Methods for the Design of Vasomotor Symptom Trials: The MsFLASH Network

Katherine M. Newton, PhD, **Janet S. Carpenter, PhD, RN, FAAN**, **Katherine A. Guthrie, PhD**, **Garnet L. Anderson, PhD**, **Bette Caan, Dr PH**, **Lee S. Cohen, MD**, **Kristine E. Ensrud, MD, MPH**, **Ellen W. Freeman, PhD**, **Hadine Joffe, MD**, **Barbara Sternfeld, PhD**, **Susan D. Reed, MD, MPH,**, **Sheryl Sherman, PhD**, **Mary D. Sammel, ScD**, **Kurt Kroenke, PhD**, **Joseph C. Larson, MS**, and **Andrea Z. LaCroix, PhD**

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflicts of interest/financial disclosures:

Author	Conflicts/Disclosures
Katherine M. Newton, PhD	**Research Support** Integrated Diagnostics
Janet S. Carpenter, PhD, RN, FAAN,	NONE
Katherine A. Guthrie, PhD,	NONE
Garnet L. Anderson, PhD,	NONE
Bette Caan, Dr PH,	NONE
Lee S. Cohen, MD,	**Research Support:** Astra-Zeneca Pharmaceuticals; Bristol-Myers Squibb; Cephalon, Inc.; GlaxoSmithKline; National Institute on Aging; National Institute of Mental Health; Ortho-McNeil Janssen; Sunovion Pharmaceuticals, Inc. **Advisory/Consulting:** PamLab LLC; Noven Pharmaceutical **Honoraria:** None **Royalty/patent, other income:** None
Kristine E. Ensrud, MD, MPH,	**Data Monitoring Committee** Merck Sharpe & Dohme.
Ellen W. Freeman, PhD,	**Research Support:** Forest Laboratories, Inc, Xanodyne Pharmaceuticals, Bionovo, Wyeth
Hadine Joffe, MD,	**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. **Speaking honoraria:** Eli Lilly, GlaxoSmithKline. **Advisory/Consulting:** Abbott Laboratories, Eli Lilly, Forest Laboratories, Inc., JDS-Noven Pharmaceuticals, Sanofi-Events, Sepracor Inc., Sunovion, Wyeth-Ayerst Pharmaceuticals.
Barbara Sternfeld, PhD,	NONE
Susan D. Reed MD, MPH,	NONE
Sheryl Sherman, PhD,	NONE
Mary D. Sammel, ScD,	**Consultant** for Swiss Precision Diagnostics GmbH; **Statistical reviewer** for Am. J. Obstetrics and Gynecology
Kurt Kroenke, PhD,	NONE
Joseph C. Larson, MS,	NONE
Andrea Z. LaCroix, PhD	**Scientific Advisory Board and Research Support:** Global Longitudinal study of Osteoporosis in Women (GLOW), unrestricted research grant to the University of Massachusetts, Center for Outcomes Research, Proctor and Gamble and Sanofi. **Attendee:** one-time consultant's meeting for Pfizer to evaluate the evidence on hormone therapy and breast density

Data Coordinating Center, Fred Hutchinson Cancer Research Center, Seattle, WA (Drs. Guthrie, Anderson, Larson, and LaCroix); Departments of Obstetrics/Gynecology and Epidemiology, University of Washington School of Medicine, Seattle, WA (Dr. Reed); Departments of Obstetrics/ Gynecology and Psychiatry (Dr. Freeman), Center for Clinical Epidemiology and Statistics (Dr. Sammel), University of Pennsylvania School of Medicine, Philadelphia, PA; Division of Research, Kaiser Permanente of Northern California, Oakland, CA (Drs. Caan and Sternfeld); Group Health Research Institute, Seattle, WA (Dr. Newton); Massachusetts General Hospital/Harvard Medical School, Boston, MA (Drs. Cohen and Joffe); School of Nursing, Indiana University, Indianapolis, IN (Dr. Carpenter); National Institute on Aging, US National Institutes of Health, Bethesda, MD (Dr. Sherman);VA Medical Center / University of Minnesota, Minneapolis, MN (Dr. Ensrud); VA Center of Excellence for Implementing Evidence-Based Practice and Indiana University School of Medicine, Indianapolis, IN (Dr. Kroenke)

Abstract

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

Methods—Established in response to a National Institute 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: 1) a randomized, double-blind, placebo-controlled clinical trial designed to evaluate effectiveness of the selective serotonin reuptake inhibitor escitalopram in reducing vasomotor symptom frequency and severity; 2) a 2×3 factorial design trial to test three different interventions (yoga, exercise, and omega-3 supplementation) for 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 compared to placebo. The network's structure and governance are also discussed.

Conclusions—The methods used *and lessons learned* in the MsFLASH trials are shared to encourage and support the conduct of similar trials and encourage collaborations with other researchers.

INTRODUCTION

The long term objective of the National Institute on Aging's RFA-AG-08-004, "New Interventions for Menopausal Symptoms" (U01) was to accelerate progress in identifying effective remedies for vasomotor symptoms 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 (NICHD), the National Center for Complementary and Alternative Medicine (NCCAM), and the Office of Research on Women's Health (ORWH) with a goal of creating a network of scientists highly knowledgeable about the menopausal transition and experienced in the conduct of women's health trials. The purpose of this paper 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 vasomotor symptoms trials in this multicenter national menopause network.

The MsFLASH network is funded through a cooperative agreement and is comprised of the Data Coordinating Center (DCC) and 5 study sites, with representation from the four funding agencies (Figure 1). Network investigators have conducted three randomized controlled trials examining six different interventions for relief of menopause symptoms. The primary outcomes for all three trials included vasomotor symptom frequency, and severity or bother. The first trial was a standard placebo-controlled study to determine the efficacy and tolerability of 10–20 mg/day escitalopram, a selective serotonin reuptake inhibitor, compared to placebo pills.[1] The second trial employed a three by two factorial design to compare the effects of yoga and exercise separately to a common wait-list control group, and simultaneously to compare omega-3 fatty acid capsules to placebo capsules. The third trial compares the efficacy of low-dose oral estradiol and the serotonin-norepinephrine uptake inhibitor venlafaxine XR to placebo.

