Menopause: The Journal of The North American Menopause Society
Vol. 21, No. 1, pp. 45-58
DOI: 10.1097/gme.0b013e31829337a4

Methods for the design of vasomotor symptom trials: the Menopausal Strategies: Finding Lasting Answers to Symptoms and Health network

Katherine M. Newton, PhD,[1] Janet S. Carpenter, PhD, RN, FAAN,[2] Katherine A. Guthrie, PhD,[3] Garnet L. Anderson, PhD,[3] Bette Caan, DrPH,[4] Lee S. Cohen, MD,[5] Kristine E. Ensrud, MD, MPH,[6] Ellen W. Freeman, PhD,[7] Hadine Joffe, MD,[5] Barbara Sternfeld, PhD,[4] Susan D. Reed, MD, MPH,[8] Sheryl Sherman, PhD,[9] Mary D. Sammel, ScD,[10] Kurt Kroenke, PhD,[11] Joseph C. Larson, MS,[3] and Andrea Z. LaCroix, PhD[3]

Abstract

Objective: This report describes the Menopausal Strategies: Finding Lasting Answers to Symptoms and Health network and methodological issues addressed in designing and implementing vasomotor symptom trials.

Methods: Established in response to a National Institutes of Health request for applications, the network was charged with conducting rapid throughput randomized trials of novel and understudied available interventions postulated to alleviate vasomotor and other menopausal symptoms. Included are descriptions of and rationale for criteria used for interventions and study selection, common eligibility and exclusion criteria, common primary and secondary outcome measures, consideration of placebo response, establishment of a biorepository, trial duration, screening and recruitment, statistical methods, and quality control. All trial designs are presented, including the following: (1) a randomized, double-blind, placebo-controlled clinical trial designed to evaluate the effectiveness of the selective serotonin reuptake inhibitor escitalopram in reducing vasomotor symptom frequency and severity; (2) a two-by-three factorial design trial to test three different interventions (yoga, exercise, and ω-3 supplementation) for the improvement of vasomotor symptom frequency and bother; and (3) a three-arm comparative efficacy trial of the serotonin-norepinephrine reuptake inhibitor venlafaxine and low-dose oral estradiol versus placebo for reducing vasomotor symptom frequency. The network's structure and governance are also discussed.

Conclusions: The methods used in and the lessons learned from the Menopausal Strategies: Finding Lasting Answers to Symptoms and Health trials are shared to encourage and support the conduct of similar trials and to encourage collaborations with other researchers.

Key Words: Menopause – Yoga – Exercise – Hot flashes – Clinical trial network – Vasomotor symptoms – Study design – Methodology.

Received December 3, 2012; revised and accepted March 4, 2013.

From the [1]Group Health Research Institute, Seattle, WA; [2]School of Nursing, Indiana University, Indianapolis, IN; [3]Data Coordinating Center, Fred Hutchinson Cancer Research Center, Seattle, WA; [4]Division of Research, Kaiser Permanente of Northern California, Oakland, CA; [5]Massachusetts General Hospital/Harvard Medical School, Boston, MA; [6]VA Medical Center/University of Minnesota, Minneapolis, MN; [7]Departments of Obstetrics/Gynecology and Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA; [8]Department of Obstetrics/Gynecology and Department of Epidemiology, School of Public Health and Community Medicine, University of Washington, Seattle, WA; [9]National Institute on Aging, US National Institutes of Health, Bethesda, MD; [10]Center for Clinical Epidemiology and Statistics, University of Pennsylvania School of Medicine, Philadelphia, PA; and [11]VA Center of Excellence for Implementing Evidence-Based Practice and Indiana University School of Medicine, Indianapolis, IN.

Clinical Trial Registration: NCT00894543, NCT 01178892, and NCT 01418209 (http://www.clinicaltrials.gov).

Funding/support: This study was funded by the National Institutes of Health as a cooperative agreement issued by the National Institute on Aging, the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Center for Complementary and Alternative Medicine, and the Office of Research on Women's Health (U01 AG032656, U01AG032659, U01AG032669, U01AG032682, U01AG032699, and U01AG032700).

Financial disclosure/conflicts of interest: K.M.N. has received research support from Integrated Diagnostics. L.S.C. has received research support from Astra-Zeneca Pharmaceuticals, Bristol-Myers Squibb, Cephalon Inc., GlaxoSmithKline, National Institute on Aging, National Institute of Mental Health, Ortho-McNeil Janssen, and Sunovion Pharmaceuticals Inc., and has served in the advisory board of or as consultant to PamLab LLC and Noven Pharmaceutical. K.E.E. has served in the Data Monitoring Committee of Merck Sharpe and Dohme. E.W.F. has received research support from Forest Laboratories Inc., Xanodyne Pharmaceuticals, Bionovo, and Wyeth. H.J. has received research support from Astra-Zeneca Pharmaceuticals, Bayer HealthCare, Cephalon/Teva, Eli Lilly, Forest Laboratories Inc., GlaxoSmithKline, Sanofi-Events (product support only), Sepracor Inc., and Wyeth-Ayerst Pharmaceuticals; has received speaking honoraria from Eli Lilly and GlaxoSmithKline; and has served in the advisory board of or as consultant to Abbott Laboratories, Eli Lilly, Forest Laboratories Inc., JDS-Noven Pharmaceuticals, Sanofi-Events, Sepracor Inc., Sunovion, and Wyeth-Ayerst Pharmaceuticals. M.D.S. has served as consultant to Swiss Precision Diagnostics GmbH and as statistical reviewer for the *American Journal of Obstetrics and Gynecology*. A.Z.L. has served in the Scientific Advisory Board of the Global Longitudinal study of Osteoporosis in Women (GLOW); has received an unrestricted research grant from the University of Massachusetts Center for Outcomes Research, Proctor and Gamble, and Sanofi; and was an attendee at a one-time consultants' meeting for Pfizer to evaluate evidence on hormone therapy and breast density. Other authors report no conflict of interest.

Address correspondence to: Katherine M. Newton, PhD, Group Health Research Institute, Seattle, WA. E-mail: newton.k@ghc.org

The long-term objective of the National Institute on Aging's (NIA's) request for applications (RFA)-AG-08-004, "New Interventions for Menopausal Symptoms" (U01), was to accelerate progress in identifying effective remedies for vasomotor symptoms (VMS) in women experiencing the menopausal transition. The RFA was sponsored by the NIA, in collaboration with the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Center for Complementary and Alternative Medicine, and the Office of Research on Women's Health, to create a network of scientists who are highly knowledgeable about the menopausal transition and experienced in the conduct of women's health trials. The purpose of this article is to describe the composition of the Menopausal Strategies: Finding Lasting Answers to Symptoms and Health (MsFLASH) network and the methodological issues addressed in the design and implementation of VMS trials in this multicenter national menopause network.

The MsFLASH network is funded through a cooperative agreement and is composed of the Data Coordinating Center (DCC) and five study sites, with representation from the four funding agencies (Fig. 1). Network investigators have conducted three randomized controlled trials examining six different interventions for relief of menopausal symptoms. The primary outcomes for all three trials included VMS frequency and severity/bother. The first trial was a standard placebo-controlled study to determine the efficacy and tolerability of escitalopram (10-20 mg/d), a selective serotonin reuptake inhibitor, compared with placebo pills.[1] The second trial used a two-by-three factorial design to separately compare the effects of yoga and exercise with a common wait-list control group and to simultaneously compare ω-3 fatty acid capsules to placebo capsules. The third trial compared the efficacies of low-dose oral estradiol and the serotonin-norepinephrine reuptake inhibitor venlafaxine XR to placebo.

