



OPEN

## Open-label placebos for menopausal hot flushes: a randomized controlled trial

Yiqi Pan<sup>1,2</sup>✉, Ramona Meister<sup>3</sup>, Bernd Löwe<sup>2</sup>, Ted J. Kaptchuk<sup>4</sup>, Kai J. Buhling<sup>5</sup> & Yvonne Nestoriuc<sup>1,2</sup>

This study investigated the efficacy of an open-label placebo (OLP) treatment for menopausal hot flushes. Women with at least five moderate or severe hot flushes per day were allocated to receive four weeks of OLP for twice a day or no-treatment. Intention-to-treat analyses included  $n = 100$  women. In comparison to no-treatment, OLP reduced the log-transformed hot flush composite score (frequency  $\times$  intensity) (mean difference in change:  $-0.32$ , 95% CI  $[-0.43; -0.21]$ ,  $p < 0.001$ , Cohen's  $d = 0.86$ ), hot flush frequency ( $-1.12$  [ $-1.81; -0.43$ ],  $p = 0.02$ , Cohen's  $d = 0.51$ ), and improved overall menopause-related quality of life ( $-2.53$  [ $-4.17; -0.89$ ],  $p = 0.02$ , Cohen's  $d = 0.49$ ). Twelve (24%) (vs. three [6%]) patients had 50% lesser hot flushes. Problem rating of hot flushes and subdomains of quality of life did not improve. After four weeks, the OLP group was further divided via randomization to continue or discontinue the treatment. Benefits were maintained at week 8 (log-transformed score:  $-0.04$  [ $-0.06; 0.14$ ],  $p = 0.45$ ). There was no difference between taking placebos for 8 or 4 weeks (log-transformed score:  $0.04$  [ $-0.17; 0.25$ ],  $p = 0.73$ ). Results indicate that open-label placebos may be an effective, safe alternative for menopausal hot flushes.

Placebos are commonly used in clinical practice deceptively<sup>1,2</sup>. In a systematic review, Fässler and colleagues reported that 17–80% of clinicians administer placebos in such an ethically problematic manner<sup>1</sup>.

A growing body of studies demonstrated that placebos without deception can have beneficial effects. Prescribing placebos in an honest manner would eliminate the ethically problematic use of deceptive placebos or active medications that have no efficacy for the treated condition (“impure placebos”) in the clinical setting<sup>2</sup>. The first randomized controlled trial (RCT) on open-label placebo (OLP) was conducted in 2010 among patients with irritable bowel syndrome<sup>3</sup>, followed by further RCTs on chronic back pain<sup>4,5</sup>, depression<sup>6</sup>, allergic rhinitis<sup>7</sup>, and cancer-related fatigue<sup>8</sup>. A meta-analysis from 2017 found an overall positive effect of 0.9, indicating a large effect of OLP compared to no-treatment<sup>9</sup>.

Currently, these studies do not allow definite clinical conclusions because of their small sample size. Nonetheless, the end goal of OLP research is to improve patient care, especially for patients whose treatment options are limited, and when OLP has the potential to fill in the gaps in care.

One such patient group is women suffering from hot flushes. Hot flushes are the most commonly reported symptoms in menopause, occurring in up to 80% of all women and affecting quality of life to various degrees<sup>10,11</sup>. Women who experience hot flushes at night (also called night sweats) additionally suffer from lack of sleep, which in turn negatively impacts their functioning during the day. For decades, hot flushes have been treated successfully with hormone substitute therapy<sup>12,13</sup>. However, the Women's Health Initiative (WHI) trial from the year 2002, investigating over 16,000 women taking hormone therapy, documented increased risks of adverse events, including breast cancer and stroke<sup>14</sup>. Since then, many clinicians and patients are seeking for non-hormonal treatment options. Although almost two decades have passed, the topicality of the WHI study remains as the association between breast cancer risk and hormone therapy has been recently validated in a meta-analysis using epidemiological data<sup>15</sup>.

<sup>1</sup>Department of Clinical Psychology, Helmut-Schmidt-University/University of the Federal Armed Forces, Hamburg, Germany. <sup>2</sup>Department of Psychosomatic Medicine and Psychotherapy, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany. <sup>3</sup>Department of Medical Psychology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany. <sup>4</sup>Program in Placebo Studies and the Therapeutic Encounter (PiPS), Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA. <sup>5</sup>Department of Gynecological Endocrinology, Clinic of Gynecology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany. ✉email: y.pan@uke.de

To treat hot flushes, patients may choose from a large variety of complementary and alternative medicines, e.g., black cohosh, phytoestrogens, wild yam, Chinese herbal medicine, etc. Herbal-based remedies were researched extensively with meta-analyses or systematic reviews demonstrating mixed or no effects over placebo<sup>16–18</sup>. Likewise, mind–body interventions, including yoga, acupuncture, paced breathing, and relaxation appear ineffective in reducing hot flushes<sup>13,19</sup>. Selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs), gabapentin, and clonidine are effective yet may cause side effects or withdrawal symptoms<sup>20</sup>. Given their safety profile, the benefits of these medications remain debatable<sup>21,22</sup>. Early studies have shown that cognitive-behavioural therapy and hypnosis may help<sup>23,24</sup>. However, the body of evidence is small.

Importantly, placebo effects in double-blind clinical trials are substantial. The mixed effectiveness of multiple remedies and large placebo effects across studies indicate that non-specific effects and related mechanisms such as expectations may play an important role<sup>25</sup>. In placebo arms, hot flushes were reduced by an average of 58% in hormone therapy trials<sup>26</sup>, and by 20–50% in non-hormonal trials<sup>20</sup>. Using pooled data from two studies, Freeman found that 33% of the women report half the hot flushes they had at baseline after taking eight weeks of placebos, indicating clinically meaningful improvements<sup>25</sup>. The magnitude of the placebo response may thus explain the inefficacy of some active treatments<sup>27</sup>. However, placebo effects in RCTs are confounded by natural history and statistical artefacts like regression to the mean since untreated groups are rarely included<sup>28</sup>. A no-treatment control can be important to separate placebo effects from these other confounders.

In this study, we investigate the efficacy of placebos without deception for menopausal hot flushes by comparing the OLP group with a no-treatment group. To further our understanding of the duration of OLP effects and to explore potential exposure–response relationships, we compare 4 and 8 weeks of placebo intake. Since its underlying mechanisms is unclear, we will additionally examine the link between expectations, hope, and optimism and OLP response. Whereas expectations are one of the most robust predictors of placebo response in laboratory studies<sup>29</sup>, optimism, and hope were additionally proposed as pertinent factors to OLP<sup>30</sup>, thus warranting scrutiny.

## Results

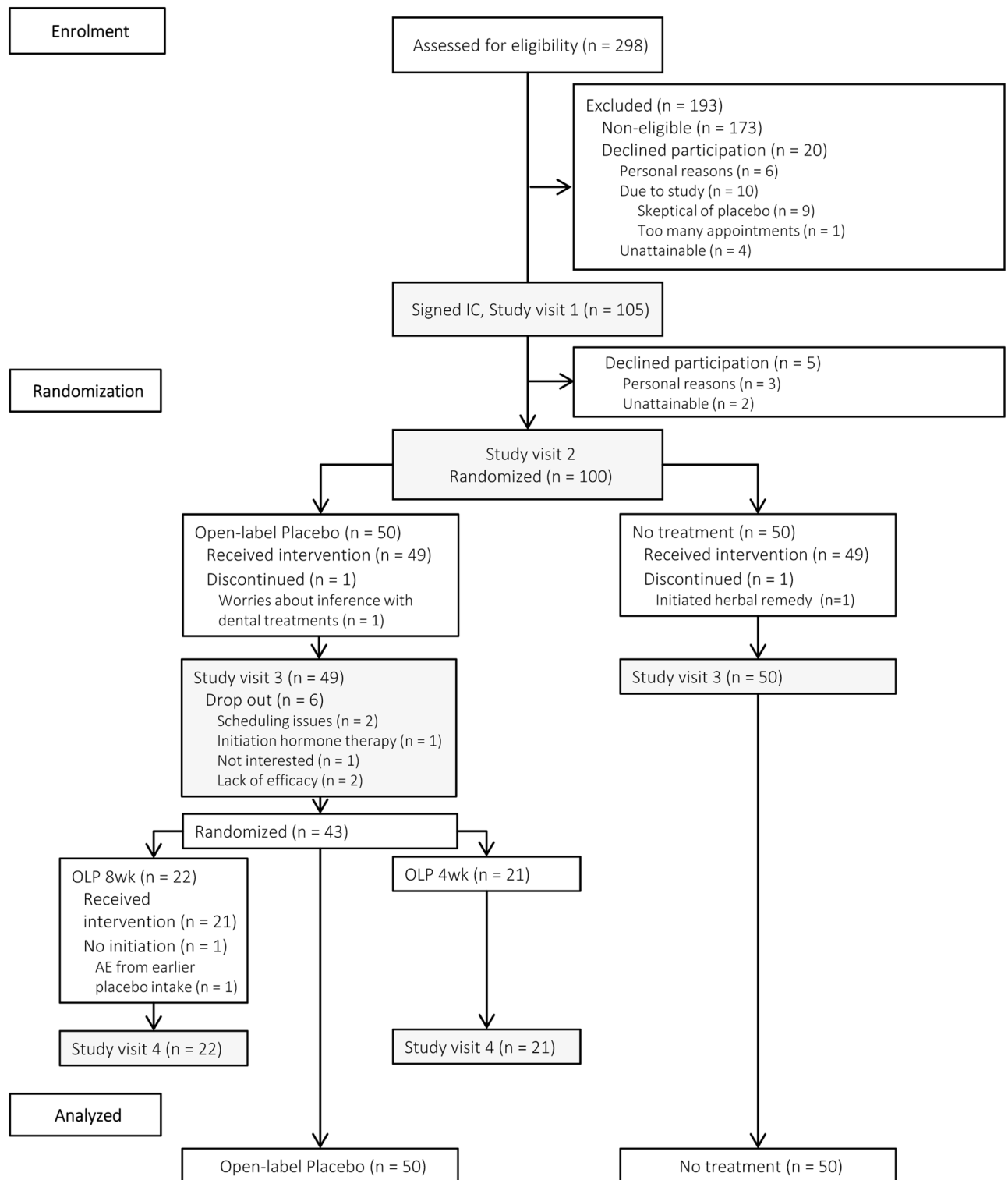
**Participant flow.** We randomized and analysed  $n = 100$  participants (Fig. 1). One group knowingly received 4 weeks of placebos (OLP), whereas the other group received no treatment. Two patients (one in each group) dropped out (2%) during the treatment phase. After 4 weeks of treatment, six patients in the OLP group refused to participate in the follow-up phase, two of which did so due to unrelieved symptoms. One patient rejected further placebo intake yet completed diary and questionnaires at study visit 4. As a result,  $n = 22$  and  $n = 21$  patients were randomized to the OLP 8 week and OLP 4 week group.