METHODOLOGICAL ISSUES IN THE DESIGN OF THE MsFLASH TRIALS

The NIH charge to the network was to conduct rapid throughput randomized trials of novel and understudied currently available interventions postulated to alleviate menopausal symptoms. Reduction in vasomotor symptoms (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 comparisons of intervention results across trials; 2) Generalizability of trial results is improved 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) The vast majority of women experiencing menopause-related vasomotor symptoms have 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 the network investigators; 7) Women who experience vasomotor symptoms for prolonged durations (> 5 years) after cessation of ovarian function are a small but important subgroup of women and network investigators are committed to advancing therapies in this group; and 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. The network investigators wanted to study a variety of pharmaceutical and behavioral interventions, particularly interventions that were frequently recommended without strong evidence (e.g., 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 drug and matching placebos, side effects, dosing issues, and known information about effectiveness. There was consideration about women's willingness to receive each planned treatment intervention given potential side effect profiles. Other considerations included adherence (ability to maintain an 8–12 week intervention), tolerability, and safety. To the extent possible, masking of interventions was incorporated,

given the established placebo response found in most studies of interventions for menopause symptoms.

Common eligibility and exclusion criteria

A detailed comparison of eligibility criteria used in previous trials by network investigators and others was undertaken and 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 and/or to exclude women for safety reasons.

Age—In the Study of Women's Health Across the Nation (SWAN) the median age at natural menopause was 51.4 years[2] and menopause symptoms peaked around the time of menopause[3]. In an earlier study McKinlay found that 96.4% of women were postmenopausal by age 54[4]. We therefore recruited women ages 40–62 years to capture women likely to be in the menopause transition or within 5 years beyond menopause, the time when symptoms are most prevalent. This was a pragmatic decision to maximize mailings to women most likely to meet our entry criteria. Over time we found that most responses to recruitment mailings were from women aged 47–62 years, and mailings were targeted to women in this age range, although women aged 40–46 who contacted us about the trial and met other eligibility criteria remained eligible.

Menopausal Status—Studies of menopause symptom treatments (including 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 menopause transition and those in the first 5 years post menopause because women frequently seek therapy for symptoms during the menopause transition. Guided by the Stages for Reproductive Aging Workshop criteria in defining the stages of menopause,[6] women were classified as "Late Menopausal Transition" when they skipped two consecutive menstrual cycles not associated with pregnancy or breast feeding and had an interval of amenorrhea of > 60 days within the last year. Women were classified as "Postmenopausal" when their last menstrual period was 12 or more months ago. 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 evaluation of treatment effectiveness during short-term trials.

Network investigators discussed at length whether to include women with oophorectomy and hysterectomy and decided that the same criteria should be applied to all women with menopausal VMS regardless of menopause type. Thus, women with *bilateral* oophorectomy were eligible if they met other inclusion criteria. For women with hysterectomy *and at least one ovary*, FSH level >20 mIU/mL, and estradiol level <=50 pg/mL, were used to include those women most likely to be post-menopausal (acknowledging that some of these women might have been classified as being in the late transition, had they had a uterus).

Vasomotor Symptom (VMS) Criteria, Assessment and Primary Outcomes—Our primary interest was in VMS frequency and severity or bother. While we considered the use of electronic diaries, the cost of electronic devices and of creating on-line 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, in which women record 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 did have concerns about the burden of even this twice daily recording,

including whether women would complete diaries, or that they might complete them retrospectively. But given the evidence about prior successful use with minimal burden and our desire to compare our results to the results of other studies we chose to employ this method for all the 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 diaries throughout the study, while others thought that this was too great a burden and suggested that diaries be done only during selected weeks. Ultimately diaries were done continuously in MsFLASH 01 and 03, and at baseline, 6 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 prior to randomization (Figure 1). In the first trial (MsFLASH 01) that tested a pharmacologic agent, we required: 28 VMS per week recorded on daily diaries for 3 weeks; VMS rated as bothersome or severe on 4 or more days per week; and that VMS frequency in week 3 did not decrease >50% from the average weekly levels in weeks 1 and 2.

We carefully considered the minimum 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 per day, or 50 to 60 per week at baseline[5]. The FDA defines moderate as a sensation of heat with sweating, but able to continue activity, while severe VMS require cessation of activity. We chose lower thresholds for both frequency and bother/severity to be more inclusive and generalizable than typical FDA monitored trials. Our rationale for requiring a minimum number of VMS perceived as bothersome or severe was that women seek treatment not only because of VMS frequency but 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, e.g. a woman may rate her VMS as severe, but 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 the placebo response in the trials.

In the second trial, we lowered the VMS criteria for two reasons. First, we were testing several non-pharmacologic interventions that might be attractive to a broader spectrum of women with VMS. Second, recruitment was challenging due to several study-specific entry criteria. When analyses of screening diaries showed that insufficient VMS frequency was the most common reason for ineligibility in MsFLASH 01, we evaluated the potential impact of using a lower threshold. We examined both the range of baseline VMS symptoms in the MsFLASH 01, and the range of baseline VMS in a prior study of herbal therapies for menopause symptoms that required only 2 VMS per day.[8] We concluded that a lower threshold would potentially increase recruitment with minimal effects on our results. We amended the criteria to: 14 VMS per week recorded on daily hot flash diaries for 3 weeks; VMS rated as bothersome or severe 4 or more times per week; and the VMS frequency in week 3 did not decrease >50% from the average weekly levels in weeks 1 and 2.At the end of the study we found that for the 67 women randomized prior to this (requiring 4 VMS / day), the mean VMS / 24 hour day was 8.5 (SD 4.0). For the 288 women randomized after the change, the mean VMS / 24 hour day was 7.4 (SD 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

using selective estrogen receptor modulators (SERMs) or aromatase inhibitors in the past 2 months. These exclusions were incorporated to avoid potential carry-over 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 a guiding principle the use of well-validated and psychometrically sound self-report questionnaires. We sought measures of mood that were not overly sensitive to somatic symptoms. We favored shorter scales over longer scales to lessen participant burden. We included selected global measures (e.g. the MenQOL) because they are widely used in other menopause studies.