METHODOLOGICAL ISSUES IN THE DESIGN OF MsFLASH TRIALS

The National Institutes of Health charged the network to conduct rapid throughput randomized trials of novel and understudied currently available interventions postulated to alleviate menopausal symptoms. Reduction in VMS (hot flashes plus night sweats) was the overall goal of the network trials. Network investigators held extensive discussions on what common eligibility criteria should be used for network trials. Consensus emerged on the following major points:

1. Having common eligibility criteria facilitates the comparison of intervention results across trials.
2. Generalizability of trial results improves when criteria are more inclusive than exclusive.
3. There is value in allowing inclusion and exclusion criteria to vary when such variation strengthens the science or safety of a specific protocol.
4. Women should experience a sufficient number of bothersome VMS to warrant treatment.
5. Most women who experience menopause-related VMS see their symptoms diminish or disappear within 5 years of becoming postmenopausal.
6. Women with premature ovarian failure, early oophorectomy, or ovarian ablation are important subpopulations of interest to network investigators.
7. Women who experience VMS for prolonged durations (>95 y) after cessation of ovarian function are a

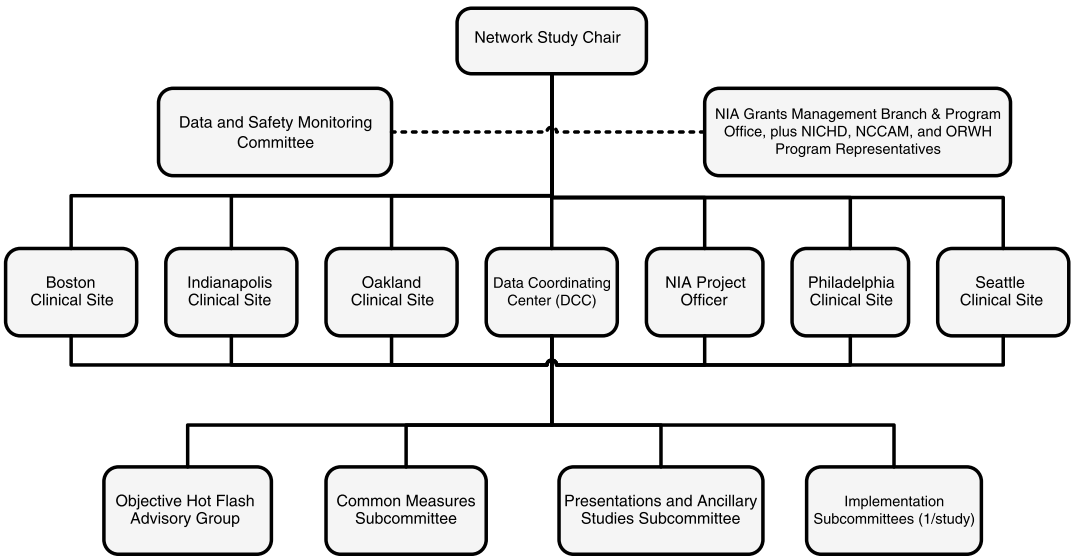

FIG. 1. Menopausal Strategies: Finding Lasting Answers to Symptoms and Health leadership structure. NIA, National Institute on Aging; NICHD, Eunice Kennedy Shriver National Institute of Child Health and Human Development; NCCAM, National Center for Complementary and Alternative Medicine; ORWH, Office of Research on Women's Health.

small but important subgroup of women, and network investigators are committed to advancing therapies in this group.

8. Individual trials may be designed to focus especially on understudied special populations of interest, such as African-American women or women with prolonged VMS.

Directed by the guidelines above, we sought to establish common study designs, eligibility/exclusion criteria, and study measures that could be used across the trials (Table 1). The goal was to include in the network trials as many women as possible who were experiencing frequent and bothersome VMS around the time of menopause, thus preserving the ability of the interventions to show treatment effects that could be generalized to the population.

Criteria for interventions and study selection

We established guidelines and discussion points on which to base our trial selection process. Network investigators wanted to study a variety of pharmaceutical and behavioral interventions, particularly interventions that were frequently recommended without strong evidence (eg, yoga, exercise, paced respiration). We wanted to maximize our ability to compare interventions both within and between trials to help women and providers in decision-making. Factors considered in choosing study medications included cost, timeline to generic status, availability of drugs and matching placebos, adverse effects, dosing issues, and known information about effectiveness. There was consideration of women's willingness to receive each planned treatment intervention given the potential adverse effect profiles. Other considerations included adherence (ability to maintain an 8-wk intervention to a 12-wk intervention), tolerability, and safety. Masking of interventions was incorporated to the extent possible, given the established placebo response found in most studies of interventions for menopausal symptoms.

Common eligibility and exclusion criteria

We undertook a detailed comparison of eligibility criteria used in previous trials by network investigators and others, and a consensus emerged for eligibility criteria in the network, with the understanding that specific protocols might have additional criteria to optimize the protocol for the specific intervention, to exclude women for safety reasons, or both.

Age

In the Study of Women's Health Across the Nation, the median age at natural menopause was 51.4 years,[2] and menopausal symptoms peaked around the time of menopause.[3] In an earlier study, McKinlay et al[4] found that 96.4% of women were postmenopausal by the age of 54 years. We therefore recruited women aged 40 to 62 years to capture women who are likely to be in the menopausal transition or within 5 years beyond menopause, the time when symptoms are most prevalent. This was a pragmatic decision to maximize mailings to women who

TABLE 1. *MsFLASH eligibility criteria for all MsFLASH trials (exceptions noted in parentheses)*

Menopausal/hot flash
1. At least 28 VMS/wk, as recorded on daily VMS diaries for each of 3 wk (MsFLASH 01; changed to 14 VMS/wk in MsFLASH 02; maintained in MsFLASH 03). On the third week of VMS diaries, the mean hot flash frequency cannot drop by more than 50% from the mean level reported for the first 2 wk.
2. At least four reports of days or nights with moderate to severe VMS per week and/or bother for each of the 3 wk.
3. Not longer than 5 y postmenopausal in the late menopausal transition.
4. Aged 40-62 y at screening.
5. No reported use of systemic hormones (contraceptives or postmenopausal hormones) in the past 2 mo (levonorgestrel intrauterine system and vaginal estrogen local therapy <4 times per week allowed).
6. Agreeing not to use hormones outside the trial for the duration of the trial.
7. No reported use of any nonhormonal hot flash therapy in the past month.
8. Hysterectomy with serum FSH level >20 mIU/mL and estradiol level <50 pg/mL at baseline (in at least one of two blood draws 2 wk apart).
9. Using the Mirena intrauterine device or had endometrial ablation but still has one ovary or both ovaries, with FSH level >20 mIU/mL and estradiol level ≤50 pg/mL (in at least one of two blood draws 2 wk apart).

Medical history
1. Generally good health as determined by medical history.
2. No severe or unstable medical illness.
3. No uncontrolled hypertension (>160/100 mm Hg).
4. Resting heart rate ≤110 beats/min.
5. Not pregnant, not intending to be pregnant, or not breast-feeding.
6. No abnormal mammogram results in the past 2 y (MsFLASH 03).
7. Body mass index ≤37 kg/m^2 based on measured height and weight.
8. No major severe depressive episode in the past 3 mo.
9. No suicide attempt in the last 3 y (MsFLASH 01 and 03).
10. No diagnosis of psychosis/psychotic disorder.

Drug/medication use (not related to menopause therapy; see above)
1. No drug/alcohol abuse in the past year.
2. Not using psychotropic medications (selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, monoamine oxidase inhibitors) or other antidepressants and anxiolytics within the past 30 d (MsFLASH 01 and 03).

Study logistics
1. Provided a written informed consent form.
2. Not participating in any another menopause trial.
3. No hypersensitivity or contraindication to study medications/treatment (varies by study).
4. Willing and able to complete study procedures.

MsFLASH, Menopausal Strategies: Finding Lasting Answers to Symptoms and Health; VMS, vasomotor symptoms; FSH, follicle-stimulating hormone.

were most likely to meet our entry criteria. Over time, we found that most responses to recruitment mailings were from women aged 47 to 62 years, and mailings were targeted to women in this age range, although women aged 40 to 46 years who contacted us about the trial and met other eligibility criteria remained eligible.

Menopause status

Studies of menopausal symptom treatments (including US Food and Drug Administration [FDA]–monitored drug trials) typically include only postmenopausal women because of the potential impact of variations in circulating hormones.[5] We included women in the late menopausal transition and those in the first 5 years postmenopause because women frequently seek therapy for symptoms during the menopausal transition. Guided by the Stages of Reproductive Aging Workshop criteria in defining the stages of menopause,[6] we classified women into "late menopausal transition" when they skipped two consecutive menstrual cycles not associated with pregnancy or breast-feeding and had an amenorrhea interval of 60 days or more within the last year. Women were classified as "postmenopausal" when their last menstrual period occurred 12 months or more previously. Network investigators considered expanding these criteria to include women in the early menopausal transition who met the VMS criteria and might benefit from these interventions. However, they were ultimately excluded because, based on the clinical judgment of study clinicians, women in the early menopausal transition are more likely to experience intermittent VMS that would obscure the evaluation of treatment effectiveness during short-term trials.

Network investigators discussed at length whether to include women who had undergone oophorectomy and hysterectomy and decided that the same criteria should be applied to all women with menopausal VMS regardless of menopause type. Thus, women who had undergone bilateral oophorectomy were eligible if they met the other inclusion criteria. Women who had undergone hysterectomy and had at least one ovary, had follicle-stimulating hormone levels higher than 20 mIU/mL, and had estradiol levels of 50 pg/mL or less were included as they were most likely to be postmenopausal (acknowledging that some of these women might have been classified into the late menopausal transition had they had a uterus).