**Recruitment.** Participants were recruited in and around Hamburg, Germany, from the general population. We recruited from October 2018 to December 2019 and followed up from November 2018 to March 2020. Women were informed via leaflets at their practitioner (gynaecologists, general practitioners, practitioners of alternative medicine), pharmacies, stores, cultural institutions, and gyms. We also used online and newspaper advertising, and the internal newsletter of the University Medical Centre (UMC) Hamburg. This study was advertised as a “novel holistic treatment” (short leaflet) or as “open-label placebo treatment for hot flushes” (detailed leaflet, study website). Women contacted the study team via phone or email. The screening interview was standardized and included a stringent decision tree based on the eligibility criteria. Study assistants conducted it over the phone. If eligible, a first study visit was scheduled. All patients were explicitly informed about the study’s objective of testing honestly prescribed placebos with informed consent.

**Baseline characteristics.** Patient characteristics are shown in Table 1. Most patients (76%) were post-menopausal, in a partnership (52%), and employed (75%). Time since onset of hot flushes varied between 30 days and 26 years ( $M = 5.2$  years,  $SD = 5.4$ ). Six patients had a history of breast cancer, of which four completed or discontinued endocrine therapy for at least one year. The majority had used some kind of hot flush treatment before (75%). No woman utilized body-mind interventions (e.g., yoga, meditation, hypnosis) for hot flushes. The number of daily hot flushes varied between 3.1 and 24.4. Baseline characteristics did not differ between the groups.

**Missing values.** Single missing values in questionnaires and the diary were indicated by 25 and 18 patients, respectively. Missing values (per patient) ranged from 0.4 to 1.6% in questionnaires and from 0.7 to 5% in the diary. Except for one person who overlooked one page (6.7% missing values; not outcome-related), missing items were scattered across the questionnaires. At least 80% of the items were present for all scales. Missing data due to dropouts were very low ( $n = 1$ ; 1%).

**Efficacy analyses.** The OLP group reported a hot flush score of  $M = 16.74$  ( $SD = 9.68$ ) at baseline, and of  $M = 10.72$  ( $SD = 9.73$ ) at week 4. The no-treatment group reported a score of  $M = 18.41$  ( $SD = 9.53$ ) at baseline, and of  $M = 15.15$  ( $SD = 7.78$ ) at week 4 (Fig. 2). In comparison to no-treatment, OLP reduced the log-transformed hot flush score and hot flush frequency with effect sizes of Cohen’s  $d = 0.86$  and  $d = 0.51$ , respectively, indicating large and moderate effects (Table 2). Back-transformed estimates of the hot flush score reduced by 6.65 points (43%) (vs. 2.99 points [19%]) in the OLP group. The onset of action appeared to be rapid, given that 40% was reduced after three weeks of intake. By week 4, estimates indicated that the OLP group had 2.90 lesser hot flushes (33%) (vs. 1.77 [20%]). Also, 31 (vs. 20) women reported to have overall improved ( $\chi^2(1, N = 50) = 4.82$ ,  $p = 0.03$ , Cramer’s  $V = 0.22$ ). Twelve women (vs. 3) obtained a clinically relevant reduction by 50% in hot flush frequency (Fisher’s Exact Test:  $p = 0.02$ , Cramer’s  $V = 0.25$ ) (Fig. 2). One case in the OLP group was identified as a multivariate outlier as she had a baseline score and frequency of 4.1 and 4.0 SD above the sample mean. After



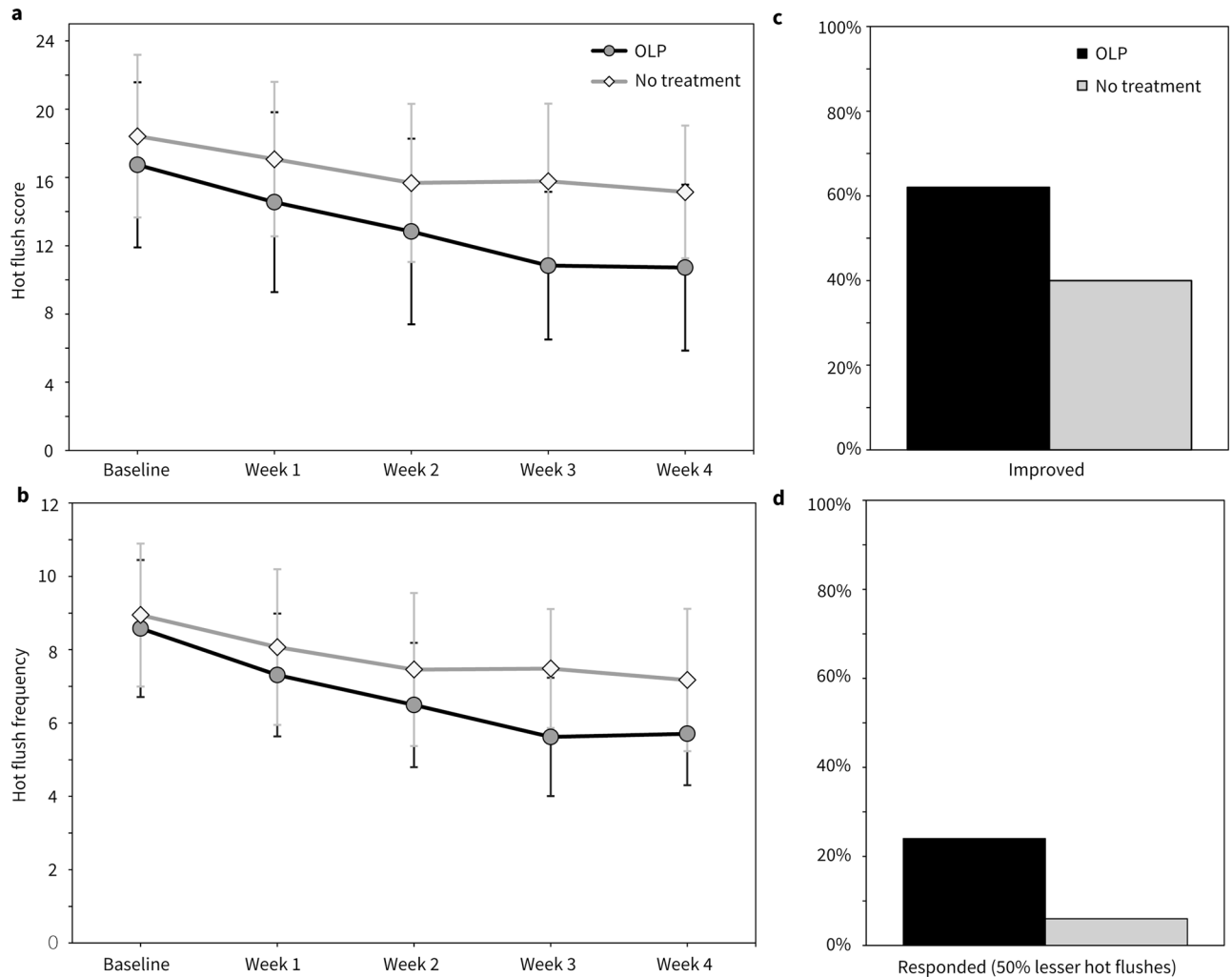
**Figure 1.** Flow of participants. *IC* informed consent, *OLP* open-label placebo, *wk* week, *AE* adverse event.

deleting this case, the results were identical to ITT. However, the effects were marginally larger (log-transformed hot flush score:  $-0.33$ , 95% CI  $[-0.45; -0.21]$ ,  $p < 0.001$ , Cohen's  $d = 0.89$ ; frequency:  $-1.19$   $[-1.87; -0.51]$ ,  $p = 0.006$ , Cohen's  $d = 0.55$ ).