Sleep Measures—Sleep disturbance and related complaints are a primary reason why women seek treatment for VMS.[9] To measure insomnia symptoms, we used the Insomnia Severity Index (ISI)[10–12], a valid and reliable self-administered instrument that measures perception of current (past two weeks) insomnia symptoms. The index has 7 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 0–4 point scale (total score 0–28), with higher scores suggesting more severe insomnia symptoms. The absence of insomnia is indicated by scores 0–7, subthreshold or mild insomnia by scores 8–14, clinical insomnia of moderate severity by scores 15–21, and severe clinical insomnia by scores 22–28. Trials of pharmacologic and behavioral interventions in patients with insomnia have suggested that the ISI is sensitive in measuring treatment response.[13, 14]

To assess self-reported sleep quality, we used the Pittsburgh Sleep Quality Index (PSQI), a validated measure of subjective sleep quality and sleep disturbances over a one-month time period Smith.[15] The PSQI assesses subjective sleep quality, latency, duration, and efficiency; sleep disturbances; use of sleeping medication; and daytime dysfunction.[15, 16] Global PSQI scores range from 0–21 with higher scores indicating poorer sleep quality. Cutoffs of 5[16] and 8[17] have been reported to indicate poor sleep quality. The PSQI has been shown to be sensitive in measuring response to cognitive behavioral therapy in randomized trials conducted in patients with insomnia.[18]

To measure sleep-wake patterns objectively, participants wore an Actiwatch 2 (Philips Respironics, Bend, OR, USA) for 7 days at baseline and follow up (8 or 12 weeks) The Actiwatch is a small device 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.[19, 20] Participants were instructed to wear the Actiwatch continuously for 7 nights/8 days one week prior to baseline and closeout visits, removing it only for bathing, or situations in which 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 any times the actigraph was removed. Sleep logs were used to aid in editing the 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 was challenging for both the women and the study staff.

Mood—Another primary reason women initiate menopausal treatment is mood disturbance. The menopause transition is a time of increased risk for new onset or reoccurrence of clinical depression and depressive symptoms.[21, 22] To assess depression, we used either the 8-item version (in behavioral intervention studies) or 9-item version (in antidepressant studies where the 9th item on assessing thoughts of death or self-harm was deemed important) of the depression module of the Patient Health Questionnaire (PHQ).[23] Both versions of the PHQ 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 GAD-7, which can likewise be scored either continuously (as an anxiety severity score) or categorically (with cutpoints that indicate a probable DSM anxiety disorder).[23] Because the GAD-7 is not 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.[24–26] We used the Perceived Stress Scale (PSS), a widely-used, validated self-report of perceived stress, to assess stress as an independent construct associated with VMS.[27]

Pain—MsFLASH investigators were interested in exploring the relationship between menopause and pain, and chose the PEG, a 3-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).[28, 29] The PEG scale has shown high internal consistency, construct validity and responsiveness to change in outpatient populations.[28] Responsiveness to change in pain in a randomized trial of adults with musculoskeletal pain has been found to be equal or superior to several longer pain scales.[29]

Quality of Life—The Menopause specific Quality of Life questionnaire (MenQOL)[30] was chosen because it is a global health-related quality-of-life scale designed specifically for use in the menopause, and because it has been used frequently in menopause research.

VMS Interference—Perceived hot flash interference was evaluated using the Hot Flash Related Daily Interference Scale.[31] This 10-item scale measures a woman's perceptions of the degree to which VMS interfere with nine daily life activities; the tenth item measures interference with overall quality of life. This scale was modeled after items on the Brief Pain Inventory[32] and Brief Fatigue Inventory[33] which assess the degree pain or fatigue interfere with similar activities. Participants rate the degree to which VMS have interfered with each item during the previous week using a 0 (do not interfere) to 10 (completely interfere) scale. Recent structural equation modeling suggests this is a uni-dimensional 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[34] and behavioral interventions.[35]

Sexual Function and Sexual Distress—Sexual function and distress are common complaints of mid-life 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 the Female Sexual Function Index (FSFI)[36] and a single item from the Female Sexual Distress (FSDS) Scale in all trials.[37]. The full FSDS was implemented in MsFLASH 02.

Methods relevant to the placebo response

Almost all studies of interventions for menopause symptoms have shown VMS decreases in both the intervention and control groups. In placebo groups, VMS frequency has been shown to decrease 20% to 60% from baseline.[7, 38–42] This variable decrease in VMS has been attributed to regression to the mean, natural resolution of symptoms, placebo response,

and fatigue with symptom recording over long time 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 in week 3 compared to weeks 1 and 2, were ineligible.

All drug interventions in the MsFLASH trials had double-blind designs using matched placebo capsules. We also sought believable and appropriate control groups for the 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 aspect of the intervention other than attention or expectancy.[43] As Gross points out, the assumption is that the attention and expectation that "good things will happen" are not active ingredients of the intervention, and can somehow be separated.[44] Arguably, for the 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.[43] 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 omega-3 fatty acid pills. Thus, all women had some expectation of effect.

Blood, urine, saliva, vaginal sample collections 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 a 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 aliquotted 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 aliquotted into each of nine 1 mL cryovials. The buffy was removed and aliquotted into two 0.5 mL cryovials. The DCC provided barcoded blood ID labels suitable for −70°C freezer storage to ensure that samples, forms, database and cryovials were linked via a unique ID and 2-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 which was provided along with a gel ice pack. Approximately 9 ml of urine was centrifuged and 0.5 mL aliquotted into each of six 1 mL cryovials for storage in the MsFLASH repository at FHCRC.