VMS criteria, assessment, and primary outcomes

Our primary interest was VMS frequency and severity/bother. Although we considered the use of electronic diaries, the cost of electronic devices and of creating online resources for these relatively short studies was prohibitive. We considered continuous real-time VMS diaries but were concerned about participant burden. There was also strong evidence that a daily diary—on which women record the number, bother, and severity of daytime hot flashes before going to sleep and record the same values for night sweats upon arising—was responsive to change.[7] We had concerns about the burden of twice-daily recording, including whether women would complete the diaries or would complete them retrospectively. However, given the evidence on prior successful use with minimal burden and our desire to compare our results with the results of other studies, we chose to use this method for all MsFLASH trials. We also considered how frequently the diaries should be completed, and there were differences of opinion among the study investigators. Some believed that it was easiest if women kept the diaries throughout the study, whereas others thought that this was too great a burden and suggested that the diaries be accomplished only during selected weeks. Ultimately, the diaries were accomplished continuously in MsFLASH 01 and 03, and at baseline, 6 weeks, and 12 weeks in MsFLASH 02.

To enroll women with stable and persistent VMS symptoms, we asked women to complete the VMS diaries for 3 weeks before randomization (Fig. 1). In the first trial (MsFLASH 01) that tested a pharmacologic agent, we required the following: 28 or more VMS/week recorded on daily diaries for 3 weeks; VMS rated as bothersome or severe on 4 days or more per week; and VMS frequency on week 3 not decreasing by more than 50% from the average weekly levels on weeks 1 and 2.

We carefully considered the minimal number of VMS for eligibility because women needed to have sufficient VMS to see a change in symptoms with therapy. For drug trials, the FDA recommends that women have 7 to 8 moderate to severe VMS/day, or 50 to 60 moderate to severe VMS/week at baseline.[5] The FDA defines moderate VMS as a sensation of heat with sweating but with the woman able to continue activity, whereas severe VMS requires cessation of activity. We chose lower thresholds for both frequency and severity/bother for the study to be more inclusive and generalizable than typical FDA-monitored trials. Our rationale for requiring a minimal number of VMS perceived as bothersome or severe was that women seek treatment not only because of VMS frequency but also because they are bothered by VMS and/or perceive them as severe. We collected both VMS severity and VMS bother, despite their high correlation, because of the subjective interpretation of bother (eg, a woman may rate her VMS as severe but may not be bothered, or be highly bothered, by mild VMS). The 3-week screening period was designed to eliminate women with highly variable VMS frequency and thus minimize placebo response in the trials.

In the second trial, we lowered the VMS criteria for two reasons. First, we were testing several nonpharmacologic interventions that might be attractive to a broader spectrum of women with VMS. Second, recruitment was challenging owing to several study-specific entry criteria. When analyses of screening diaries showed that insufficient VMS frequency was the most common reason for ineligibility for MsFLASH 01, we evaluated the potential impact of using a lower threshold. We examined both the range of baseline VMS symptoms in MsFLASH 01 and the range of baseline VMS in a prior study of herbal therapies for menopausal symptoms that required only 2 VMS/day.[8] We concluded that a lower threshold would potentially increase recruitment with minimal effects on our results. We amended the criteria to the following: 14 or more VMS/week recorded on daily hot flash diaries for 3 weeks; VMS rated as bothersome or severe four

or more times per week; and VMS frequency on week 3 not decreasing by more than 50% from the average weekly levels on weeks 1 and 2. At the end of the study, we found that for the 67 women randomized before this (requiring 4 VMS/d), the mean (SD) VMS/24-hour day was 8.5 (4.0). For the 288 women randomized after the change, the mean (SD) VMS/24-hour day was 7.4 (3.8). Thus, changing our criteria increased enrollment and generalizability with minimal difference in baseline VMS. The criteria were retained for the third trial.

Menopausal and hormonal therapies

We excluded women if they had used over-the-counter or herbal therapies specifically for VMS in the past 30 days or if they had used hormone therapy or hormonal contraceptives in the past 2 months. We also excluded women who used selective estrogen receptor modulators or aromatase inhibitors in the past 2 months. These exclusions were incorporated to avoid potential carryover or withdrawal effects that might obscure intervention results.

Common secondary outcome measures

In addition to VMS, secondary domains of interest were sleep, depression, anxiety, pain, quality of life, sexual function, sexual distress, and perceived stress (Table 2). Measures were chosen based on several factors. At the outset, we established as guiding principle the use of well-validated

TABLE 2. *MsFLASH common measures and outcomes*

Outcome	Description	Measure selected	Measurement details	Number of items	When collected
Vasomotor symptoms	Subjective VMS frequency	Diary: twice-daily (day/night) estimate of the number of VMS	Self-reported paper diary	2 per day	Daily
	Subjective VMS severity	Diary: twice-daily rating (mild, moderate, severe)	Self-reported paper diary	2 per day	Daily
	Subjective VMS bother[9]	Diary: twice-daily rating using Study of Women's Health Across the Nation response categories (not at all, a little, moderately, a lot)[9]	Self-reported paper diary	2 per day	Daily
	Subjective VMS interference	Hot Flash Related Daily Interference Scale[10]	Self-reported	10	Baseline, follow-up
Sleep	Sleep quality and disturbance	Pittsburgh Sleep Quality Index[11]	Self-reported	18	Baseline, follow-up
	Insomnia symptoms	Insomnia Severity Index[11]	Self-reported	7	Baseline, follow-up
	Sleep, wake, and nap times; Actiwatch use	Diary: twice daily (day/night); device	Self-reported	≥2 per day	Baseline, follow-up, daily
	Objective sleep quality	Actiwatch real-time recording for 3 d	Actigraphic monitoring	0	Baseline, follow-up
Mood state/depression/ anxiety/stress	Depressive symptoms	PHQ (PHQ-9 or PHQ-8)[12-14]	Self-reported	8-9	Baseline, follow-up
	Anxiety symptoms	Generalized Anxiety Disorder-7[12]	Self-reported	7	Baseline
		Hopkins Symptom Checklist[15-17]	Self-reported	10	Baseline, follow-up
	Perceived stress	Perceived Stress Scale[18]	Self-reported	10	Baseline, follow-up
Physical pain	Pain severity and interference	PEG (from the Brief Pain Inventory)[19,20,21]	Self-reported	3	Baseline, follow-up
Quality of life: menopause-specific	Presence and bother associated with menopausal symptoms	Menopause-Specific Quality of Life questionnaire[22]	Self-reported	29	Baseline, follow-up
Quality of life: overall	Importance and satisfaction with life	Quality of Life Enjoyment and Satisfaction Questionnaire[23]	Self-reported	15	Baseline, follow-up
Sexual function/vaginal dryness	Key dimensions of female sexual function	Female Sexual Function Index[24]	Self-reported	18	Baseline, follow-up
Sexual distress	Distinguish between sexual dysfunction and no sexual dysfunction; sexually related personal distress in women with hypoactive sexual desire disorder	Female Sexual Distress Scale—Revised[25]	Self-reported	1	Baseline, follow-up

MsFLASH, Menopausal Strategies: Finding Lasting Answers to Symptoms and Health; VMS, vasomotor symptoms; PHQ, Patient Health Questionnaire; PEG, a three-item questionnaire adapted from the Brief Pain Inventory that assesses average pain intensity (P), interference with enjoyment of life (E), and interference with general activity (G).

and psychometrically sound self-report questionnaires. We sought measures of mood that were not overly sensitive to somatic symptoms. We favored shorter scales to longer scales to lessen participant burden. We included selected global measures (eg, Menopause-Specific Quality of Life questionnaire) because they are widely used in other menopause studies.