Problem rating, i.e., the perceived burden by hot flushes, decreased within both groups (OLP:  $-1.86$ , 95% CI  $[-2.38; -1.34]$ ,  $p < 0.001$ ; No-treatment:  $-1.43$   $[-1.94; -0.91]$ ,  $p < 0.001$ ), yet did not differ *between* the groups ( $-0.43$   $[-1.01; 0.15]$ ,  $p = 0.24$ , Cohen's  $d = 0.23$ ). We found a significant group difference of moderate size (Cohen's  $d = 0.51$ ) for overall menopause-related quality of life (MRS-II). However, there were no differences

	OLP (n = 50)	No-treatment (n = 50)
Age (M, range)	54.2 (44-76)	54.9 (46-71)
<b>Marital status (N)</b>		
In a partnership	27	25
Single	17	15
Divorced	6	7
Widowed	0	3
<b>Highest education (N)</b>		
Middle school	22	17
High school	15	15
University	13	18
<b>Employment status (N)</b>		
Employed	35	40
Unemployed/home maker	8	6
Retired	7	4
<b>Menopausal status (N)<sup>a</sup></b>		
Transition period (interval of amenorrhea $\geq$ 60 days)	14	10
Early post menopause (2-6 years since FMP)	33	35
Late post menopause (remaining life span)	3	5
<b>Years since onset of hot flush (N)</b>		
< 1 year	7	7
1-2 years	6	8
2-4 years	15	17
4-6 years	6	7
> 6 years	16	11
<b>Gynaecological history (N)</b>		
History of breast cancer	1	5
Hysterectomy	7	8
<b>Numbers of previous hot flush treatment (N)</b>		
None	16	9
1-3	30	36
4 or more	4	5
<b>Class of previous hot flush treatment<sup>b</sup> (N)</b>		
Black cohosh	18	20
Hormone therapy	5	10
Homeopathic remedy	8	7
Sage	5	8
Siberian rhubarb roots	4	3
Other <sup>c</sup>	13	21
<b>Body Mass Index, categorized (N)<sup>d</sup></b>		
Underweight	4	1
Normal weight	19	29
Pre-obesity	14	15
Obese	13	5
<b>Smoking status</b>		
Never	23	18
Current	11	14
Former smoker	16	18
<b>Physical activity</b>		
Never	4	4
1-3 days per month	12	7
1-2 days per week	24	23
3+ days per week	10	16
<b>Number of hot flushes per day<sup>e</sup></b>		
3-5	11	10
6-7	10	7
8-10	17	17
11+	12	16

**Table 1.** Patient characteristics. *OLP* open-label placebo, *FMP* final menstrual period. <sup>a</sup>Classified after the Stages of Reproductive Aging Workshop (STRAW) criteria. <sup>b</sup>Numbers do not sum to 100 since multiple entries were possible. <sup>c</sup>Other treatments include bioidentical hormones, chaste berry, citric acid, dandelion, devil's claw, evening primrose, hemp seeds, Iceland moos, lady's mantle, linseeds, pollen extract, pomegranate, red clover, soy, St. John's wort, and wild yam. <sup>d</sup>Classified after the World Health Organization. <sup>e</sup>Based on the diary. Eligibility criteria required patients to have at least five hot flushes a day yet mean hot flushes during the baseline week might have differed from reports at the screening.



**Figure 2.** Changes in hot flushes and overall improvement (observed values). Changes in (a) hot flush score, and (b) hot flush frequency over time. (c) Proportion of patients who indicated improvement, and (d) proportion of patients with clinically meaningful improvements. Error bars indicate standard deviations. OLP Open-label placebo.

in the menopause-related quality of life subdomains (WHQ). Only for the subdomain somatic symptoms, the groups differed in trend (4.92 [0.05; 9.98],  $p=0.05$ , Cohen's  $d=0.39$ ). Since overall and specific quality of life scales demonstrated different results, we deleted the "vasomotor symptoms" item from the overall scale for further inspection. We found the same results as with the original scale ( $-2.47$  [ $-3.99$ ;  $-0.95$ ],  $p=0.01$ , Cohen's  $d=0.51$ ).

As some scenarios of the pre-planned sensitivity analyses did not apply to this sample, we omitted three of the four sensitivity analyses. Eleven patients (OLP: 5, no-treatment: 6) made lifestyle changes during the treatment yet removing them did not modify the results (log-transformed hot flush score:  $-0.32$  [ $-0.44$ ;  $-0.19$ ],  $p<0.001$ , Cohen's  $d=0.86$ ). Changes in lifestyle were non-substantial, i.e., women were less physically active or drank less alcohol in comparison to baseline. We skipped re-running the analyses after imputing missing values of dropped cases, as there was only one patient who did not complete assessments at treatment end. This low percentage of missing data is unlikely to create biases in estimates of linear mixed models<sup>31</sup>. We also skipped re-running the analyses after deleting non-adherers since all patients adhered to the treatment. Another sensitivity analysis would include stress and years since onset of hot flushes as covariates given its role as potential confounders. However, in this sample, neither stress ( $r=0.11$ ,  $p=0.30$ ), nor years since onset of hot flushes ( $r=-0.04$ ,  $p=0.71$ ) correlated with the change of log-transformed hot flush score, rendering further analyses futile. We omitted sensitivity analyses with problem rating, as there was no treatment effect.

**Duration and time course of OLP effects.** Within the OLP group, the log-transformed hot flush score did not differ between week 4 and week 8 ( $-0.04$  [ $-0.06$ ;  $0.14$ ],  $p=0.45$ ). This pattern was also found within OLP 4 week ( $n=21$ ), for which benefits remained at week 8 ( $-0.03$  [ $-0.11$ ;  $0.17$ ],  $p=0.69$ ). Correspondingly, problem rating did not differ between week 4 and week 8, neither in the OLP group ( $-0.001$  [ $-0.46$ ;  $0.46$ ],  $p=0.99$ ), nor in the OLP 4 week subgroup ( $0.05$  [ $-0.59$ ;  $0.69$ ],  $p=0.89$ ). Furthermore, there were no differences between the OLP 8 week and OLP 4 week at week 8 (log-transformed hot flush score:  $0.04$  [ $-0.17$ ;  $0.25$ ],  $p=0.73$ ; problem

Outcomes, metric	Baseline (observed)				Mean change from baseline to treatment end (week 4) (estimates)				
	OLP (N = 50)		No-treatment (N = 50)		OLP	No-treatment	Group difference [95% CI]	p	Cohen's d
	M	SD	M	SD	M	M			
Log-transformed hot flush score (diary) <sup>a</sup>	2.75	0.49	2.85	0.51	-0.52	-0.20	-0.32 [-0.43; -0.21]	<0.001***	0.86
Problem rating (HFRS) <sup>b</sup>	6.45	1.24	7.12	1.20	-1.86	-1.43	-0.43 [-1.01; 0.15]	0.24	0.23
Hot flush frequency (Diary)	8.58	3.90	8.95	3.73	-2.90	-1.77	-1.12 [-1.81; -0.43]	0.02*	0.51
QoL overall (MRS-II) <sup>c</sup>	18.06	9.25	18.86	7.87	-1.75	0.78	-2.53 [-4.17; -0.89]	0.02*	0.49
QoL Anxiety and Depression (WHQ) <sup>d</sup>	75.67	21.59	75.43	16.22	2.22	0.86	1.37 [-1.82; 4.56]	0.50	0.14
QoL Well-being (WHQ)	80.78	16.48	80.33	14.16	-2.01	-1.00	-1.01 [-4.73; 2.71]	0.67	0.09
QoL Somatic symptoms (WHQ)	69.47	20.39	69.60	15.60	3.18	-1.73	4.92 [0.05; 9.98]	0.05	0.39
QoL Memory/concentration (WHQ)	55.56	29.78	50.67	24.81	4.70	1.78	2.92 [-1.63; 7.49]	0.31	0.20
QoL Sleep problems (WHQ)	42.33	26.77	43.33	30.12	5.38	1.33	4.05 [-2.00; 10.09]	0.29	0.20

**Table 2.** Change from baseline to week 4 in hot flush symptoms and quality of life. Baseline values were obtained from the observed means and standard deviation. The mean change from baseline to week 4 was obtained from estimates of linear mixed models. *OLP* Open-label placebo, *HFRS* Hot Flush Rating Scale, *QoL* Quality-of-Life, *MRS-II* Menopause Rating Scale II, *WHQ* Women's Health Questionnaire, *PGIC* Patient Global Impression of Change. \* $p < 0.05$ , \*\*\* $p < 0.001$ . <sup>a</sup>Hot flush score is a composite scale of frequency  $\times$  severity (1 = mild, 2 = moderate, 3 = severe). <sup>b</sup>Scale range: 1–10. <sup>c</sup>Scale range: 0–55. Higher scores indicate severe menopausal symptoms and great impact of such symptoms on quality of life. <sup>d</sup>Scale range of all WHQ subdomains: 0–100. Higher scores indicate better health.