In MsFLASH 02, saliva samples (for cortisol) were collected. Four samples were obtained on each of 2 consecutive days at baseline and study completion (week 12) (16 samples total). Salivettes were placed in a zip-lock bag in the participant's freezer until they were transported back to the clinical site. At the clinical site, the samples were processed and placed in a -70 degree Celsius freezer, and then batched for transport to the DCC. The DCC

provided a form and sample ID labels that linked the salivary samples to the participant and visit.

In MsFLASH 03 women were consented separately for an ancillary study of the vaginal microbiome. Vaginal swabs were collected at baseline. Women were asked to collect swabs at home (using techniques perfected by Mitchell and Fredricks with excellent compliance, safety and specimen quality)[45] on Days 1–14 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 lab 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 the MsFLASH investigators was that all trials would employ both subjective and objective VMS monitoring. Objective VMS measurement has been recommended as an adjunct to subjective measurement of frequency, severity, bother, and/or duration. The potential advantages of objective monitoring are that results should be unbiased by placebo effects[34], sleep-wake cycles[46, 47], and reporting biases[46, 48]. We evaluated three potential monitors for use in the MsFLASH trials:[49] the Freedman monitor, the Bahr Monitor™ and the Biolog™. 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 missing data. The Biolog™ performed more consistently but there were ongoing problems with electrode availability. Therefore the decision was made not to use objective VMS monitoring in the MsFLASH studies.[49] While objective monitoring devices can be purchased, and despite NCCAM'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 the clinical trial designs

MsFLASH 01 was a randomized, double-blind, placebo-controlled clinical trial designed to evaluate effectiveness of an SSRI (escitalopram) in reducing VMS frequency and severity. In order to mimic clinical practice the design included a blinded dose escalation for non-responders halfway through the 8-week trial. Because the dose escalation was dependent on response in the first 4 weeks, the trial does not provide a randomized comparison of the two doses but rather an estimate of realistic effectiveness of the drug within a narrow range of doses. This trial also included a VMS symptom assessment at 3 weeks following 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 compared to white women. To improve power for this interaction test, accrual was restricted to assure 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% versus 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 the 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 the urine was unpleasant and they often refused. We eliminated urine collection from MsFLASH 02 and 03.

MsFLASH 02 applied a 2×3 factorial design to test three different interventions (yoga, exercise and omega-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 (omega-3 or placebo) for 12 weeks. The factorial design was selected as an efficient approach for testing all three interventions, and to assure that all women would participate in at least one intervention (omega-3 supplement or placebo capsules) during the primary intervention period and hence would have some expectancy of benefit (see discussion of placebo effect above). We assumed that the effects of any of these approaches would be relatively modest in reducing VMS and that the interventions would likely operate through independent pathways such that any interaction between omega 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 one larger behavioral control group to make two active arm comparisons, with sample size ratios 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.[50] Ninety participants in each of the yoga and exercise groups and 135 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 omega-3 fatty acid supplementation (45 in yoga, 45 in exercise and 68 in control arm) provided 90% power for a difference in mean reduction of 0.40 SD units, with 2-sided significance level of 0.025. The total enrollment goal was 374, allowing 10% inflation for loss to follow-up and an extra 10% in the yoga and exercise groups to account for increased variability in the outcome measurements due to a range of compliance to the behavioral interventions. This trial included an ancillary study to measure heart rate variability and salivary cortisol to evaluate potential associations with outcomes and baseline participant characteristics.

MsFLASH 02 taught us several things we might do differently in future similar behavioral trials. Because of the factorial design, and the types of interventions, the facilities used were different for yoga and exercise. This created limitations due to women's inability to travel to one or the other site. Using the same site for both interventions, and finding sites on public transit lines would have lessoned these challenges. In MsFLASH 02 we also asked women to wear multiple devices – an actigraph for sleep and an accelerometer for activity measures. We did this 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 (i.e. weekly yoga classes with home practice, 3 times per week in-person exercise sessions, and omega three capsules) created efficiencies, but also meant that women who might have been eligible for one intervention, if studies alone, were excluded because they were ineligible based on another intervention (e.g. 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 (SNRI) venlafaxine and low-dose oral estradiol versus placebo for reducing VMS frequency compared to placebo. 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 SNRI versus low-dose oral estradiol comparison would have been a test of non-inferiority, 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 ratios of 2:2:3 were used to gain efficiency by creating one 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 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 change in VMS frequency between groups equivalent to an effect size of 0.52 SD using a two-sided alpha = 0.025 test for each comparison (Table 3).

Trial Duration

The MsFLASH trials focused on short-term (8 to 12 week) effects. Factors that went into this decision included expected time to efficacy, and the goal of trying to produce relief rather than determining how long an intervention might maintain relief. Shorter trials are less expensive, have a 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 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 the MsFLASH trials followed a similar protocol: 1) mass mailings to potential participants who responded via phone or email: 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 labs were drawn, study measures were taken (including, when appropriate, screening ECG and treadmill), and participants were given one additional week of VMS diaries to complete at home; 6) a second screening visit where diaries and questionnaires completed in the prior week were reviewed and data-entered, and eligible participants were randomized (Figure 2).

Experience in large RCTs suggests that the backbone of 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 Northern California and Group Health in Washington State). We created informational letters, flyers, and large post cards that were mailed to invite women to the study. These materials included a "hot line" phone number for women to call and express interest. We initially also created an email account but found that women left incomplete information and retired this method. Mailings were sent to each woman up to three times, depending on response rates for each trial. In MsFLASH 01 each study site first mailed and conducted screening phone calls for its own participants, but we found this was 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 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 they had experience from prior studies, while 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 rates of recruitment at each study site, and greatly facilitated the careful monitoring of yields from each step in the recruitment process.