Sleep measures

Sleep disturbance and related complaints are a primary reason why women seek treatment of VMS.[26] To measure insomnia symptoms, we used the Insomnia Severity Index,[27-29] a valid and reliable self-administered instrument that measures the perception of current (past 2 wk) insomnia symptoms. The index has seven items assessing difficulty falling asleep, difficulty staying asleep, problems with early awakening, satisfaction with current sleep pattern, interference of sleep problem with daily functioning, noticeability of impairment attributed to the sleep problem, and degree of distress caused by the sleep problem. Each item is rated on a scale from 0 to 4 (total score, 0-28), with higher scores suggesting more severe insomnia symptoms. The absence of insomnia is indicated by scores 0 to 7; subthreshold or mild insomnia is indicated by scores 8 to 14; clinical insomnia of moderate severity is indicated by scores 15 to 21; and severe clinical insomnia is indicated by scores 22 to 28. Trials of pharmacologic and behavioral interventions in women with insomnia have suggested the Insomnia Severity Index to be sensitive in measuring treatment response.[30,31]

To assess self-reported sleep quality, we used the Pittsburgh Sleep Quality Index (PSQI), a validated measure of subjective sleep quality and sleep disturbances, during a 1-month period.[11] The PSQI assesses subjective sleep quality, latency, duration, and efficiency; sleep disturbances; use of sleeping medication; and daytime dysfunction.[11,32] Global PSQI scores range from 0 to 21, with higher scores indicating poorer sleep quality. Cutoffs of 5[32] and 8[33] have been reported to indicate poor sleep quality. The PSQI has been shown to be sensitive in measuring responses to cognitive-behavioral therapy in randomized trials conducted in women with insomnia.[34]

To objectively measure sleep-wake patterns, participants wore an Actiwatch 2 (Philips Respironics, Bend, OR) for 7 days at baseline and follow-up (8 or 12 wk). The Actiwatch is a small device that is similar in appearance to a wristwatch. An accelerometer within the Actiwatch measures movement several times per second and digitally stores the information every minute. Actigraphy has been shown to provide an objective and reliable estimate of sleep-wake patterns.[35,36] Participants were instructed to wear the Actiwatch continuously for 7 nights/8 days 1 week before baseline and closeout visits, removing it only for bathing or in situations where it might get submerged in water. They were also asked to keep a sleep log in which they recorded their normal sleep-wake patterns, as well as their time to bed, time of final arising, and times that the actigraph was removed. Sleep logs were used to aid in the editing of actigraph data.

In MsFLASH 02, women were also asked to wear an accelerometer for 7 days at baseline and follow-up to monitor free-living physical activity. Tracking and maintenance of two devices were challenging for both the women and the study staff.

Mood

Another primary reason that women initiate menopausal treatment is mood disturbance. Menopausal transition is a time of increased risk for a new onset or a reoccurrence of clinical depression and depressive symptoms.[37,38] To assess depression, we used either the eight-item version (in behavioral intervention studies) or the nine-item version (in antidepressant studies where the ninth item assessing thoughts on death or self-harm was deemed important) of the depression module of the Patient Health Questionnaire.[12] Both versions of the Patient Health Questionnaire depression scale can be scored either continuously (as a depression severity score) or categorically (to indicate a probable DSM-IV depressive diagnosis). We evaluated anxiety using the Generalized Anxiety Disorder-7, which can likewise be scored either continuously (as an anxiety severity score) or categorically (with cutpoints that indicate a probable DSM anxiety disorder).[12] Because the Generalized Anxiety Disorder-7 has not been validated for change in response to treatment, we also included at baseline and follow-up the anxiety factor of the Hopkins Symptom Checklist as a validated and sensitive measure of change in response to treatment.[15-17] We used the Perceived Stress Scale, a widely used validated self-report of perceived stress, to assess stress as an independent construct associated with VMS.[18]

Pain

MsFLASH investigators were interested in exploring the relationship between menopause and pain, and chose PEG, a three-item questionnaire adapted from the Brief Pain Inventory that assesses average pain intensity (P), interference with enjoyment of life (E), and interference with general activity (G).[19,39] The PEG scale has shown high internal consistency, construct validity, and responsiveness to changes in outpatient populations.[19] In a randomized trial of adults with musculoskeletal pain, responsiveness to change in pain has been found to be equal or superior to several longer pain scales.[39]

Quality of life

The Menopause-Specific Quality of Life questionnaire[22] was chosen because it is a global health-related quality-of-life scale designed specifically for use in menopause and has been used frequently in menopause research.

VMS interference

Perceived hot flash interference was evaluated using the Hot Flash Related Daily Interference Scale.[10] This 10-item scale measures a woman's perceptions of the degree to which VMS interfere with nine daily life activities; the 10th item measures interference with overall quality of life. This scale was modeled after items on the Brief Pain Inventory[40] and Brief Fatigue

Inventory,[41] which assess the degree to which pain or fatigue interferes with similar activities. Participants rate the degree to which VMS have interfered with each item during the previous week using a scale from 0 (*does not interfere*) to 10 (*completely interferes*). Recent structural equation modeling suggests that this is a unidimensional scale best represented by an overall mean score (sum of items / 10). The Hot Flash Related Daily Interference Scale has been shown to be sensitive to the effects of pharmacologic interventions[42] and behavioral interventions.[43]

Sexual function and sexual distress

Sexual function and sexual distress are common complaints of midlife women but have not received widespread attention in research or clinical practice, and we viewed our trials as an opportunity to further explore these issues. We therefore included in all trials the Female Sexual Function Index[24] and a single item from the Female Sexual Distress Scale.[25] The full Female Sexual Distress Scale was implemented in MsFLASH 02.

Methods relevant to placebo response

Almost all studies of interventions for menopausal symptoms have shown VMS decreases in both intervention and control groups. In placebo groups, VMS frequency has been shown to decrease by 20% to 60% from baseline.[8,44-47] This variable decrease in VMS has been attributed to regression to the mean, natural resolution of symptoms, placebo response, and fatigue with symptom recording for long periods. We sought to minimize this phenomenon in our studies through screening procedures. Participants recorded VMS daily for 2 weeks; those who met study criteria recorded VMS for a third week. Those with a greater-than-50% decrease in VMS frequency on week 3 compared with weeks 1 and 2 were ineligible.

All drug interventions in MsFLASH trials had double-blind designs using matched placebo capsules. We also sought believable and appropriate control groups for behavioral studies, and we discussed at length whether we should create attention-control groups for our behavioral interventions (exercise, yoga). The comparison of an intervention group to an attention-control group tests the hypothesis that the effect of the intervention is attributable to some aspects of the intervention other than attention or expectancy.[48] As Gross[49] pointed out, the assumption is that attention and the expectation that "good things will happen" are not active ingredients of the intervention and can somehow be separated. Arguably, for MsFLASH behavioral interventions, it was difficult to imagine why attention would not be considered an inherent and important component of the intervention. Furthermore, it is critical to the appropriate use of attention-control groups that they provide a believable intervention.[48] We were unable to create an attention-control behavioral intervention that we trusted would believably hold women's interest for 12 weeks. Thus, rather than using attention-control groups, we designed a study where all women would be engaged in a believable intervention. Our second trial involved two behavioral interventions: exercise and yoga. The factorial design of the trial (see below) simultaneously randomized all women to either active or placebo ω-3 fatty acid pills. Thus, all women had some expectation of effect.

Collection of blood, urine, saliva, and vaginal samples, and biorepository

Trial participants were asked to contribute a variety of biologic samples, including blood, urine, saliva, and vaginal swabs, for future studies. These are being maintained in the MsFLASH Biobank for ancillary studies. An a priori decision was to collect fasting (overnight) blood specimens at baseline and follow-up for every trial. Blood samples were processed on site; aliquoted into cryovials for serum, plasma, and buffy coat; frozen; and shipped frozen to the central biospecimen repository at the Fred Hutchinson Cancer Research Center's Specimen Processing Laboratory for later analyses. Approximately 8 mL of blood was collected in a (10 mL) red-top tube; after processing, 0.5 mL of serum was aliquoted into each of nine 1-mL cryovials. Approximately 8 mL of blood was collected in a 10-mL lavender-top EDTA tube; after processing, 0.5 mL of EDTA plasma was aliquoted into each of nine 1-mL cryovials. The buffy was removed and aliquoted into two 0.5-mL cryovials. The DCC provided barcoded blood ID labels suitable for −70°C freezer storage to ensure that samples, forms, databases, and cryovials were linked via a unique ID and a two-digit cryovial number to the participant's study ID and type of visit.

In MsFLASH 01, an overnight urine sample was collected at baseline and study completion (week 8). Participants were instructed to keep their specimen in the refrigerator during the collection period and to return it in a cooler that was provided along with a gel ice pack. Approximately 9 mL of urine was centrifuged, and 0.5 mL was aliquoted into each of six 1-mL cryovials for storage in the MsFLASH repository at the Fred Hutchinson Cancer Research Center.

In MsFLASH 02, saliva samples (for cortisol) were collected. Four samples were obtained on each of two consecutive days at baseline and study completion (week 12; total, 16 samples). Salivettes were placed in a Ziploc bag in the participant's freezer until they were transported back to the clinical site. At the clinical site, the samples were processed, placed in a −70°C freezer, and batched for transport to the DCC. The DCC provided a form and sample ID labels that linked the salivary samples to the participant and the visit.

