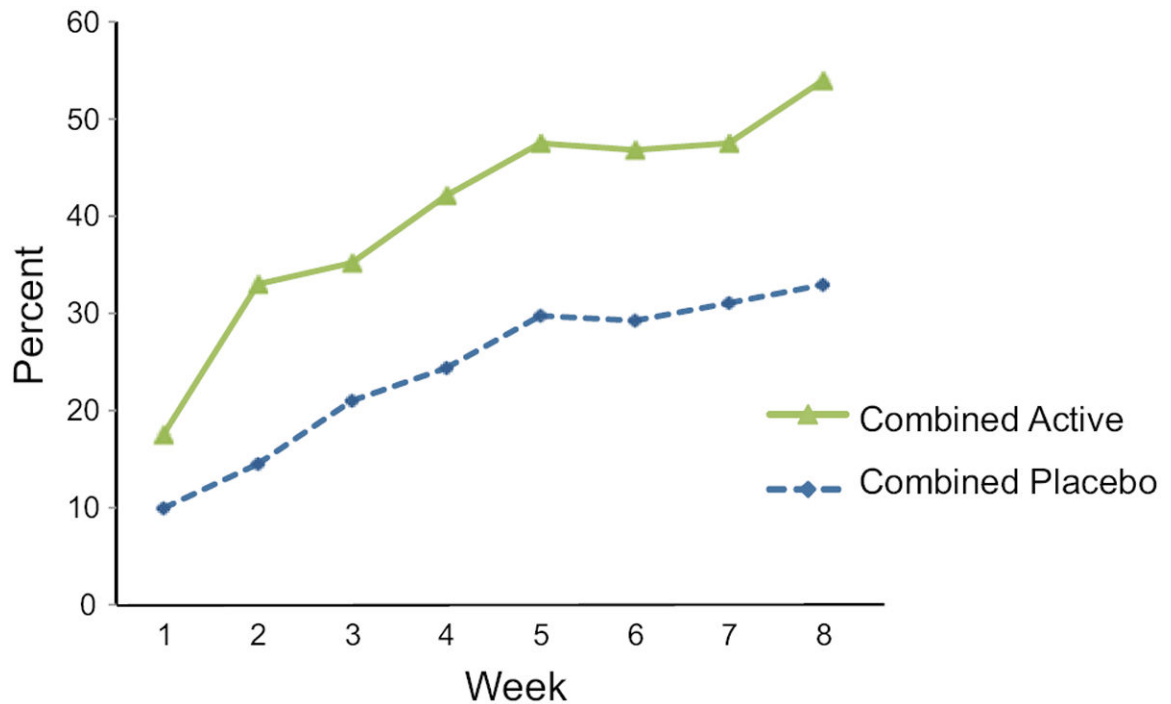


Figure 1. Percent of Participants Reporting Clinically Significant Improvement in Hot Flashes From Baseline by Intervention Arm

Data points are the percent improved at each time point in the placebo group (N=247) and in the active drug group (N=297). Not all participants had data at all weeks.

Repeated measures logistic regression modeled 50% reduction in HF from baseline (yes/no) as a function of intervention (active drug vs. placebo), week (1-8) as a continuous variable and intervention*week, adjusted for study and clinic center.

Main effects of intervention and week, $P < 0.001$ for each; interaction between intervention and week, $P = 0.88$.