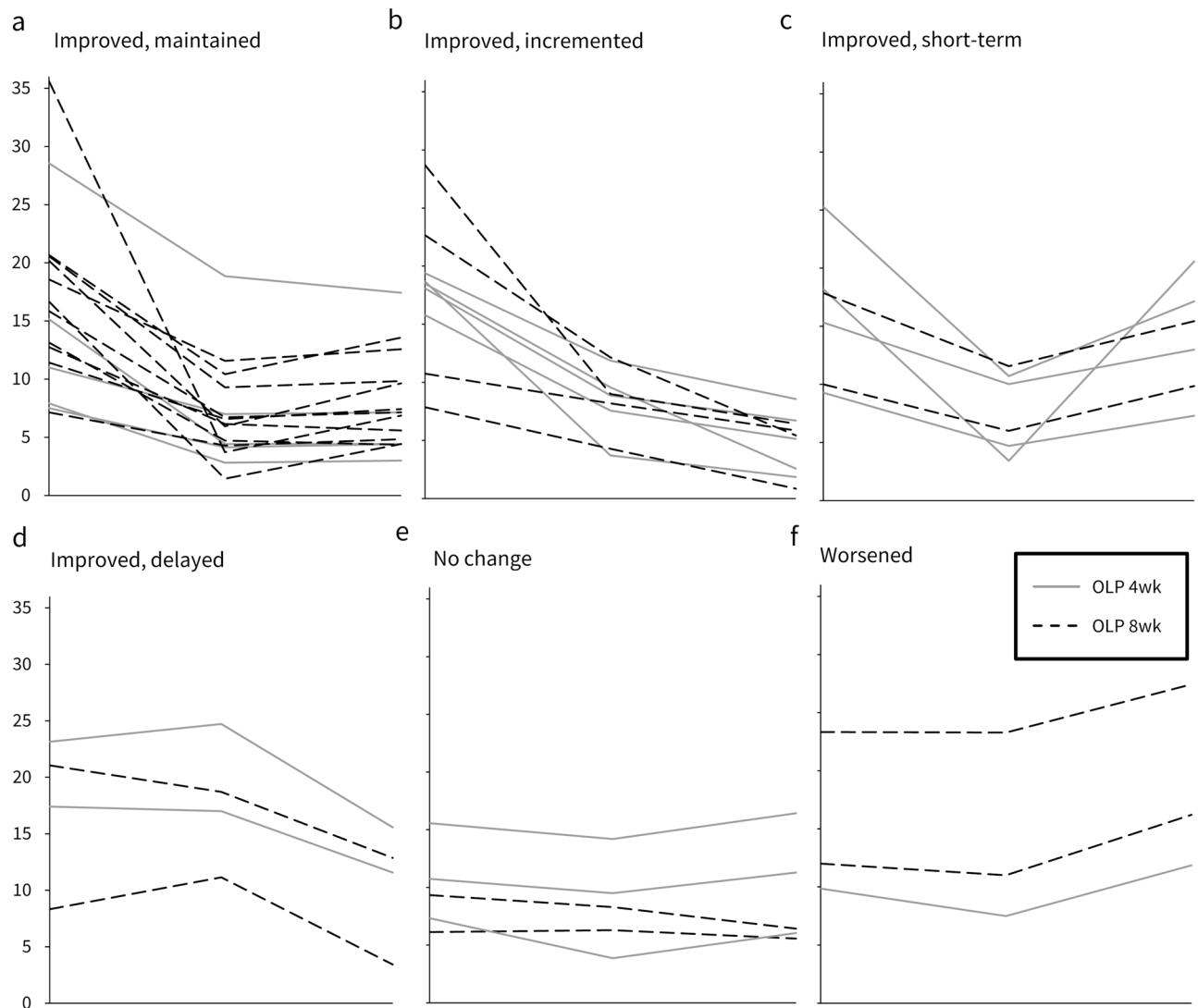
rating:  $-0.07 [-1.17; 1.08]$ ,  $p = 0.90$ ). Taken together, these results indicate that benefits remained after intake was stopped. Taking the placebo for 8 weeks was not superior to taking the placebo for 4 weeks. Measures of distribution of all outcomes in the follow-up period are listed in the Supplementary Table S1.

When examining the individual trajectories of the hot flush score over time, we found six different patterns (Fig. 3): Improved, maintained (37%), Improved, incremented (21%), Improved, short-term (14%), Improved, delayed (9%), No change (12%), and Worsened (7%). Descriptive statistics displayed no distinct group differences considering the symptom course. In particular, of OLP 8 week and OLP 4 week, 16 (73%) and 13 (62%) women reported maintained, incremented, or delayed improvements. Summarized, visualizations of the symptom course validated the statistical analyses.

**Role of expectation, hope, and optimism.** On average, hope was higher than expectations at all time points (Table 3). Expectations at baseline were right-skewed, i.e., 24% patients expected 'no change' after they were randomized. In contrast, hope was left-skewed, i.e., 33% wished for maximum improvement. Table 3 shows that the treatment group predicted expectations after the first randomization ( $p = 0.005$ ), indicating that patients expected more improvement when allocated to OLP. After the second randomization, expectations did not differ between OLP 8 week and OLP 4 week ( $p = 0.48$ ). This indicated that treatment continuation or discontinuation did not affect expecting improvements. Hope was affected by neither of the two randomizations.

Although the randomization influenced expectations, expectations had no effect on hot flushes. For the complete sample, we found that the baseline to week 4 change in the log-transformed hot flush score did not correlate with expectations at baseline ( $r = -0.17$ ,  $p = 0.09$ ), nor with hope ( $r = -0.03$ ,  $p = 0.78$ ), or optimism ( $r = -0.06$ ,  $p = 0.55$ ) at baseline. We found the same results when correlating expectations with both the non-transformed hot flush score and frequency at week 4. Among these variables, hope correlated significantly with expectations ( $r = 0.39$ ,  $p < 0.001$ ) and optimism ( $r = 0.25$ ,  $p < 0.01$ ). No associations were found between expectations and optimism ( $r = 0.09$ ,  $p = 0.40$ ). As indicated by fixed effects of linear mixed models, there was no main effect of expectations on the log-transformed hot flush score ( $F(1, 94) = 2.30$ ,  $p = 0.13$ ). Also, there was no interaction effect of Expectations  $\times$  Group  $\times$  Time on the log-transformed hot flush score ( $F(8, 381) = 0.97$ ,  $p = 0.46$ ). Correspondingly, there was no main effect ( $F(1, 94) = 1.11$ ,  $p = 0.30$ ), or interaction effect ( $F(8, 380) = 1.01$ ,  $p = 0.43$ ) of hope. We found the same pattern for optimism (Main:  $F(1, 94) = 0.59$ ,  $p = 0.45$ ; Interaction:  $F(8, 381) = 0.68$ ,  $p = 0.71$ ). As there was no treatment effect of problem rating, analyses to assess potential moderation effects by expectations, hope, and optimism were omitted.

**Adherence.** At week 4, 33 (66%) and seventeen patients (34%) reported having taken 'all' and 'almost all' pills, respectively. No patients took less than 'almost all' pills. Therefore, all patients were adherent according to our criterion. Of those who took 'almost all' pills, six additionally specified that they had forgotten one or two pills. Others did not give further information. At week 8, eighteen (82%) and four patients (18%) indicated to have taken 'all' and 'almost all' pills, respectively. Adherence at week 8 was reported by the OLP 8 week group only.



**Figure 3.** Individual trajectories of hot flush score from baseline, week 4, to week 8 (observed values). Each line represents one case. *OLP 4wk* OLP intake from baseline to week 4, *OLP 8wk* OLP intake from baseline to week 8.

**Adverse events.** We examined whether there were placebo-related adverse events. As shown in Table 4, the no-treatment group reported *more* symptoms at treatment end than the OLP group, indicating that these symptoms may be everyday life complaints. It should be mentioned that some of these symptoms—which are frequent adverse events—might also be related to menopause, including sleeping problems, fatigue, nervousness, and mood swings.

Six and three patients in the OLP and the no-treatment group respectively reported that their hot flushes had gotten worse (“much worse” or “minimally worse”). One patient in the no-treatment group got “very much worse”. Three cases dropped out due to symptom aggravation, of which one (no-treatment group) dropped out during the treatment phase, and two (OLP) dropped out after the treatment phase, i.e., did not participate in the follow-up. One patient refused to continue the treatment for another 4 weeks after increased rates of diarrhoea during the first 4 weeks of treatment. Abdominal symptoms are unlikely to be caused by the pills’ lactose loads<sup>32</sup> yet may be associated with nocebo effects.

**Opinion of OLP.** Patients’ opinions were assessed at week 4, i.e., after the treatment phase. Of 97 patients who stated an opinion, 70% thought that OLP usage in clinical practice is ‘definitely legitimate’. Another 25% indicated ‘rather yes’ whereas 8% indicated ‘rather no’. Some women specified their choice. Quotes are shown in Supplementary Table S2.

## Discussion

In this first study to investigate open-label placebos for menopausal hot flushes, we found that hot flush symptoms decreased significantly in comparison to no-treatment. The primary outcome hot flush score decreased by 43% after 4 weeks of placebo intake, which was significantly better than a 19% reduction in the no-treatment group.