In MsFLASH 03, women consented separately for an ancillary study of vaginal microbiome. Vaginal swabs were collected at baseline. Women were asked to collect swabs at home (using techniques perfected by Srinivasan et al[50] with excellent compliance, safety, and specimen quality) on days 1 to 14 and then weekly for the remaining 6 weeks of the trial. Women were also asked to complete a vaginal symptoms questionnaire at baseline and study completion. Participants mailed vaginal swabs to the study laboratory weekly and returned diaries to the research clinic at study completion (week 8). The swabs were stored for analysis after the end of the trial.

Objective hot flash monitoring

The original intent of MsFLASH investigators was that all trials would use both subjective and objective VMS monitoring. Objective VMS measurement has been recommended as an adjunct to the subjective measurement of frequency, severity, bother, and/or duration. The potential advantage of objective monitoring is that results should be unbiased by placebo effects,[42] sleep-wake cycles,[51,52] and reporting biases.[51,53] We evaluated three potential monitors for use in MsFLASH trials[54]: Freedman monitor, Bahr monitor, and Biolog monitor. Briefly, none of the tested monitors were found to be suitable for ambulatory clinical trials. In our tests, the Freedman monitor did not adequately distinguish VMS events from ambient humidity. The Bahr monitor recorded data inconsistently, with large sections of poor quality or having missing data. The Biolog monitor performed more consistently, but there were ongoing problems with electrode availability. Therefore, the decision not to use objective VMS monitoring in MsFLASH studies was made.[54] Although objective monitoring devices can be purchased and despite the National Center for Complementary and Alternative Medicine's efforts to move this technology forward, a reliable and affordable VMS monitoring device for ambulatory studies remains elusive. Such a device would be a meaningful addition to VMS studies.

Aspects of clinical trial designs

MsFLASH 01 was a randomized, double-blind, placebo-controlled clinical trial designed to evaluate the effectiveness of a selective serotonin reuptake inhibitor (escitalopram) in reducing VMS frequency and severity. To mimic clinical practice, we included blinded dose escalation in the design for nonresponders halfway through the 8-week trial. Because the dose escalation was dependent on response in the first 4 weeks, the trial did not provide a randomized comparison of the two doses but rather an estimate of the realistic effectiveness of the drug within a narrow range of doses. This trial also included VMS symptom assessment at 3 weeks after therapy cessation to identify return of symptoms.[1] An important secondary objective of this trial was to examine potential differences in treatment effects in African-American women compared with white women. To improve power for this interaction test, we restricted accrual to ensure that at least 95 African-American women would be randomized. A target sample size of 200 was chosen to provide at least 90% power to detect a 24% difference in hot flash frequency reduction (52% vs 28%) and a 0.52-SD unit difference in mean change in severity scores, with a two-sided 2.5% level test and allowing for up to 10% loss to follow-up (Table 3). We also collected urine specimens from all women for banking and later analysis. However, overnight urine collections were difficult to implement at all sites. Women who used public transportation or who came to the study clinic directly from work found that transporting urine was unpleasant, and they often refused. We eliminated urine collection from MsFLASH 02 and 03.

MsFLASH 02 applied a two-by-three factorial design to test three different interventions (yoga, exercise, and ω-3 supplementation) for improvement of VMS frequency and bother. We chose bother, as opposed to severity, as a primary outcome for this trial because we believed that yoga, with its meditative component, might specifically affect women's perception of bothersomeness. Randomized participants received both a behavioral intervention (yoga, exercise, or wait list for their choice of intervention) and a supplement (ω-3 or placebo) for 12 weeks. The factorial design was selected as an efficient approach to testing all three interventions and to ensure that all women would participate in at least one intervention (ω-3 supplement or placebo capsules) during the primary intervention period and, hence, would have some expectancy of benefit (see discussion of placebo effects above). We assumed that the effects of any of these approaches on VMS reduction would be relatively modest and that the interventions would probably operate through independent pathways such that any interaction between ω-3 and the two behavioral interventions was likely to be negligible. The MsFLASH 02 study design also incorporated unbalanced sample sizes to gain efficiency by using a larger behavioral control group to make two active arm comparisons with a sample size ratio of 2:2:3. As an additional benefit, an unbalanced design reduced overall study costs by randomizing fewer women to the behavioral intervention arms that were more expensive to implement.[55] Ninety participants in each of the yoga and exercise groups and 135 participants in the usual activity group were planned to provide 90% power to detect a 0.49-SD unit difference in mean change in VMS scores, based on a *t* test with a two-sided 2.5% significance level (Table 3). The marginal sample size of 158 participants with and without ω-3 fatty acid supplementation (45 in yoga, 45 in exercise, and 68 in control arm) provided 90% power for

TABLE 3. *Comparisons and sample sizes for MsFLASH clinical trials*

Trial	Intervention group	Comparison group	Effect size[a]
01	Escitalopram (n = 100)	Placebo (n = 100)	2.1
02	Exercise and ω-3/placebo (n = 112)	Usual activity and ω-3/placebo (n = 150)	2.0
	Yoga and ω-3/placebo (n = 112)	Usual activity and ω-3/placebo (n = 150)	2.0
	ω-3 and exercise/yoga/usual activity (n = 187)	Placebo and exercise/yoga/usual activity (n = 187)	1.6
03	Venlafaxine (n = 87)	Placebo (n = 130)	2.1
	Estradiol (n = 87)	Placebo (n = 130)	2.1

MsFLASH, Menopausal Strategies: Finding Lasting Answers to Symptoms and Health.

[a]Measured as between-group difference in vasomotor symptom frequency per day change from baseline to follow-up, assuming an SD of four vasomotor symptoms per day. Sample size calculations were based on *t* tests with a power of 90% and a two-sided α of 0.025.

a difference in mean reduction of 0.40 SD units, with a two-sided significance level of 0.025. The total enrollment goal was 374, allowing for a 10% inflation for loss to follow-up and for an extra 10% in the yoga and exercise groups to account for increased variability in outcome measurements attributable to a range of compliance to the behavioral interventions. This trial included an ancillary study measuring heart rate variability and salivary cortisol to evaluate potential associations with outcomes and baseline participant characteristics.

MsFLASH 02 taught us several things that we might do differently in future similar behavioral trials. Because of the factorial design and types of interventions, the facilities used were different for yoga and exercise. This created limitations attributable to women's inability to travel to one site or the other. Using the same site for both interventions and finding sites on public transit lines would have lessened these challenges. In MsFLASH 02, we also asked women to wear multiple devices—an actigraph for sleep and an accelerometer for activity measures—because each device was superior for the specific measures of interest. However, coordinating these two devices was challenging for both the women and the staff. Using a single device for both measures would have been simpler and less costly. Finally, having three different interventions (ie, weekly yoga classes with home practice, thrice-weekly in-person exercise sessions, and ω-3 capsules) created efficiencies but also meant that women who might have been eligible for one intervention (if studied alone) were excluded because they were ineligible based on another intervention (eg, seafood allergy). It was also challenging to describe the complexity of the study to potential participants.

MsFLASH 03 was a three-arm comparative efficacy trial of the serotonin-norepinephrine reuptake inhibitor venlafaxine and low-dose oral estradiol versus placebo for reducing VMS frequency. The trial was designed to test two interventions concurrently for 8 weeks against a common placebo group. We elected not to conduct the trial as a direct head-to-head comparison of these two interventions because the hypothesis of interest for the serotonin-norepinephrine reuptake inhibitor versus low-dose oral estradiol comparison would have been a test of noninferiority and would have required a much larger sample size. Nevertheless, data from these two arms conducted in parallel will provide important comparative data. The unbalanced sample size ratio of 2:2:3 was used to gain efficiency by creating a larger placebo group for two active arm comparisons. The target enrollment for this study was 304 women, with 87 assigned to each active arm and with 103 assigned to placebo. Assuming a rate of no more than 10% loss to follow-up, this sample size provides 90% power to detect a difference in the change in VMS frequency between groups equivalent to an effect size of 0.52 SD using a two-sided α of 0.025 for each comparison (Table 3).

Trial duration

MsFLASH trials focused on short-term (8-12 wk) effects. Factors that went into this decision included expected time to efficacy and the goal of producing relief rather than determining how long an intervention might maintain relief. Shorter trials are less expensive, have lower participant and staff burden, take advantage of the higher adherence rate associated with the early weeks of trial participation, and are more likely to avoid (in the case of VMS) the confounding effects of a spontaneous relief of symptoms. Limitations of these shorter trials, however, include lack of information on maintenance of effects and longer-term safety considerations.

Screening and recruitment

Study recruitment for all MsFLASH trials followed a similar protocol: (1) mass mailings to potential participants who responded via telephone or e-mail; (2) telephone prescreening; (3) mailing of screening questionnaires, including 2 weeks of hot flash diaries, which were returned by mail; (4) review of diaries and questionnaires for eligibility; (5) a study visit for those still eligible, where consent was completed, study laboratories were drawn, study measures were taken (including, when appropriate, screening electrocardiogram and treadmill), and participants were given an additional week of VMS diaries to complete at home; and (6) a second screening visit where diaries and questionnaires completed in the prior week were reviewed and data-entered, and eligible participants were randomized (Fig. 2).