Although we estimated a 3.2% response rate based on responses to the Herbal Alternatives Trial (HALT)[39], conducted in 2001–2003, response rates to mass mailing ranged from 0.5–1% for the 3 network trials. The recruitment process was driven by participant response at key points and reports showing yields at each of these points (Figure 2) were discussed during twice-monthly calls. By carefully tracking response rates at every level of the

recruitment cascade we were able to quickly adjust recruitment mailings to achieve target recruitment numbers without over-recruiting. By MsFLASH 03 we front loaded recruitment with mailings larger than are 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 phone screening prior to brining women into the research clinic for any testing was critical to minimizing costs, since many women were either 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 hot flash diary for two 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[51] that stratified on network site and race for the first study, and site only for the second and third studies. To randomize a woman, documentation of consent and all eligibility data were entered into the database where it was 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 intervention group was accessible in the database only for site staff involved in the implementation of the intervention.

Our design and analysis principles rely on the intent-to-treat (ITT) 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.[52] 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 an ITT approach by providing 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 potential for staff or participant unblinding, a one-week phone call after initialization of treatment regimen to address concerns and promote adherence, collection of side effect reports and encouragement to call clinics if needed, and modest monetary compensation for completing visits.

Three potential methods of quantifying outcome measures were considered: 1) post-treatment outcome adjusted for the baseline measure; 2) percent change from baseline; and 3) a threshold in percent change (e.g. proportion of women with at least a 50% decrease in VMS frequency). Using data from the first trial (baseline average HF frequency of 9.4 per day), hypothesized VMS outcomes (standard deviation (SD) of change in HF frequency 3.5, effect size of 0.52 SD) and a range of correlations between baseline and post-treatment scores, we simulated data and analyzed the three potential methods of quantifying outcome measures to inform our choice of optimal analytical method for MsFLASH trials. Based on these assumptions, the simulations showed that analyzing the post-treatment 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 percent change or a threshold in percentage

change (Table 4).[53] This analysis method was applied to each of the VMS outcomes and to other continuous secondary outcomes. To further enrich the analytic efficiency and compensate for drop-out, all analyses included outcome data collected mid-study with generalized estimating equations applied to account for repeated measures from each participant. Percent change from baseline and a "clinical" definition of improvement as 50% or greater reduction in VMS are calculated to aid in 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 on-line data entry system maintained at the DCC. The exercise intervention required weekly monitoring of exercise protocols at each site. The yoga intervention required monitoring of the intervention delivery by a research specialist who confirmed adherence to the yoga protocols. For the drug trials, pill counts were done 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 comprised of the network principal investigators, an external Steering Committee Chair and a representative from the NIA (Figure 1). The network investigators are experts in mid-life 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 Committee was charged with identifying exposure and outcome measures that would be collected across all the 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 vasomotor symptom trials. Each trial had trial principal investigators (PIs), typically those who had suggested a specific intervention for study, as well as site PI(s) who participated in an Intervention and Implementation committee to assure that implementation was standardized and across participating sites and intervention-specific safety concerns were addressed. The Intervention and Implementation sub-committee for each trial had representation from DCC scientists and was responsible for 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 the MsFLASH trials. The DSMB is an independent, multidisciplinary group consisting of a biostatistician, an epidemiologist, a nurse scientist, and clinicians that collectively have experience in management of symptoms related to the menopausal transition and in the conduct and monitoring of randomized clinical trials. The 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 Boards (IRB) of each participating site and all network committees before implementation.

All MsFLASH trials are registered at ClinicalTrials.gov (http://clinicaltrials.gov/): MsFLASH-01 Escitalopram for Menopausal Symptoms in Midlife Women (NCT 00894543); MsFLASH02 Interventions for Relief of Menopausal Symptoms: A 3-by-2 Factorial Design Examining Yoga, Exercise, and Omega-3 Supplementation, (NCT 01178892); MsFLASH 03 Comparative Efficacy of Low-Dose Estradiol and Venlafaxine XR for Treatment of Menopausal Symptoms (NCT 01418209).

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

LESSONS LEARNED

The MsFLASH trials have provided many opportunities for improving our methods for RCTs for menopause symptoms. Our lessons learned include:

1. Centralized recruitment can reduce costs, improve the ability to control rates of recruitment at each study site, and facilitate the careful monitoring of yields from each step in the recruitment process.

2. Front loading recruitment mailings, as opposed to ramping up, provides quick estimates for ongoing mailing numbers and assists in keeping recruitment on target or ahead of schedule.

3. Detailed phone screening prior to bringing women into the research clinic for any testing can be critical to minimizing costs, since many women will be either ineligible or unwilling after hearing more about the study requirements.

4. Requiring 4 or more VMS per day over a 3-week screening period is a high threshold of VMS for recruitment, particularly if it must be met over a 3-week period. Lowering the threshold to 2 VMS per day can improve recruitment and still provide sufficient distribution of VMS. The ultimate choice must balance the effect size one wants to detect, and the sample size one can afford, since starting at a higher threshold may allow a smaller sample size.

5. Regardless of the VMS threshold chosen, requiring 3 weeks of screening was important for reducing the placebo response. For example, in MsFLASH 01 the percent decrease in VMS from baseline to 8 week was 33%[1]. In other trials of interventions for VMS the placebo response has been as high as 46%[54].

6. Although counter-intuitive, continuing to rate daily VMS, vs. intermittent diaries, is preferable. It requires tremendous staff time to conduct intermittent ratings because subjects need to be reminded and contact each time there is a new start-up point. Many subjects 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, would decrease data entry costs, and date/time stamping would provide data about back-filling. While a wearable event marker could fill part of this need, an online/phone application for VMS measurement could allow data entry of variables such as intensity, bother, and concurrent activities (e.g. asleep, physically active, sitting, concurrent stresses). MsFLASH investigators did discuss 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 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 menopause transition by allowing a young age (e.g., as young as 40 years) may find that their study population is predominantly older women (in our case age 48–62). 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 the urine is unacceptable.

10. In considering a multifactorial design for behavioral interventions, weigh the potential cost savings related to a shared control group with the potential complexities. These may include limitations of access for women due to facility location, screening complexities because women must be eligible for all the interventions, and the challenge of explaining complex protocols to potential participants.