	Enrolment (M, SD) <sup>a</sup>	Baseline (M, SD) <sup>a</sup>	OLP vs. no-treatment	Week 4 (M, SD) <sup>b</sup>	OLP 8 week vs. OLP 4 week	Week 8 (M, SD) <sup>c</sup>
<b>Expectations</b>						
No-treatment	2.28 (2.65)	2.88 (2.62)	$\beta = -0.24,$ $p = 0.005$	2.88 (3.02)		
OLP	3.22 (2.84)	4.76 (2.93)		See subgroups		
OLP 8 week				5.18 (3.54)	$\beta = -0.12, p = 0.20$	4.64 (3.66)
OLP 4 week				4.43 (3.20)		3.29 (3.44)
<b>Hope</b>						
No-treatment	5.70 (3.18)	6.88 (3.09)	$\beta = -0.11, p = 0.48^d$	6.33 (3.21)		
OLP	5.80 (3.35)	7.58 (2.44)		See subgroups		
OLP 8 week				6.95 (2.85)	$\beta = -0.04,$ $p = 0.79^d$	5.86 (3.66)
OLP 4 week				6.86 (2.43)		6.81 (2.58)

**Table 3.** Group differences in expectations of and hope for hot flush improvement. Patients were asked whether they expected [hoped] that hot flushes would improve in the ‘next week’ (enrolment) or ‘in the next 4 weeks’ (baseline, week 4, week 8). Both scales range from 0 ‘no improvement’ to 10 ‘maximum improvement’. At baseline and week 4, expectation and hope were assessed *after* the randomization. Tests of group differences were adjusted for the respective value at enrolment. OLP Open-label placebo. <sup>a</sup>No-treatment: N = 50; OLP: N = 50. <sup>b</sup>No-treatment: N = 48; OLP 8 week: n = 22; OLP 4 week: n = 21. <sup>c</sup>OLP 8 week: n = 22; OLP 4 week: n = 21. <sup>d</sup>Due to violation of homoscedasticity, significance tests were based on standard errors obtained using bootstrapping.

	Open-label placebo (n = 50)	No-treatment (N = 50)
Decreased appetite	3	3
Dry mouth	5	9
Sleeping problems	2	5
Nausea	2	9
Dizziness	7	7
Fatigue	5	12
Constipation	6	3
Nervousness	6	8
Mood swings	6	10

**Table 4.** Baseline-adjusted occurrence of adverse events.

We also found a decrease in daily hot flushes by 33% and improvements in overall menopause-related quality-of-life. Overall improvement was reported by 62% of women in the OLP group, whereas a clinically meaningful improvement, i.e., 50% lesser hot flushes, was obtained by 24%. In comparison, many women in the no-treatment also reported overall improvement (40%), yet only 6% achieved hot flush reductions that were clinically meaningful. Compared with control, OLP reduced the number of hot flushes per day by 1.12 points. This change is similar to reductions achieved with SSRI/SNRI (1.13), soy isoflavones (1.15), and clonidine (0.95)<sup>33</sup>, yet minor in relation to hormonal therapy (16.8–22.4, depending on the active ingredient)<sup>26,33</sup>.

The treatment effect on the hot flush score, which we log-transformed in the analyses, was large. The magnitude of this effect aligns with previous OLP studies of irritable bowel syndrome<sup>3</sup> or chronic low back pain<sup>4</sup>. It is well established that placebo effects in double-blind hot flush trials are substantial, with rates ranging between 20 and 50%<sup>20</sup>. Our study demonstrated that openly administered placebos show benefits similar to placebos given under double-blind uncertainty.

We found no treatment effect on problem rating. Problem rating refers to the degree to which hot flushes are regarded as problematic, distressing, and causing interference with day-to-day life. It is argued that problem rating, not numbers or severity of hot flushes per se, determines whether a patient would seek treatment<sup>11</sup>. This outcome was mainly included in psychological interventions for hot flushes. Improvements in problem rating were achieved by cognitive-behavioural therapy (CBT)<sup>24,34</sup> but not by mindfulness-based stress reduction (MBSR)<sup>35</sup>. During CBT, patients learn to better manage their hot flushes, e.g., in social situations in which women often feel embarrassed when flushed or considering sleeping problems<sup>34</sup>. Similar to MBSR, our treatment did not directly target hot flush management, which presumably explains the lack of a treatment effect. However, problem ratings were significantly reduced within both groups. In particular, problem rating decreased by 1.9 points in the OLP group, and by 1.4 points in the no-treatment group. The benefits in the OLP group are noteworthy, given that a 2-point reduction is considered clinically significant<sup>34</sup>. We hypothesize these improvements to be related to the three study visits in which patients interacted with a warm, empathetic clinician. A positive patient-clinician



relationship is known to be beneficial for health outcomes<sup>36,37</sup> and is discussed to have also caused improvements in the no-treatment arm of previous OLP studies<sup>3</sup>.

There was a positive effect on the overall quality of life, but not on the subdomains of quality of life. These included anxiety and depression, well-being, somatic symptoms, memory/concentration, and sleep problems. A reduction in hot flushes might have brought about collateral benefits to other complaints like insomnia, fatigue, irritability, and depressed mood-symptoms that were recorded with the overall quality of life scale. However, these secondary benefits were seemingly not large enough to cause significant reductions in each subdomain. These findings are comparable to the first OLP RCT with irritable bowel syndrome patients in which the groups differed in trend considering quality of life<sup>3</sup>.

Notably, the benefits of the placebo persisted beyond intake. On average, and for most patients (52%) who discontinued after 4 weeks, improvements remained for another 4 weeks. Interestingly, taking the placebo for 4 or 8 weeks did not affect whether the benefits remained at follow-up. Freeman has reported comparable findings for double-blind drug trials of hot flushes<sup>25</sup>. After discontinuing or tapering placebo pills, most patients remained improved. Sustained OLP effects were also found in the first two studies including exploratory follow-up phases with cancer patients with chronic fatigue after undergoing chemotherapy<sup>8</sup> and chronic back pain patients<sup>5</sup>.

There is an ongoing debate about the underlying mechanisms of open-label placebos. Expectations are—alongside with conditioning—the most researched and established mechanisms of placebo effects in general<sup>38</sup>. Researchers have argued that, by providing all patients with a briefing about the placebo effect at the beginning, i.e. prior to randomization, positive expectations might have been elicited<sup>39</sup>. Expecting treatment benefits in turn might influence actual treatment outcomes. Indeed, after randomization, the OLP group expected more improvements than the no-treatment group. However, there was no link between expectations and change in hot flush scores. These results align with a previous study on acupuncture for hot flushes, in which expectations were also not predictive of symptom change<sup>40</sup>. After taking 4 weeks of placebo, patients were randomized again; this time to either continue or discontinue the OLP treatment. Although descriptively, expectations were higher when treatment continued, the difference between the groups was nonetheless smaller in comparison to the first allocation. Whereas patients had to rely on “external” information when formulating their expectations at the first allocation, they were able to incorporate first-hand experiences in the second allocation. Overall, our results indicate that the relationship between expectations and treatment benefits is not as straightforward, i.e., simply expecting to get better is not the reason why someone does get better. However, we should note that our design is not suited or aimed at testing underlying mechanisms of OLP effects. This is better investigated in studies in which OLP is paired with different kinds of information that aim at eliciting different degrees of expectations<sup>41,42</sup>. Aside from expectation, we also looked at hope for improvement and optimism. Hope was not affected by the group assignment and was higher than expectations, which aligns with qualitative studies on complementary and alternative therapies<sup>43,44</sup>.

Like expectations, hope also did not predict change in hot flush score or moderate the treatment effect. Moreover, we found no link between optimism and the outcome, which aligns with a previous study that reported no predictive value of optimism on OLP effects in healthy participants<sup>45</sup>. Some researchers have begun to propose that OLP relies on non-conscious Bayesian cortical processing that can significantly shape perceptions of symptoms<sup>46</sup>. Taken together, understanding the underlying mechanisms of OLP requires further investigation.

Some limitations should be noted. First, our follow-up period was rather short. The 4-week follow-up allows for first insights considering the time course of placebo effects yet is too short to evaluate clinical usefulness. The European Medicines Agency, for example, requires a follow-up period of 12-week for market approval of hot flush remedies<sup>47</sup>. Second, we recruited via the general population. Therefore, symptoms were not validated by gynaecologists. However, our objective was to reach those women who suffered from hot flushes yet did not see a practitioner since they rejected hormone therapy<sup>48</sup>. As reflected in our baseline characteristics, most patients have used remedies like for example black cohosh, homoeopathic remedies, sage, etc. which are sold as over-the-counter medicines or food supplements and do not require seeing a practitioner. Considering that practitioners also mostly rely on women's self-reports—hormonal tests for women over 45 with irregular or absent menses to determine menopausal status are explicitly discouraged in the guidelines<sup>21,22</sup>—we believe our procedure to be valid. Besides, we excluded other potential causes of flushing via the structured screening interview. The third limitation concerns adherence. First of all, we assessed adherence only through self-report. Secondly, we defined adherence as taking more than 80% of all pills, which applied to all patients. However, 34% indicated to have taken “almost all” pills. Based on the participants' feedback during the study visits, we presume that this patient sample is not used to taking medications daily, thus making it likely for women to forget one or two pills. Other frequent reasons for non-adherence such as complexity of regimen and side effect burden do not apply to our treatment<sup>49</sup>. We also aimed to establish a supportive patient-clinician relationship based on trust, which in turn can enhance adherence. However, since forgetfulness is one of the most common reasons for non-adherence<sup>49</sup>, future studies might consider using medication-related strategies to increase adherence, like providing pill boxes or notifications via an App. Safeguarding adherence is especially relevant for future OLP studies with treatment periods longer than three or four weeks. Lastly, in our sample, the total symptom duration ranged between 30 days and 26 years, thus constituting both a potential limitation and a strength with regards to ecological validity.