Experience in large randomized controlled trials suggests that the backbone of a successful recruitment of generally healthy individuals is direct mailing to potential participants. Thus, all MsFLASH sites used mass mailings to targeted samples based on age and area of residence, either through purchase of commercially available mailing lists or computerized health plan membership files (Kaiser Permanente of Northern California and Group Health Research Institute in Washington State). We created informational letters, flyers, and large postcards that were mailed to invite women to the study. These materials included a "hotline" telephone number for women to call and express interest. We initially also created an e-mail account but found that women left incomplete information and retired this method. Mailings were sent to each woman up to three times, depending on the response rates for each trial. In MsFLASH 01, each study site first mailed to and conducted screening telephone calls on its own participants, but we found this to be a cumbersome and inefficient approach. To reduce costs and improve efficiency, we implemented a centralized mailing and screening protocol for subsequent trials. One site (Kaiser Permanente of Northern California) assumed responsibility for purchasing mailing lists and for printing and managing the mailings of all invitation materials through a commercial mailing firm with whom it had experience from prior studies, whereas a second site (Group Health Research Institute) assumed responsibility for all initial screening calls to women who responded to recruitment materials. Centralized recruitment methods reduced costs, improved our ability to control the rates of recruitment at each study site, and greatly facilitated the careful monitoring of yields from each step in the recruitment process.

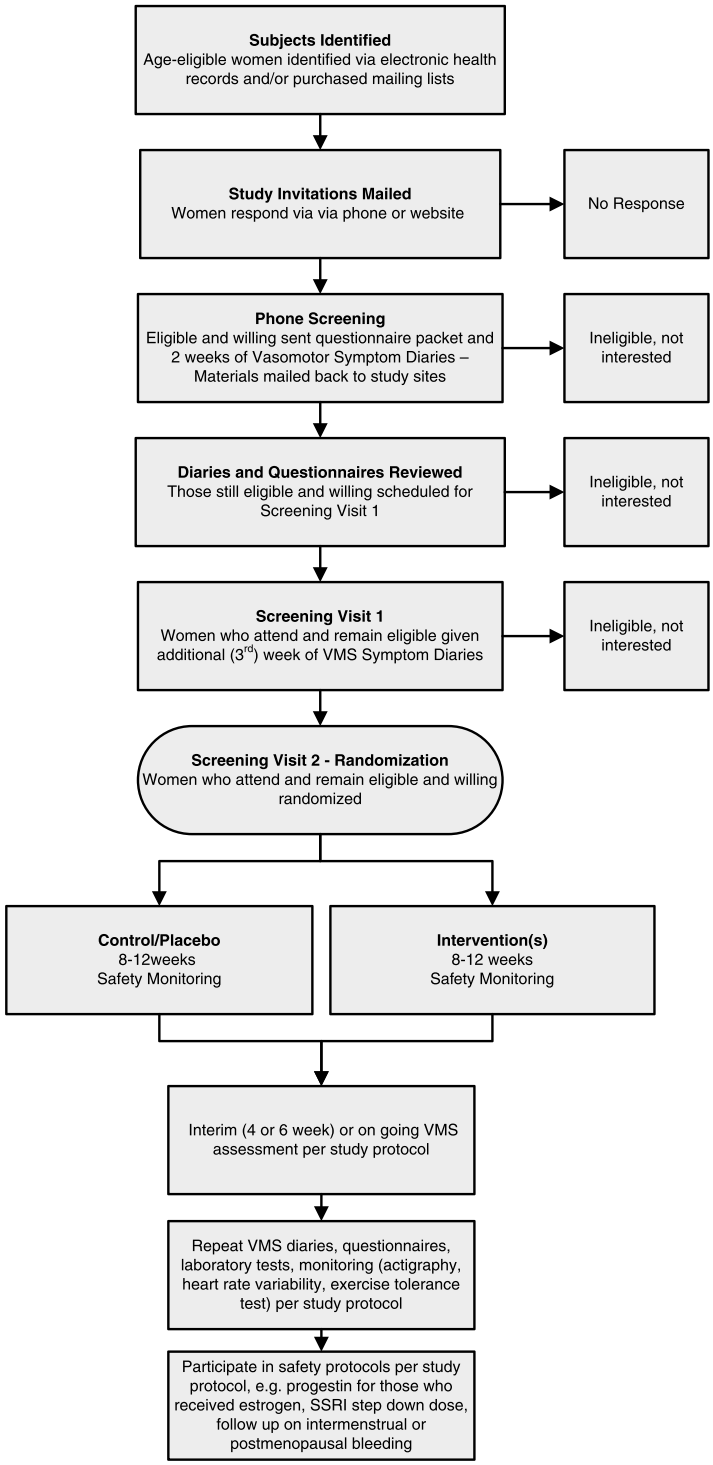

FIG. 2. Menopausal Strategies: Finding Lasting Answers to Symptoms and Health recruitment and follow-up. VMS, vasomotor symptoms; SSRI, selective serotonin reuptake inhibitor.

Although we estimated a 3.2% response rate based on responses to the Herbal Alternatives Trial[8] conducted in 2001-2003, the response rates to mass mailing ranged from 0.5% to 1% for the three network trials. The recruitment process was driven by participant response at key points, and reports showing yields at each of these points (Fig. 2) were discussed during twice-monthly calls. By carefully tracking the response rates at every level of the recruitment cascade, we were able to quickly adjust recruitment mailings to achieve target recruitment numbers without overrecruiting. By MsFLASH 03, we front-loaded recruitment with mailings larger than what calculations indicated we would need. This

worked well to quickly provide the necessary estimates for ongoing mailing numbers and to keep us ahead of schedule for recruitment.

The detailed telephone screening that we performed before bringing women into the research clinic for any testing was critical to minimizing costs because many women were found to be either ineligible or unwilling after hearing more about the study requirements. We also found that there was a much larger drop-off in the behavioral trials than in the drug trials, even after women completed their hot flash diary for 2 weeks and were deemed eligible, perhaps because they realized the amount of time the behavioral interventions would require.

Statistical methods

Randomization was implemented through the network's Web-based relational database, developed and maintained at the DCC, using a dynamic balancing algorithm[56] that stratified on network site and race for the first study and on site only for the second and third studies. For the randomization of women, documentation of consent and all eligibility data were entered into the database, where they were checked using a database algorithm. Once eligibility was established, the database randomization function was executed. The database provided a secure link between the randomization assignment and a medications inventory system that supported blinded study pill dispensation at each site. For the behavioral interventions, the randomized allocation to the intervention group was accessible in the database only for the site staff involved in the implementation of the intervention.

Our design and analysis principles rely on the intent-to-treat approach; we strive to evaluate and include all randomized participants in the primary analysis, regardless of adherence to treatment assignment or protocol requirements. Statistical research has established that exclusion of randomized participants or observed outcomes from analysis can lead to biased results of unknown magnitude or direction.[57] Furthermore, although some amount of missing data is inevitable in any study, the validity of results from any data imputation method rests on statistical assumptions that cannot be tested. Our primary approach to missing data was prevention. Follow-up data collection was required for all randomized participants, regardless of their adherence to study treatments. Our study designs and implementation further supported the intent-to-treat approach by providing the following: clear inclusion and exclusion criteria; automated eligibility verification and randomization data systems processes to reduce error; an adherence run-in with VMS diaries; a pill dispensation system that reduces the potential for staff or participant unblinding; a 1-week telephone call after the initialization of treatment regimen to address concerns and to promote adherence; a collection of adverse effect reports and encouragement to call clinics if needed; and modest monetary compensation for completing visits.

Three potential methods for quantifying outcome measures were considered: (1) posttreatment outcome adjusted for baseline measure; (2) percentage change from baseline; and (3) a threshold in percentage change (eg, proportion of women with at least a 50% decrease in VMS frequency). Using data from the first trial (baseline average hot flash frequency, 9.4 per day), hypothesized VMS outcomes (change in hot flash frequency, 3.5 SD; effect size, 0.52 SD), and a range of correlations between baseline and posttreatment scores, we simulated data and analyzed the three potential methods for quantifying outcome measures to inform our choice of optimal analytical method for MsFLASH trials. Based on these assumptions, the simulations showed that analyzing the posttreatment outcome as a function of treatment group adjusted for the baseline measure increased statistical power by up to 19 percentage points more than either the percentage change or the threshold in percentage change (Table 4).[58] This analysis method was applied to each of the VMS outcomes and to other continuous secondary outcomes. To further enrich analytical efficiency and to compensate for dropout, all analyses included outcome data collected at midstudy, with generalized estimating equations applied to account for repeated measures from each participant. Percentage change from baseline and a "clinical" definition of improvement as a 50% or greater reduction in VMS are calculated to aid in the interpretation of study results.