11. Following women after an intervention is discontinued can provide important data about the speed with which symptoms rebound. This information is clinically relevant for letting women know what to expect should they choose to use an intervention. Ongoing follow up can also provide information about the placebo effect. For example, in MsFLASH 01, after medication was discontinued at 8 weeks there was a VMS rebound among women in the esitalopram group, but not in the placebo group[1]. Women were not unblinded until week 12, and at 11 weeks VMS frequency in the two groups was identical. It is also possible to gain information about women's intention for continue treatment after they are unblinded.

SUMMARY

The first NIH funded menopause network (MsFLASH) has successfully established network operations - designing and conducting 3 randomized controlled trials that collectively studied 6 interventions in randomized, controlled clinical trials, over 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 menopause symptoms. The use of standardized methods to determine eligibility and assess VMS and related outcomes will greatly improve the ability to compare the effectiveness of various treatment modalities across trials and therefore enable women and their health care providers to make more informed choices when choosing treatments to relieve menopause 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.

Acknowledgments

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. [PubMed: 21245182]

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. [PubMed: 11323317]

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. [PubMed: 10981461]

4. McKinlay S, Jefferys M, Thompson B. An investigation of the age at menopause. J Biosoc Sci. 1972; 4:161–173. [PubMed: 5030375]

5. U.S. 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. [PubMed: 11770182]

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. [PubMed: 11731510]

8. Newton KM, Reed SD, LaCroix AZ, Grothaus LC, Ehrlich K, Guiltinan J. Treatment of vasomotor symptoms of menopause with black cohosh, multibotanicals, soy, hormone therapy, or placebo: a randomized trial. Ann Intern Med. 2006; 145:869–879. [PubMed: 17179056]

9. 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. [PubMed: 9279834]

10. 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. [PubMed: 11438246]

11. 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. [PubMed: 11696076]

12. Morin, C., et al. Insomnia: Psychological Assessment and Management. New York: Guilford Press; 1993.

13. 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. [PubMed: 19454639]

14. 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. [PubMed: 10086433]

15. Smith MT, Wegener ST. Measure of sleep. Arthritis & Rheumatism. 2003; 49:S184–S196.

16. 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. [PubMed: 2748771]

17. Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh Sleep Quality Index. J Psychosom Res. 1998; 45:5–13. [PubMed: 9720850]

18. 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. [PubMed: 21263078]

19. 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. [PubMed: 12627737]

20. 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. [PubMed: 12749557]

21. 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. [PubMed: 16585467]

22. 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. [PubMed: 14706945]

23. 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. [PubMed: 20937930]

24. 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]

25. 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. [PubMed: 4607278]

26. Lipman RS, Covi L, Shapiro AK. The Hopkins Symptom Checklist (HSCL)--factors derived from the HSCL-90. J Affect Disord. 1979; 1:9–24. [PubMed: 162184]

27. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav. 1983; 24:385–396. [PubMed: 6668417]

28. 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. [PubMed: 19418100]

29. 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. [PubMed: 20856144]

30. Hilditch JR, Lewis J, Peter A, et al. A menopause-specific quality of life questionnaire: development and psychometric properties. Maturitas. 1996; 24:161–175. [PubMed: 8844630]

31. 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. [PubMed: 11738160]

32. 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. [PubMed: 6646795]

33. 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. [PubMed: 9610214]

34. 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. [PubMed: 17227907]

35. 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. [PubMed: 17048223]

36. 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. [PubMed: 10782451]

37. 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. [PubMed: 18042215]

38. 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. [PubMed: 15668596]

39. 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. [PubMed: 12100799]

40. 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. [PubMed: 11896107]

41. Loprinzi L, Barton DL, Sloan JA, et al. Pilot evaluation of gabapentin for treating hot flashes. Mayo Clin Proc. 2002; 77:1159–1163. [PubMed: 12440550]

42. 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. [PubMed: 21900849]

43. Bootzin, RR. The role of expectancy in behavior change. In: White, L.; Turskey, B.; Schwartz, G., editors. Placebo: Theory, research and mechanisms. New York: The Guilford Press; 1985. p. 196-210.

44. Gross D. On the merits of attention-control groups. Res Nurs Health. 2005; 28:93–94. [PubMed: 15779061]

45. Srinivasan S, Liu C, Mitchell CM, et al. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. PLoS ONE. 2010; 5:e10197. [PubMed: 20419168]

46. Carpenter JS, Andrykowski MA, Freedman RR, Munn R. Feasibility and psychometrics of an ambulatory hot flash monitoring device. Menopause. 1999; 6:209–215. [PubMed: 10486790]

47. 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. [PubMed: 15572497]

48. Thurston RC, Blumenthal JA, Babyak MA, Sherwood A. Emotional antecedents of hot flashes during daily life. Psychosom Med. 2005; 67:137–146. [PubMed: 15673636]

49. 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

50. Anderson GL, Prentice RL. Individually randomized intervention trials for disease prevention and control. Stat Methods Med Res. 1999; 8:287–309. [PubMed: 10730335]

51. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics. 1975; 31:103–115. [PubMed: 1100130]

52. Friedman, LM.; Furberg, CD.; DeMets, DL., et al. Fundamentals of Clinical Trials. 3rd ed.. New York: Springer-Verlag; 1998.