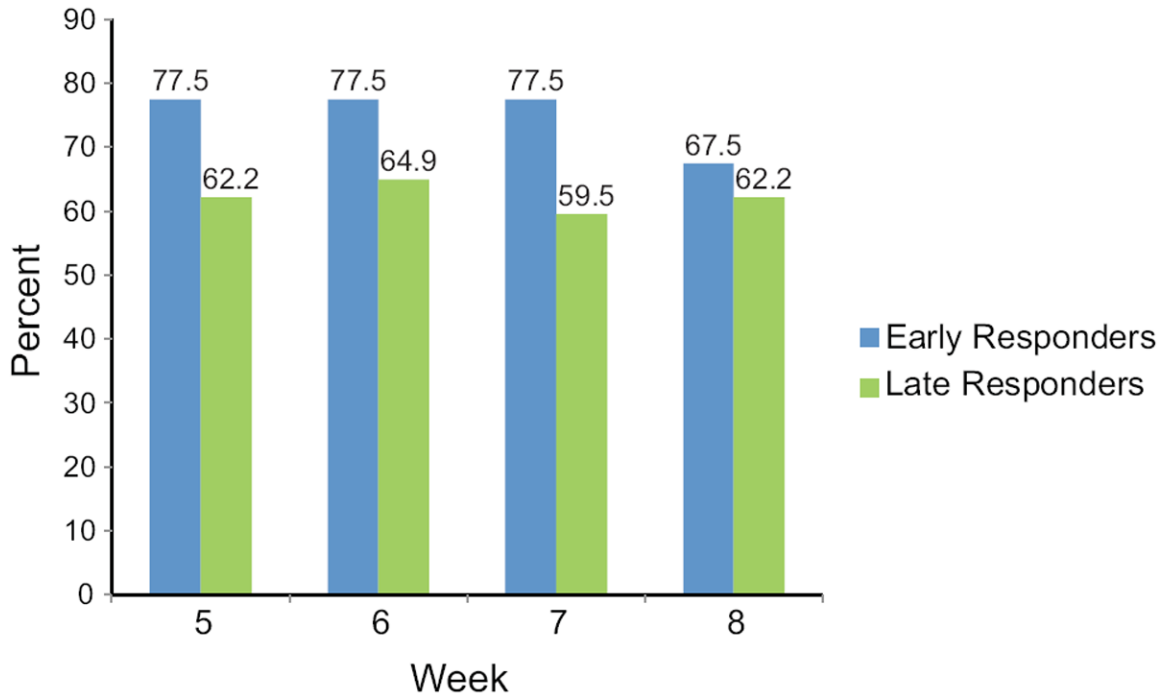


Figure 2. Percent of Early Placebo Responders Reporting Clinically Significant Improvement in Weeks 5-8

P=0.076 for the main effect of difference in clinically significant improvement in weeks 5-8 compared between early responders (responded in weeks 1-2, n=40) versus later responders (responded in weeks 3-4, n=37).

(P values for responder group × week: week 5: P=0.095; week 6: P=0.14; week 7: P=0.045; week 8: P=0.43).

P values are from a repeated measures logistic regression model modeling 50% reduction in HF from baseline (yes/no) as a function of response type (early/late), adjusted for week (5,6,7,8) and clinic center.

Table 1

Baseline Characteristics by Intervention Arm

Variable	Combined Active N=297		Combined Placebo N=247		p-value ^d
	n	%	N	%	
Age, mean (SD)	54.4 (4.1)		54.3 (3.8)		0.88
Ethnicity					0.86
White	166	55.9	139	56.3	
African American	117	39.4	94	38.1	
Other / Unknown	14	4.7	14	5.7	
Menopause Status					0.86
Menopause transition	53	17.8	40	16.2	
Postmenopause	216	72.7	182	73.7	
Indeterminate ²	28	9.4	25	10.1	
Years of hot flashes, mean (SD)	6.5 (6.0)		6.3 (5.5)		0.66
PHQ-9 Depression >=5	83	27.9	61	24.7	0.39
SCL Anxiety > 0.3	70	23.6	66	26.7	0.65
BMI, mean (SD)	28.8 (6.6)		28.4 (6.7)		0.58
30 kg/m ² (n, %)	102	34.3	83	33.6	0.95
High school / GED	46	15.5	47	19.0	0.30
Full-time employment	140	47.1	115	46.6	0.55
1 alcoholic drink / day	46	15.5	44	17.8	0.28
Current smoker	52	17.5	50	20.2	0.71
Screening HF Variables					
Average HF / Day, mean (SD)	8.9 (5.9)		8.5 (5.0)		0.38

Variable	Combined Active N=297		Combined Placebo N=247		p-value ¹
	n	%	N	%	
HF Frequency SD, mean (SD)	2.0 (1.3)		1.9 (1.1)		0.50
Average HF severity, mean (SD)	2.1 (0.5)		2.1 (0.5)		0.89
Average HF bother, mean (SD)	3.1 (0.5)		3.1 (0.5)		0.96