Quality control

Quality control was maintained in several ways. The DCC staff led in-person trainings for every trial before the trials were launched. The DCC also produced a detailed manual of operations for every trial and performed in-person audits at study sites. Data were entered via an online data entry system maintained at the DCC. The exercise intervention required

TABLE 4. *Statistical power to detect a treatment difference in vasomotor symptom frequency, based on 5,000 Monte Carlo simulations per correlation level*

Correlation[a]	Posttreatment[b]	Percentage change from baseline[c]	50% change from baseline[d]
0.3	0.91	0.81	0.72
0.5	0.95	0.92	0.81
0.7	0.99	0.99	0.91

Vasomotor symptom change from baseline to follow-up values was assumed to follow a normal distribution; sample size of 90 per treatment group. Analysis includes one follow-up time-point measure.
[a]Correlation between baseline and posttreatment measurements.
[b]Models posttreatment outcome as a continuous outcome, adjusted for baseline measure.
[c]Models percentage change from baseline as a continuous outcome.
[d]Models proportion with at least a 50% decrease in vasomotor symptoms.

weekly monitoring of exercise protocols at each site. The yoga intervention required monitoring of intervention delivery by a research specialist who confirmed adherence to the yoga protocols. For the drug trials, pill counts were performed at the end of each trial to track compliance. The DCC sent out a monthly staff newsletter that included tips to promote protocol adherence.

NETWORK STRUCTURE AND GOVERNANCE

All major scientific decisions are made by the steering committee, which is composed of a network of principal investigators, an external steering committee chair, and a representative from the NIA (Fig. 1). Network investigators are experts in midlife women's health with broad expertise in a range of disciplines.

Four subcommittees guide methods and the practical work of the network: (1) Common Measures; (2) Objective Hot Flash Monitoring Device; (3) Intervention and Implementation; and (4) Publications and Ancillary Studies. The Common Measures subcommittee was charged with identifying exposure and outcome measures that would be collected across all MsFLASH trials. The Objective Hot Flash Monitoring advisory group provided leadership on the evaluation and selection of an objective hot flash monitoring device to be considered for use in VMS trials. Each trial had trial principal investigators, typically those who had suggested a specific intervention for study, as well as site principal investigators who participated in the Intervention and Implementation committee to ensure that implementation was standardized across participating sites and that intervention-specific safety concerns were addressed. The Intervention and Implementation subcommittee for each trial had representation from DCC scientists and was responsible for the oversight of trial conduct, including study design, recruitment, intervention delivery, data collection, and trial operations. The Publications and Ancillary Studies committee developed policies and procedures related to the review of manuscript proposals, review of presentations from the network, and review of ancillary study proposals.

The NIA established an independent data and safety monitoring board (DSMB) to monitor MsFLASH trials. The DSMB is an independent, multidisciplinary group consisting of a biostatistician, an epidemiologist, a nurse scientist, and clinicians who collectively have experience in the management of symptoms related to the menopausal transition and in the conduct and monitoring of randomized clinical trials. DSMB membership was restricted to individuals with no apparent institutional, financial, scientific, or regulatory conflicts of interest. Each protocol was reviewed and approved by the DSMB, the investigational review board of each participating site, and all network committees before implementation.

All MsFLASH trials are registered with ClinicalTrials.gov (http://clinicaltrials.gov/): MsFLASH-01, Escitalopram for Menopausal Symptoms in Midlife Women (NCT 00894543); MsFLASH 02, Interventions for Relief of Menopausal Symptoms: A 3-by-2 Factorial Design Examining Yoga, Exercise, and ω-3 Supplementation (NCT 01178892); MsFLASH 03, Comparative Efficacy of Low-Dose Estradiol and Venlafaxine XR for Treatment of Menopausal Symptoms (NCT 01418209).

A public Web site (www.msflash.org) and a private SharePoint Web site were created on the first year for the dissemination of information to the public and for internal network communication and sharing of documents.

LESSONS LEARNED

MsFLASH trials have provided many opportunities for improving our methods for randomized controlled trials of menopausal symptoms. Lessons learned include the following:

1. Centralized recruitment can reduce costs, improve the ability to control the rates of recruitment at each study site, and facilitate the careful monitoring of yields from each step in the recruitment process.
2. Front-loading (as opposed to ramping up) recruitment mailings provides quick estimates for ongoing mailing numbers and assists in keeping recruitment on target or ahead of schedule.
3. Performing detailed telephone screening before bringing women into the research clinic for any testing can be critical to minimizing costs because many women will be either ineligible or unwilling after hearing more about the study requirements.
4. Requiring 4 or more VMS/day during a 3-week screening period is a high threshold of VMS for recruitment, particularly if it must be met during a 3-week period. Lowering the threshold to 2 VMS/day can improve recruitment and still provide a sufficient distribution of VMS. The ultimate choice must balance the effect size one wants to detect and the sample size one can afford because starting at a higher threshold may allow for a smaller sample size.
5. Regardless of the VMS threshold chosen, requiring 3 weeks of screening was important for reducing placebo response. For example, in MsFLASH 01, the percentage decrease in VMS from baseline to 8 week was 33%.[1] In other trials of interventions for VMS, placebo response has been as high as 46%.[60]
6. Although counterintuitive, continuing to rate VMS daily (vs intermittent diaries) is preferable. Conducting intermittent ratings requires tremendous staff time because women need to be reminded and contacted each time there is a new start-up point. Many women forget to reinitiate reporting and need a personal reminder at each time point.
7. An electronic, preferably mobile, diary could be an important aid in VMS studies. It might be more convenient for participants and would decrease data entry costs, and date/time stamping would provide data about backfilling. Although a wearable event marker could fill in part of this need, an online/phone application for VMS measurement could allow data entry of variables such as intensity, bother, and concurrent activities (eg, asleep, physically active,

sitting, concurrent stresses). MsFLASH investigators discussed this option but were concerned at the time that women would be overloaded with the number of devices we would ask them to wear (at the time, we still believed that women would be wearing a VMS monitor). Currently, there are symptom applications available, and it is now much easier to synchronize data from these applications into secure databases for research purposes.

8. Even VMS studies that attempt to include women in the menopausal transition by allowing young age (eg, as young as 40 y) may find that their study population is predominantly older (in our case aged 48-62 y). Targeting this older group may decrease recruitment costs.
9. Overnight urine collection may be difficult to implement because women who use public transportation or who come to the study clinic directly from work may find that transporting urine is unacceptable.
10. In considering a multifactorial design for behavioral interventions, weigh the potential cost savings related to a shared control group against potential complexities. These may include access limitations for women owing to facility location, screening complexities (women must be eligible for all interventions), and the challenge of explaining complex protocols to potential participants.
11. Following women after an intervention has been discontinued can provide important data about the speed with which symptoms rebound. This information is clinically relevant to letting women know what to expect should they choose to use an intervention. Ongoing follow-up can also provide information about placebo effects. For example, in MsFLASH 01, after the medication was discontinued at 8 weeks, there was VMS rebound among women in the escitalopram group, but not among women in the placebo group.[1] Women were not unblinded until week 12, and VMS frequency in the two groups was identical at 11 weeks. It is also possible to gain information about women's intention to continue treatment after they are unblinded.

CONCLUSIONS

The first National Institutes of Health–funded menopause network (MsFLASH) has successfully established network operations—designing and conducting three randomized controlled trials that collectively studied six interventions in randomized controlled clinical trials during 5 years of funding. We designed, conducted, and analyzed these VMS trials according to the most rigorous principles of randomized trials. We standardized the methods across trials to promote broader comparisons and are publishing the details of these methods to assist in comparisons with other studies.

We share these experiences to encourage and support others as they design and conduct similar randomized trials, particularly those trials focused on interventions for the relief of menopausal symptoms. The use of standardized methods to determine eligibility and to assess VMS and related outcomes will greatly improve the ability to compare the effectiveness of various treatment modalities across trials and therefore will enable women and their healthcare providers to make more informed choices when choosing treatments to relieve menopausal symptoms.

We welcome collaborations with other researchers seeking to find safe and effective treatments for menopausal symptoms and other health concerns of midlife and older women, as well as proposals for ancillary studies and analyses using the MsFLASH Biobank and database.