53. 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. [PubMed: 11459516]

54. Sood R, Sood A, Wolf SL, et al. Paced breathing compared with usual breathing for hot flashes. Menopause. 2013; 20:179–184. [PubMed: 22990758]

55. Avis NE, Kaufert PA, Lock M, McKinlay SM, Vass K. The evolution of menopausal symptoms. Baillieres Clin Endocrinol Metab. 1993; 7:17–32. [PubMed: 8435051]

56. 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. [PubMed: 17688760]

57. 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. [PubMed: 17874169]

58. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore. 1994; 23:129–138. [PubMed: 8080219]

59. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. J Pain. 2004; 5:133–137. [PubMed: 15042521]

60. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharmacol Bull. 1993; 29:321–326. [PubMed: 8290681]
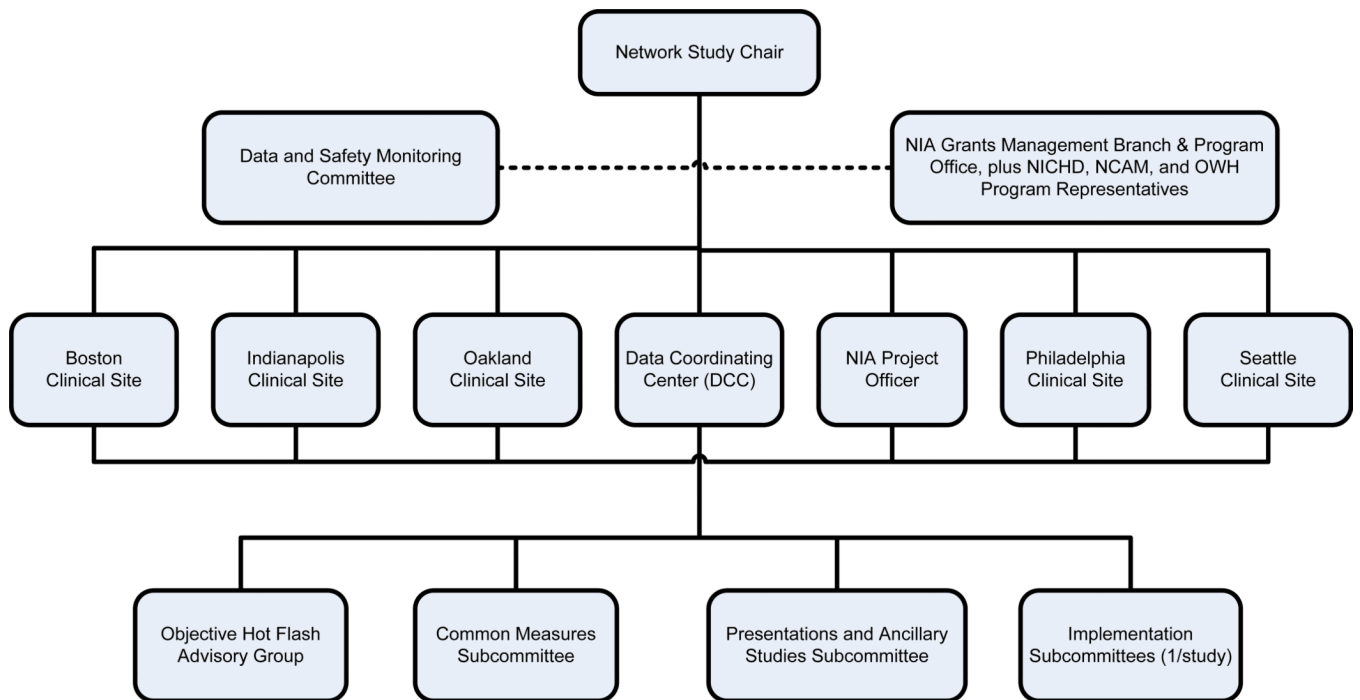
FIG. 1. Menopausal Strategies: Finding Lasting Answers to Symptoms and Health leadership structure
NIA, National Institute on Aging; NICHD, Eunice Kennedy Shriver National Institute of Child Health and Human Development; NCCAM, National Center for Complementary and Alternative Medicine; OWH, Office of Research on Women's Health.
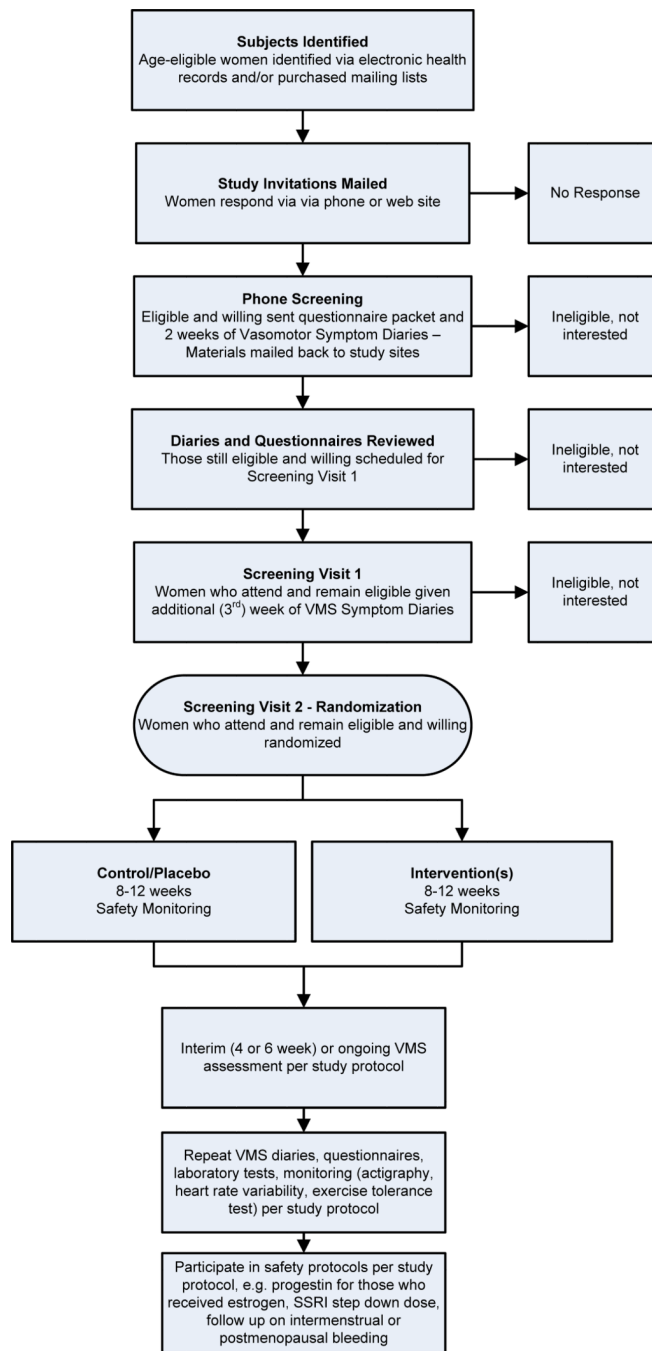
FIG. 2. Menopausal Strategies: Finding Lasting Answers to Symptoms and Health recruitment and follow-up
VMS, vasomotor symptoms; SSRI, selective serotonin reuptake inhibitor.