¹ p-values from t-tests for continuous and chi-square tests for categorical characteristics

² Menses indeterminate, met hormone or other criteria

Table 2

First Clinically Significant Improvement by Week¹

Week	Active (n=200)			Placebo (n=114)		
	n	Percent	Cumulative Percent	n	Percent	Cumulative Percent
1	50	25.0	25.0	23	20.2	20.2
2	53	26.5	51.5	17	14.9	35.1
3	20	10.0	61.5	21	18.4	53.5
4	23	11.5	73.0	16	14.0	67.5
5	22	11.0	84.0	17	14.9	82.5
6	9	4.5	88.5	4	3.5	86.0
7	9	4.5	93.0	7	6.1	92.1
8	14	7.0	100.0	9	7.9	100.0

¹ Participants who had clinically significant improvement (>=50% reduction in hot flash frequency) at any time in the 8-week study (n=314).

Table 3
Predictors of 50% Reduction in Hot Flash Frequency after 8 Weeks of Placebo Intervention

Univariate Predictor (unit increase)	Univariate ¹		First Multivariable Model ^{1,4}		Final Model ^{1,4}	
	OR	p-value	OR	p-value	OR	p-value
Ethnicity		<0.001		0.002		<0.001
White	1.00		1.00		1.00	
African American	7.52 (3.31, 17.08)		5.09 (2.07, 12.48)		5.61 (2.41, 13.07)	
Other / unknown	0.89 (0.23, 3.47)		0.89 (0.21, 3.86)		0.88 (0.22, 3.56)	
Current smoking		<0.001		0.086		0.038
No	1.00		1.00		1.00	
Yes	3.50 (1.70, 7.23)		2.05 (0.90, 4.66)		2.30 (1.05, 5.06)	
Screening HF severity ²		<0.001		0.037		0.047
1-2	1.79 (1.27, 2.52)		1.50 (1.02, 2.20)		1.45 (1.00, 2.10)	
3-4	0.62 (0.42, 0.93)	0.021	0.75 (0.47, 1.20)	0.23		
BMI, kg/m ²		0.11		0.99		
<30 (ref)	1.00		1.00		1.00	
30-34.9	1.63 (0.90, 2.95)		1.00 (0.50, 2.00)			
PHQ-9 Depression ≥5		0.11		0.096		
No	1.68 (0.90, 3.17)		1.86 (0.90, 3.86)			
Yes	1.00	0.16	1.00	0.92		
Full-time employment		0.16		0.92		
No	1.00		1.00			
Yes	0.67 (0.38, 1.18)		0.97 (0.50, 1.89)			
High school / GED		0.18		0.99		
No	1.00		1.00			
Yes	1.64 (0.80, 3.34)		1.00 (0.42, 2.38)			
Menopause Status		0.36				
Transition (ref)						
Postmenopause	0.62 (0.29, 1.30)					

Univariate Predictor (unit increase)	Univariate ¹		First Multivariable Model ^{1,4}		Final Model ^{1,4}	
	OR	p-value	OR	p-value		
Indeterminate	0.93 (0.31, 2.75)					
SCL Anxiety > 0.3	1.23 (0.66, 2.29)	0.52				
AE ² reported at Week 4		0.55				
No	1.00					
Yes	0.84 (0.47, 1.50)					
Screening HF frequency standard deviation ²	1.06 (0.83, 1.34)	0.66				
Years of hot flashes ²	1.04 (0.80, 1.34)	0.79				
Alcohol use						
< 1 drink / day	1.00					
1 drink / day	0.93 (0.45, 1.94)	0.85				

¹ Combined placebo group, N=247. Models adjusted for MS FLASH Protocol and clinic site.

² Odds ratio indicates the likelihood of improvement based on a unit increase in the continuous variable as follows: HF severity, 0.5 points; age, 5 years; HF frequency standard deviation, 1; years of HF, 5 years.

³ Adverse events included fatigue, nausea, insomnia, headache, dry mouth, vivid dreams, appetite changes, drowsiness, and increased sweating.

⁴ The first model included all variables entered with P<=0.20 in univariate analysis. The final model included all variables at P<0.05 in multivariable analysis.