Does this study warrant the use of OLP in clinical practice? Since this is the first study to investigate OLP for hot flushes with a short follow-up period and a rather small sample size in comparison to trials of other hot flush remedies, we believe that further evidence is needed to draw definite recommendations. However, our findings have clinical value as they provide a basis for re-evaluating the role of placebo treatments in medical practice. Given that complementary and alternative treatments do not outperform placebos in RCTs but are nonetheless commonly used by menopausal women<sup>48,50</sup>, it may be conceivable to consider OLP as an additional option. In the German S3-clinical guidelines of treating peri- and postmenopausal symptoms, placebos are already placed

as an optional hot flush treatment next to, among others, phytoestrogen and CBT<sup>21</sup>. Therefore, in Germany, specifically, our study provides a basis for clinicians to administer placebos under ethical values. It is important to note that we tested the efficacy of OLP within a clinical context. That is, we are uncertain about whether women self-administering placebo treatment would be as effective as prescribing placebo by a health care provider, given that evidence strongly suggests that placebo effects are enhanced by a supportive relationship<sup>36,37,51</sup>. In summary, the positive effects of OLP in this trial indicate a potential for OLP to be a safe, effective treatment for women with moderate and severe hot flushes.

## Methods

**Trial design.** The study protocol<sup>52</sup> provides a detailed description of the design. Briefly, in this parallel group, two-arm, superiority trial, participants were allocated 1:1 to receive 4 weeks of open-label placebos or no treatment. During the study period, participants completed a hot flush diary. A phone interview that was conducted by study assistants screened for eligibility. We formed a stringent questionnaire that either qualified or disqualified women for study participation. Enrolled women received three study visits in which they interacted with a clinician and completed questionnaires. The OLP group was randomized a second time at the third study visit to either take another four weeks of open-label placebos (OLP 8 week) or to discontinue their OLP treatment (OLP 4 week). To obtain a baseline on hot flushes, women completed the diary for 7 days before randomization, i.e., between study visit 1 and 2. The first randomization took place at study visit 2. In turn, study visits 3, and 4 (OLP 8 week and OLP 4 week only) took place at 4 and 8 weeks after the first randomization. The clinician conducted short calls inquiring for symptom status, potential adverse effects, and adherence (OLP only) at 2 and 6 weeks after the first randomization. Ethical approval was given by the ethics committee of the Hamburg Medical Association (“Hamburger Ärztekammer”) (trial number: PV5787), and experiments were performed in accordance with the Declaration of Helsinki<sup>53</sup>. The trial was registered on 12/02/2019 at ClinicalTrials.gov (registration number: NCT03838523). Written informed consent was given by all participants prior to enrolment. This trial was conducted at the University Medical Centre Hamburg-Eppendorf, Department of Psychosomatic Medicine.

**Participants.** We included women who were post-menopausal (12 months from last menstruation), or in menopausal transition (amenorrhea  $\geq 60$  days in the past year). Eligibility criteria required women to have at least five daily hot flushes of moderate or severe intensity which were considered problematic (16+ on the problem rating scale of the Hot Flush Rating Scale<sup>54</sup>). Women who used any kind of hot flush treatment or SSRI/SSNI within the past 6 weeks were excluded. Also, other potential causes of flushes were excluded: risk of an alcohol use disorder (5+ on the Alcohol Consumption Questions<sup>55</sup>), cancer diagnosis within the past ten years, current endocrine therapy for breast cancer, untreated hyperthyroidism, high levels of either depression, or anxiety, or both (9+ on the Patient Health Questionnaire-4<sup>56</sup> or 5+ on one of its subscales). Mind-body interventions like yoga, meditation, etc., were permitted if they were ongoing throughout the treatment. At informed consent, all patients were encouraged to refrain from lifestyle changes during their participation in the trial.

**Interventions.** Both groups underwent the same interactions with the clinician (female psychologist) and received the same information about open-label placebos. The no-treatment group controls for spontaneous improvement, regression to the mean, and normal fluctuations in symptoms. The clinician pursued to treat all patients in an empathetic, warm manner.

The clinician informed all participants that the placebo contained no active ingredient and was made of lactose. At study visit 1, i.e., prior randomization, the clinician takes a history and inquiries about relevant biopsychosocial information (10–15 min). Afterwards, each patient receives the following information about placebos<sup>3</sup>: (1) the placebo effect is powerful; patients reported symptom improvement after taking placebos in double-blind drug trials, including hot flush trials. However, these participants were unaware of whether they were receiving a placebo or medicine. (2) A few studies have shown that placebos without deception can have beneficial effects. The body may react to the pill intake automatically (Depending on the patient's prior knowledge, an example is given, e.g., the Pavlov dog or food poisoning for which certain foods cause queasiness and nausea). Being positive and believing in a positive effect can help but is not necessary. (4) Taking the pills faithfully twice a day is crucial since the custom of pill intake can contribute to the effect. Also, adhering to the instructions is important to maintain the quality of the study, i.e., if some patients would only take half their pills, we would be comparing several subgroups with the no-treatment group. (5) Finally, we do not know whether placebos without deception can reduce hot flushes. Therefore, we encourage patients to simply “give it a try.”

The groups differ solely considering pill administration plus intake and inquiries about adherence. At study visit 2, patients were randomized. Patients in the OLP group received a glass bottle containing the pills labelled “Placebos for menopausal hot flushes” with the original medication leaflet. The placebo pills (white and uncoated) were manufactured by the company “Zentiva Pharma GmbH” and bottled by the UMC Hamburg pharmacy for this study. The no-treatment group was reminded about the importance of the control group and to complete the diary faithfully. The no-treatment group received placebo pills at the end of the study if desired.

**Outcomes.** Given the novelty of this study, we included two primary outcomes: group difference in hot flush score change from baseline to treatment end (week 4), and group difference in problem rating change measured with the Hot Flush Rating Scale (HFRS<sup>54</sup>). The hot flush score (frequency  $\times$  severity), measured with a diary, is the most often used outcome in hot flush trials<sup>57</sup>. The diary is considered a gold standard to assess hot flushes<sup>58</sup>. Flushes during daytime were recorded upon occurrence, and night sweats were recorded upon awakening. To facilitate recording, patients received separate diaries for day and night. Categories of each hot flush (1 = mild, 2 = moderate, 3 = severe) were pre-defined following the European Medicine Agency's guidelines<sup>47</sup>. Problem rat-

ing refers to the perceived hot flush burden considering cognition, emotion, and day-to-day interference. A 2-point improvement is regarded as clinically relevant<sup>34</sup>. Both the diary and the HFRS have demonstrated good reliability and validity<sup>54,58</sup>.

Secondary outcomes include change in hot flush frequency, change in health-related quality of life, overall improvement (PGIC)<sup>59</sup>, and number of responders (50% reduction in frequency from baseline to treatment end). Frequency is defined as the number of hot flushes per 24 h. The overall quality of life was measured with the Menopause Rating Scale-II (MRS-II)<sup>60,61</sup>, whereas subdomains of quality of life were measured with the Women's Health Questionnaire (WHQ-23)<sup>62</sup>. The MRS-II evaluates the severity of 11 common menopausal complaints. Of the original eight WHQ scales, we excluded the optional scales 'menstrual symptoms' and 'sexual behaviour' due to low relevance in this sample (latter scale completed by  $n = 40$  women). The 'vasomotor symptoms' scale was omitted due to redundancy. Analyses were conducted with the subdomains 'anxiety and depression', 'well-being' ["feels well with herself (good appetite, feels physically attractive)"], 'somatic symptoms', 'memory/concentration', and 'sleep problems'. Each subdomain ranges from 1 to 100, with higher scores indicating better health. Improvements of 10–20 points are considered clinically relevant<sup>63</sup>. For both quality of life questionnaires, validity, good reliability and sensitivity to change have been demonstrated<sup>60,62</sup>. PGIC assessed the subjective overall improvement of hot flushes within the past four weeks from 1 'very much worse' to 7 'very much better'<sup>59</sup>. We calculated a binary scale for easier interpretation, with scores larger than 4 ('no change') being categorized as 'improved'. To obtain the number of responders, we compared the hot flush frequency at baseline with the frequency at treatment end.

**Further assessments.** We assessed expectations and hope of symptom improvement at each study visit on a scale from 0 'no change' to 10 'maximum change'. As there was no validated questionnaire when the study was designed, we created one-item scales based on interview questions in a relevant qualitative study<sup>64</sup>. At baseline, we additionally measured stress (Perceived Stress Scale-10<sup>65</sup>), and dispositional optimism (Life orientation test-revised<sup>66</sup>). At treatment end, patients indicated whether they made any lifestyle changes considering smoking, alcohol, exercise, diet, and mind-body treatments. Patients also rated whether OLP use in clinical practice is acceptable from 1 'definitely yes' to 4 'definitely not'<sup>67</sup>. Adherence is assessed 4 and 8 weeks after randomization using one item ("How many placebo pills have you taken in the last week?") ranging from 0 'none' to 6 'all'<sup>68</sup>. Patients who indicated to have taken 'almost all' and 'all' their pills were considered adherent. Adverse events (e.g., nocebo effects) were recorded using the Generic Assessment of Side Effects<sup>69</sup>. We chose the nine most frequent symptoms that were reported by women of the German general population, i.e., reduced appetite, dry mouth, sleeping problems, nausea, dizziness, fatigue, constipation, nervousness, and mood swings<sup>69</sup>. Symptoms that were already indicated at baseline were not classified as adverse events.