Table 1

MsFLASH Eligibility Criteria Applied to All MsFLASH trials (exceptions noted in parentheses)

Menopausal / Hot Flash

1. At least 28 VMS / week as recorded on daily VMS diary for each of 3 weeks (MsFLASH 01 1) (Changed to 14 VMS per week in MsFLASH 02, maintained in MsFLASH 03). During a 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 weeks.

2. At least 4 reports / week of days or nights with moderate to severe VMS severity and or bother for each of 3 weeks

3. In late menopausal transition, not longer than 5 years post-menopausal Age 40–62 years at screening

4. No reported use of systemic hormones (contraceptives or postmenopausal hormones) in the past 2 months (levonorgestrel intrauterine system and vaginal estrogen local therapy < 4 times/week allowed)

5. Agrees not to use hormones outside of trials for duration of trial

6. No reported use of any non-hormonal hot flash therapy in the past month

7. Hysterectomized woman w/ serum FSH >20 mIU/mL and estradiol <50 pg/mL at baseline (on at least one of two blood draws two weeks apart).

8. Using the Mirena IUD or had an endometrial ablation and still have one or both ovaries, with FSH level > 20 mIU/mL and estradiol 50 pg/mL (on at least one of two blood draws two weeks apart)

Medical History

1. In general good health as determined by medical history

2. No severe or unstable medical illness

3. No uncontrolled hypertension (>160/100)

4. Resting heart rate 110

5. Not pregnant, intending pregnancy, or breast feeding

6. No abnormal mammogram in prior 2 years (MsFLASH 03)

7. Body mass index 37, based on measured height and weight

8. No major severe depressive episode in the past 3 months

9. No suicide attempt in last 3 years (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. No psychotropic medications use (SSRIs, SNRIs, MAOIs) or other antidepressants and anxiolytics, within the past 30 days (MsFLASH 01 and 03)

Study Logistics

1. Provide written informed consent

2. No participation in another menopause trial

3. No hypersensitivity or contraindication to study medications / treatment (varies by study)

4. Willing and able to complete study procedures

Table 2

MsFLASH Common Measures and Outcomes

Outcome	Description	Measure Selected	Measurement Details	# Items	When Collected
Vasomotor symptoms	Subjective VMS frequency	Diary – Twice daily (day/night) estimate of # 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 Avis 1993	Diary - Twice daily rating using SWAN response categories (not at all, a little, moderately, a lot)[55]	Self-reported paper diary	2 per day	Daily
	Subjective VMS interference	Hot Flash Related Daily Interference Scale (HFRDIS)[31]	Self-reported	10	Baseline Follow-up
Sleep	Sleep quality & disturbance	Pittsburgh Sleep Quality Index (PSQI)[15]	Self-reported	18	Baseline Follow-up
	Insomnia symptoms	Insomnia Severity Index (ISI)[15]	Self-reported	7	Baseline Follow-up
	Sleep and Wake and Nap Times Actiwatch usage	Diary - Twice daily (day/night) and Device	Self-reported	2+/day	Baseline Follow-up Daily
	Objective sleep quality	Actiwatch real time recording for 3 days	Actigraphic monitoring	0	Baseline Follow-up
Mood state/depression/anxiety/stress	Depressive symptoms	Patient Health Questionnaire (PHQ-9 or PHQ-8)[23,56,57]	Self-reported	8–9	Baseline Follow-up
	Anxiety symptoms	Generalized Anxiety Disorder (GAD-7)[23] Hopkins Symptom Checklist (HSCL)[24-26]	Self-reported	7 10	Baseline Baseline Follow-up
	Perceived stress	Perceived Stress Scale (PSS)[27]	Self-reported	10	Baseline Follow-up
Physical pain	Pain severity and interference	PEG (from the Brief Pain Inventory)[28,58,59]	Self-reported	3	Baseline Follow-up
Quality of Life – Menopause Specific	The presence and bother associated with menopausal symptoms	Menopause Specific Quality of Life (MENQOL)[30]	Self-reported	29	Baseline Follow-up
Quality of Life - Overall	Importance and satisfaction with life	Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q)[60]	Self-reported	15	Baseline Follow-up
Sexual function/vaginal dryness	Key dimensions of female sexual function	Female Sexual Function Index (FSFI) Rosen[36]	Self-reported	18	Baseline Follow-up
Sexual Distress	Distinguish between sexual dysfunction & no sexual dysfunction, sexually related personal distress in women with hypoactive sexual desire disorder (HSDD)	Female Sexual Distress Scale (FSDS-R)[37]	Self-reported	1	Baseline Follow-up

VMS =vasomotor symptoms

Table 3

Comparisons and sample sizes for MsFLASH clinical trials

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

[*] Measured as difference between groups in VMS frequency per day change from baseline to follow-up, assuming a standard deviation of 4.0 VMS per day. Sample size calculations were based on t-tests with power of 90% and 2-sided alpha = 0.025.

Table 4

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

Correlation[1]	Post-treatment[2]	Percent change from baseline[3]	50% change from baseline[4]
0.3	0.91	0.81	0.72
0.5	0.95	0.92	0.81
0.7	0.99	0.99	0.91

VMS change from baseline to follow-up values assumed to follow a normal distribution; sample size of 90 per treatment group; analysis includes one follow-up time point measure.

[1] correlation between baseline and post-treatment measurements

[2] models post-treatment outcome as a continuous outcome, adjusted for the baseline measure

[3] models percent change from baseline as a continuous outcome

[4] models proportion with at least a 50% decrease in VMS symptoms