REFERENCES

1. 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.
2. Gold EB, Bromberger J, Crawford S, et al. Factors associated with age at natural menopause in a multiethnic sample of midlife women. *Am J Epidemiol* 2001;153:865-874.
3. Gold EB, Sternfeld B, Kelsey JL, et al. Relation of demographic and lifestyle factors to symptoms in a multi-racial/ethnic population of women 40-55 years of age. *Am J Epidemiol* 2000;152:463-473.
4. McKinlay S, Jefferys M, Thompson B. An investigation of the age at menopause. *J Biosoc Sci* 1972;4:161-173.
5. US Food and Drug Administration. *Guidance for Industry: Estrogen and Estrogen/Progestin Drug Products to Treat Vasomotor Symptoms and Vulvar and Vaginal Atrophy Symptoms—Recommendations for Clinical Evaluation*. Rockville, MD: Division of Drug Information (HDF-240), Center for Drug Evaluation and Research, 2003.
6. Soules MR, Sherman S, Parrott E, et al. Executive summary: Stages of Reproductive Aging Workshop (STRAW). *Climacteric* 2001;4:267-272.
7. Sloan JA, Loprinzi CL, Novotny PJ, Barton DL, Lavasseur BI, Windschitl H. Methodologic lessons learned from hot flash studies. *J Clin Oncol* 2001;19:4280-4290.
8. Newton KM, Buist DS, Keenan NL, Anderson LA, LaCroix AZ. Use of alternative therapies for menopause symptoms: results of a population based survey. *Obstet Gynecol* 2002;100:18-25.
9. Avis NE, Kaufert PA, Lock M, McKinlay SM, Vass K. The evolution of menopausal symptoms. *Baillieres Clin Endocrinol Metab* 1993;7:17-32.
10. 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.
11. Smith MT, Wegener ST. Measure of sleep. *Arthritis Rheum* 2003;49: S184-S196.
12. Kroenke K, Zhong X, Theobald D, Wu J, Tu W, Carpenter JS. Somatic symptoms in patients with cancer experiencing pain or depression: prevalence, disability, and health care use. *Arch Intern Med* 2010;170: 1686-1694.
13. Gilbody S, Richards D, Barkham M. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *Br J Gen Pract* 2007;57:650-652.
14. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007;22:1596-1602.
15. 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.
16. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL). A measure of primary symptom dimensions. *Mod Probl Pharmacopsychiatry* 1974;7:79-110.
17. Lipman RS, Covi L, Shapiro AK. The Hopkins Symptom Checklist (HSCL)—factors derived from the HSCL-90. *J Affect Disord* 1979;1:9-24.
18. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983;24:385-396.
19. Krebs EE, Lorenz KA, Bair MJ, et al. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. *J Gen Intern Med* 2009;24:733-738.
20. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singapore* 1994;23:129-138.
21. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. *J Pain* 2004;5:133-137.

22. Hilditch JR, Lewis J, Peter A, et al. A menopause-specific quality of life questionnaire: development and psychometric properties. *Maturitas* 1996;24:161-175.
23. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. *Psychopharmacol Bull* 1993;29:321-326.
24. Rosen R, Brown C, Heiman J, et al. The Female Sexual Function Index (FSFI): a multidimensional self-report instrument for the assessment of female sexual function. *J Sex Marital Ther* 2000;26:191-208.
25. Derogatis L, Clayton A, Lewis-D'Agostino D, Wunderlich G, Fu Y. Validation of the Female Sexual Distress Scale—Revised for assessing distress in women with hypoactive sexual desire disorder. *J Sex Med* 2008;5:357-364.
26. Newton K, LaCroix A, Leveille S, Rutter C, Keenan N, Anderson L. Women's beliefs and decisions about hormone replacement therapy. *J Womens Health* 1997;6:459-465.
27. 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.
28. Smith S, Trinder J. Detecting insomnia: comparison of four self-report measures of sleep in a young adult population. *J Sleep Res* 2001;10:229-235.
29. Morin C. *Insomnia: Psychological Assessment and Management*. New York, NY: Guilford Press, 1993.
30. Morin CM, Vallieres A, Guay B, et al. Cognitive behavioral therapy, singly and combined with medication, for persistent insomnia: a randomized controlled trial. *JAMA* 2009;301:2005-2015.
31. Morin CM, Colecchi C, Stone J, Sood R, Brink D. Behavioral and pharmacological therapies for late-life insomnia: a randomized controlled trial. *JAMA* 1999;281:991-999.
32. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989;28:193-213.
33. Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh Sleep Quality Index. *J Psychosom Res* 1998;45:5-13.
34. Buysse DJ, Germain A, Moul DE, et al. Efficacy of brief behavioral treatment for chronic insomnia in older adults. *Arch Intern Med* 2011; 171:887-895.
35. de Souza L, Benedito-Silva AA, Pires ML, Poyares D, Tufik S, Calil HM. Further validation of actigraphy for sleep studies. *Sleep* 2003;26:81-85.
36. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* 2003;26:342-392.
37. Cohen LS, Soares CN, Vitonis AF, Otto MW, Harlow BL. Risk for new onset of depression during the menopausal transition: the Harvard Study of Moods and Cycles. *Arch Gen Psychiatry* 2006;63:385-390.
38. Freeman EW, Sammel MD, Liu L, Gracia CR, Nelson DB, Hollander L. Hormones and menopausal status as predictors of depression in women in transition to menopause. *Arch Gen Psychiatry* 2004;61:62-70.
39. Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. *Med Care* 2010;48:1007-1014.
40. Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. *Pain* 1983;17:197-210.
41. Hann DM, Jacobsen PB, Azzarello LM, et al. Measurement of fatigue in cancer patients: development and validation of the Fatigue Symptom Inventory. *Qual Life Res* 1998;7:301-310.
42. Carpenter JS, Storniolo AM, Johns S, et al. Randomized, double-blind, placebo-controlled crossover trials of venlafaxine for hot flashes after breast cancer. *Oncologist* 2007;12:124-135.
43. Elkins G, Marcus J, Stearns V, Hasan Rajab M. Pilot evaluation of hypnosis for the treatment of hot flashes in breast cancer survivors. *Psychooncology* 2007;16:487-492.
44. Suvanto-Luukkonen E, Koivunen R, Sundstrom H, et al. Citalopram and fluoxetine in the treatment of postmenopausal symptoms: a prospective, randomized, 9-month, placebo-controlled, double-blind study. *Menopause* 2005;12:18-26.
45. Loprinzi CL, Sloan JA, Perez EA, et al. Phase III evaluation of fluoxetine for treatment of hot flashes. *J Clin Oncol* 2002;20:1578-1583.
46. Loprinzi L, Barton DL, Sloan JA, et al. Pilot evaluation of gabapentin for treating hot flashes. *Mayo Clin Proc* 2002;77:1159-1163.
47. Pruthi S, Qin R, Terstreip SA, et al. A phase III, randomized, placebo-controlled, double-blind trial of flaxseed for the treatment of hot flashes: North Central Cancer Treatment Group N08C7. *Menopause* 2012;19:48-53.
48. Bootzin RR. The role of expectancy in behavior change. In: White L, Turskey B, Schwartz G, eds. *Placebo: Theory, Research and Mechanisms*. New York, NY: Guilford Press, 1985:196-210.
49. Gross D. On the merits of attention-control groups. *Res Nurs Health* 2005;28:93-94.
50. Srinivasan S, Liu C, Mitchell CM, et al. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PLoS ONE* 2010;5:e10197.
51. Carpenter JS, Andrykowski MA, Freedman RR, Munn R. Feasibility and psychometrics of an ambulatory hot flash monitoring device. *Menopause* 1999;6:209-215.
52. Carpenter JS, Monahan PO, Azzouz F. Accuracy of subjective hot flush reports compared with continuous sternal skin conductance monitoring. *Obstet Gynecol* 2004;104:1322-1326.
53. Thurston RC, Blumenthal JA, Babyak MA, Sherwood A. Emotional antecedents of hot flashes during daily life. *Psychosom Med* 2005; 67:137-146.
54. Carpenter JS, Newton KM, Sternfeld B, et al. Laboratory and ambulatory evaluation of vasomotor symptom monitors from the Menopause Strategies Finding Lasting Answers for Symptoms and Health network. *Menopause* 2012;19:664–671.
55. Anderson GL, Prentice RL. Individually randomized intervention trials for disease prevention and control. *Stat Methods Med Res* 1999;8:287-309.
56. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 1975; 31:103-115.
57. Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials* 3rd ed. New York, NY: Springer-Verlag, 1998.
58. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Med Res Methodol* 2001;1:6.
59. Sood R, Sood A, Wolf SL, et al. Paced breathing compared with usual breathing for hot flashes. *Menopause* 2013;20:179-184.