**Sample size.** Previous OLP studies have found large effects of Cohen's  $d = 0.80$ <sup>3,4</sup> and hot flush double-blind trials have reported placebo improvements of moderate size ( $d = 0.40$ )<sup>70</sup>. Given a presumed effect of  $d = 0.60$ , an alpha of 0.05 (two-tailed), 80% power, and finally, an attrition rate of 10%, we obtained our required sample of  $N = 100$ . The sample size was calculated with G-Power.

**Randomization.** The clinician conducted the randomization at study visit 2 by opening a sequentially numbered, opaque envelope. The group affiliation was communicated to the patient immediately after the reveal. We used a permuted block randomization with block sizes of 4 and 6.

We conducted a second randomization at study visit 3 among the OLP group, stratified by improvement (yes/no). After inquiring about the stratum, the clinician draws the next envelope from the corresponding deck and reveals the group affiliation to the patient. In alignment with the first randomization, envelopes were sequentially numbered and opaque. We used a stratified permuted block randomization with block sizes of 2 and 4.

Both sequences were generated using an online program (Sealed Envelope). A research assistant who had no interaction with patients generated the sequences and prepared the envelopes. The sequences were locked away before enrolment to avert access by study personnel. Similarly, block sizes remained unknown to the study personnel throughout the study.

**Blinding.** Due to the nature of the study, patients and the clinician were not blinded. Assessments of primary and secondary outcomes were conducted by research assistants who were blinded to allocation (assessors). At each study visit, clinician consultation and assessment took place in separate rooms. Both the clinician and the assessor reminded patients of not revealing their group affiliation when completing the questionnaires, verbally or by accident (e.g., putting placebo pills next to the questionnaire after receiving them). Group affiliation in questionnaires was masked by using letters (A vs. B and C vs. D) instead of group names.

**Handling missing values.** Provided that there is 80% of data on each scale, missing items were substituted by the mean of the non-missing items<sup>71</sup>. The single-item scales expectations and hope were not replaced. Missing data points in the diary were imputed separately for daytime and night-time. That is, missing data points of night sweats were substituted by the average night sweats of the rest of the week; computation of the 24 h-value followed hereafter. A previous study has found no considerable difference between a 3-day and a 7-day weekly mean of the hot flush diary<sup>72</sup>. Missing values (per scale or data point) were not imputed since linear mixed models allow statistical evaluation of incomplete data<sup>31</sup>. Missing values of categorical variables (if < 5%) were imputed with the most likely answer. That is, in the case of dropout, the categorical outcomes 'improved (yes/no)' and 'responded (yes/no)' were imputed with 'no'.

**Statistical analyses.** *Efficacy.* For metric variables, we conducted linear mixed models for repeated measures. Measurement points were nested within participants. For hot flush score and frequency, we computed a weekly average per 24 h, thus obtaining five measurement points for each outcome. The hot flush score is a composite score and was right-skewed. Since the values were arbitrary, we performed a log-transformation to obtain distributions closer to normal. Problem rating and quality of life scales were assessed at baseline and treatment end, yielding two measurement points per outcome. We included the fixed effects Time (Factor), Group, Time × Group, and modelled the Intercept as a random effect. For the hot flush score, frequency, problem rating, and overall quality of life (MRS-II), we included the respective grand-mean centred baseline value as a covariate. Models of the quality of life subdomains (WHQ) did not converge when the respective baseline value of each domain was included. Consequently, we included overall quality of life (MRS-II) at baseline as a covariate for all quality of life variables. Estimates were obtained with the restricted maximum likelihood (REML) method. A heterogeneous autoregressive covariance structure type was set as the default. If the model did not converge, we chose an easier model, i.e., proceeded with a more restrictive covariance structure type until convergence was reached (heterogeneous autoregressive to homogenous autoregressive to variance components). Categorical variables, including improvement (yes/no) and responding to treatment (yes/no), were analysed with Chi-Square tests and Fisher's Exact Test (cell number < 5), respectively.

Sensitivity analyses for primary outcomes were pre-defined in the study protocol and included: (1) imputing missing values of dropped cases, (2) deleting cases who were non-adherent, (3) statistically adjusting for perceived stress and years of hot flushes (potential confounders)<sup>11</sup>, and (4) excluding patients who made lifestyle changes.

*Expectations, hope, and optimism.* We conducted regression analyses to examine whether expectations and hope differed between the groups. All analyses were adjusted for its respective value at enrolment. To investigate whether expectation at baseline predicted or/and moderated the treatment effect, we extended the model of our primary outcome by Expectation, Expectation × Group, and Expectation × Group × Time. Matching analyses were conducted for hope and optimism.

*Maintenance and duration.* We were interested in the following questions: First, do benefits sustain in the OLP group at follow-up end (week 8)? Second, do benefits sustain in the OLP 4 week group at follow-up end? Third, do OLP 8 week and OLP 4 week differ at 8 weeks? Fourth, which kinds of symptom trajectories are present? For questions 1–3, we conducted a linear mixed model to compare the week 4 to week 8 change in log-transformed hot flushes. We included the fixed effects Time (factor with five categories: week 4, 5, 6, 7, 8), Intake Duration (OLP 8 week vs. OLP 4 week), Time × Intake Duration, and the centred log-transformed score at week 4. For question one, we compared the score at week 4 and week 8 within OLP. For question two, we conducted the same comparison within OLP 4 week. For question three, we compared OLP 8 week and OLP 4 week at week 8. Model specifications corresponded with the ones for the efficacy analyses. For question 4, we grouped symptom trajectories based on visual examination and patients' improvement ratings. All analyses on the follow-up period were pre-defined in the protocol and limited to primary outcomes. Given the small sample size of the OLP subgroups, we prioritize descriptive data over statistical significance when interpreting the data.

*Assumptions.* Prior analyses, we inspected model assumptions of linear modelling: Homoscedasticity, normality of residuals, no multicollinearity or multivariable outliers. To comply with standards of intention-to-treat analyses, we report results with the full sample. In the case of multivariate outliers, results are compared between the analyses, including and excluding the outlier. In case homoscedasticity was not met, we obtained robust errors by using bootstrap methods (1000 samples, bias and accelerated method). Normality of residuals was met for all analyses.

IBM SPSS 26.0 was used for all analyses. We used the MIXED command to conduct linear mixed models and the EMMEANS subcommand for significance testing. Effect sizes of pairwise comparisons of linear mixed models are calculated by dividing the mean group difference by the pooled standard deviation<sup>73</sup>. Tests were two-tailed with an alpha error set at 0.05. Effects of 0.2, 0.5, 0.8 (Cohen's *d*) and 0.1, 0.3, 0.5 (Cramer's *V*, *df* = 1) were considered small, moderate, large effects<sup>73</sup>.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 8 April 2020; Accepted: 6 November 2020

Published online: 18 November 2020

## References

1. Fassler, M., Meissner, K., Schneider, A. & Linde, K. Frequency and circumstances of placebo use in clinical practice—A systematic review of empirical studies. *BMC Med.* **8**, 15. <https://doi.org/10.1186/1741-7015-8-15> (2010).
2. Tilburt, J. C., Emanuel, E. J., Kaptchuk, T. J., Curlin, F. A. & Miller, F. G. Prescribing, “placebo treatments”: Results of national survey of US internists and rheumatologists. *BMJ* **337**, a1938. <https://doi.org/10.1136/bmj.a1938> (2008).
3. Kaptchuk, T. J. *et al.* Placebos without deception: A randomized controlled trial in irritable bowel syndrome. *PLoS ONE* **5**, e15591. <https://doi.org/10.1371/journal.pone.0015591> (2010).
4. Carvalho, C. *et al.* Open-label placebo treatment in chronic low back pain: A randomized controlled trial. *Pain* **157**, 2766–2772. <https://doi.org/10.1097/j.pain.0000000000000700> (2016).

5. Kleine-Borgmann, J., Schmidt, K., Hellmann, A. & Bingel, U. Effects of open-label placebo on pain, functional disability, and spine mobility in patients with chronic back pain: A randomized controlled trial. *Pain* **160**, 2891–2897. <https://doi.org/10.1097/j.pain.0000000000001683> (2019).
6. Kelley, J. M., Kaptchuk, T. J., Cusin, C., Lipkin, S. & Fava, M. Open-label placebo for major depressive disorder: A pilot randomized controlled trial. *Psychother. Psychosom.* **81**, 312–314. <https://doi.org/10.1159/000337053> (2012).
7. Schaefer, M., Harke, R. & Denke, C. Open-label placebos improve symptoms in allergic rhinitis: A randomized controlled trial. *Psychother. Psychosom.* **85**, 373–374. <https://doi.org/10.1159/000447242> (2016).
8. Hoenemeyer, T. W., Kaptchuk, T. J., Mehta, T. S. & Fontaine, K. R. Open-label placebo treatment for cancer-related fatigue: A randomized-controlled clinical trial. *Sci. Rep.* **8**, 2784. <https://doi.org/10.1038/s41598-018-20993-y> (2018).
9. Charlesworth, J. E. G. *et al.* Effects of placebos without deception compared with no treatment: A systematic review and meta-analysis. *J. Evid. Based Med.* **10**, 97–107. <https://doi.org/10.1111/jebm.12251> (2017).
10. Avis, N. E. *et al.* Duration of menopausal vasomotor symptoms over the menopause transition. *JAMA Intern. Med.* **175**, 531–539. <https://doi.org/10.1001/jamainternmed.2014.8063> (2015).
11. Hunter, M. S. *et al.* Prevalence, frequency and problem rating of hot flushes persist in older postmenopausal women: Impact of age, body mass index, hysterectomy, hormone therapy use, lifestyle and mood in a cross-sectional cohort study of 10,418 British women aged 54–65. *BJOG* **119**, 40–50. <https://doi.org/10.1111/j.1471-0528.2011.03166.x> (2012).
12. Sherif, K. *Hormone Therapy: A Clinical Handbook* (Springer, New York, 2013).
13. The North American Menopause Society. Nonhormonal management of menopause-associated vasomotor symptoms: 2015 position statement of The North American Menopause Society. *Menopause* **22**, 1155–1172. <https://doi.org/10.1097/GME.0000000000000546> (2015).
14. Rossouw, J. E. *et al.* Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* **288**, 321–333. <https://doi.org/10.1001/jama.288.3.321> (2002).
15. Collaborative Group on Hormonal Factors in Breast Cancer. Type and timing of menopausal hormone therapy and breast cancer risk: Individual participant meta-analysis of the worldwide epidemiological evidence. *Lancet* **394**, 1159–1168. [https://doi.org/10.1016/s0140-6736\(19\)31709-x](https://doi.org/10.1016/s0140-6736(19)31709-x) (2019).
16. Leach, M. J. & Moore, V. Black cohosh (*Cimicifuga* spp.) for menopausal symptoms. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD007244.pub2> (2012).
17. Lethaby, A. *et al.* Phytoestrogens for menopausal vasomotor symptoms. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD001395.pub4> (2013).
18. Franco, O. H. *et al.* Use of plant-based therapies and menopausal symptoms: A systematic review and meta-analysis. *JAMA* **315**, 2554–2563. <https://doi.org/10.1001/jama.2016.8012> (2016).
19. Dodin, S. *et al.* Acupuncture for menopausal hot flushes. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD007410.pub2> (2013).
20. Nelson, H. D. *et al.* Nonhormonal therapies for menopausal hot flashes: Systematic review and meta-analysis. *JAMA* **295**, 2057–2071. <https://doi.org/10.1001/jama.295.17.2057> (2006).
21. German Society of Gynecology and Obstetrics. *Guideline: Peri- and Postmenopause—Diagnosis and Interventions* (S-3 Level, AWMF Registry No. 015-062, October 2018). (2018).
22. NICE guideline. *Menopause: Diagnosis and Management* (NG23, last updated: December 2019). (2015).
23. Elkins, G. R., Fisher, W. I., Johnson, A. K., Carpenter, J. S. & Keith, T. Z. Clinical hypnosis in the treatment of postmenopausal hot flashes: A randomized controlled trial. *Menopause* **20**, 291–298. <https://doi.org/10.1097/GME.0b013e31826ce3ed> (2013).
24. Ayers, B., Smith, M., Hellier, J., Mann, E. & Hunter, M. S. Effectiveness of group and self-help cognitive behavior therapy in reducing problematic menopausal hot flushes and night sweats (MENOS 2): A randomized controlled trial. *Menopause* **19**, 749–759. <https://doi.org/10.1097/gme.0b013e31823fe835> (2012).
25. Freeman, E. W. *et al.* Placebo improvement in pharmacologic treatment of menopausal hot flashes: Time course, duration, and predictors. *Psychosom. Med.* **77**, 167–175. <https://doi.org/10.1097/PSY.0000000000000143> (2015).
26. Maclennan, A. H., Broadbent, J. L., Lester, S. & Moore, V. Oral oestrogen and combined oestrogen/progestogen therapy versus placebo for hot flushes. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD002978.pub2> (2004).
27. Kaptchuk, T. J. & Miller, F. G. Open label placebo: Can honestly prescribed placebos evoke meaningful therapeutic benefits?. *BMJ* **363**, k3889. <https://doi.org/10.1136/bmj.k3889> (2018).
28. Finnis, D. G., Kaptchuk, T. J., Miller, F. & Benedetti, F. Biological, clinical, and ethical advances of placebo effects. *Lancet* **375**, 686–695. [https://doi.org/10.1016/s0140-6736\(09\)61706-2](https://doi.org/10.1016/s0140-6736(09)61706-2) (2010).
29. Petrie, K. J. & Rief, W. Psychobiological mechanisms of placebo and nocebo effects: Pathways to improve treatments and reduce side effects. *Annu. Rev. Psychol.* **70**, 599–625. <https://doi.org/10.1146/annurev-psych-010418-102907> (2019).
30. Kaptchuk, T. J. Open-label placebo: Reflections on a research agenda. *Perspect. Biol. Med.* **61**, 311–334. <https://doi.org/10.1353/pbm.2018.0045> (2018).
31. Snijders, T. A. B. & Bosker, R. J. *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. (Sage, 1999).
32. Suarez, F. L., Savaiano, D. A. & Levitt, M. D. A comparison of symptoms after the consumption of milk or lactose-hydrolyzed milk by people with self-reported severe lactose intolerance. *N. Engl. J. Med.* **333**, 1–4. <https://doi.org/10.1056/NEJM199507063330101> (1995).
33. Stearns, V. Clinical update: New treatments for hot flushes. *Lancet* **369**, 2062–2064. [https://doi.org/10.1016/s0140-6736\(07\)60959-3](https://doi.org/10.1016/s0140-6736(07)60959-3) (2007).
34. Mann, E. *et al.* Cognitive behavioural treatment for women who have menopausal symptoms after breast cancer treatment (MENOS 1): A randomised controlled trial. *Lancet Oncol.* **13**, 309–318. [https://doi.org/10.1016/s1470-2045\(11\)70364-3](https://doi.org/10.1016/s1470-2045(11)70364-3) (2012).
35. Carmody, J. F. *et al.* Mindfulness training for coping with hot flashes: Results of a randomized trial. *Menopause* **18**, 611–620. <https://doi.org/10.1097/gme.0b013e318204a05c> (2011).
36. Kelley, J. M., Kraft-Todd, G., Schapira, L., Kossowsky, J. & Riess, H. The influence of the patient-clinician relationship on healthcare outcomes: A systematic review and meta-analysis of randomized controlled trials. *PLoS ONE* **9**, e94207. <https://doi.org/10.1371/journal.pone.0094207> (2014).
37. Kaptchuk, T. J. *et al.* Components of placebo effect: Randomised controlled trial in patients with irritable bowel syndrome. *BMJ* **336**, 999–1003. <https://doi.org/10.1136/bmj.39524.439618.25> (2008).
38. Wager, T. D. & Atlas, L. Y. The neuroscience of placebo effects: Connecting context, learning and health. *Nat. Rev. Neurosci.* **16**, 403–418. <https://doi.org/10.1038/nrn3976> (2015).
39. Colloca, L. & Howick, J. Placebos without deception: Outcomes, mechanisms, and ethics. *Int. Rev. Neurobiol.* **138**, 219–240. <https://doi.org/10.1016/bs.irn.2018.01.005> (2018).
40. Ee, C. C. *et al.* Expectancy after the first treatment and response to acupuncture for menopausal hot flashes. *PLoS ONE* **12**, e0186966. <https://doi.org/10.1371/journal.pone.0186966> (2017).
41. Kube, T. *et al.* Deceptive and nondeceptive placebos to reduce pain. *Clin. J. Pain* **36**, 68–79. <https://doi.org/10.1097/ajp.0000000000000781> (2020).
42. Locher, C. *et al.* Is the rationale more important than deception? A randomized controlled trial of open-label placebo analgesia. *Pain* **158**, 2320–2328. <https://doi.org/10.1097/j.pain.0000000000001012> (2017).

43. Eaves, E. R. *et al.* A qualitative study of changes in expectations over time among patients with chronic low back pain seeking four CAM therapies. *BMC Complement Altern. Med.* **15**, 12. <https://doi.org/10.1186/s12906-015-0531-9> (2015).
44. Kaptchuk, T. J. *et al.* “Maybe I made up the whole thing”: Placebos and patients’ experiences in a randomized controlled trial. *Cult. Med. Psychiatry* **33**, 382–411. <https://doi.org/10.1007/s11013-009-9141-7> (2009).
45. Locher, C., Frey Nascimento, A., Kossowsky, J., Meyer, A. & Gaab, J. Open-label placebo response—Does optimism matter? A secondary-analysis of a randomized controlled trial. *J. Psychosom. Res.* **116**, 25–30. <https://doi.org/10.1016/j.jpsychores.2018.11.009> (2019).
46. Ongaro, G. & Kaptchuk, T. J. Symptom perception, placebo effects, and the Bayesian brain. *Pain* **160**, 1–4. <https://doi.org/10.1097/j.pain.0000000000001367> (2019).
47. European Medicines Agency. *Guideline on clinical investigation of medicinal products for hormone replacement therapy of oestrogen deficiency symptoms in postmenopausal women* (EMA, London, 2005).
48. Posadzki, P. *et al.* Prevalence of complementary and alternative medicine (CAM) use by menopausal women: A systematic review of surveys. *Maturitas* **75**, 34–43. <https://doi.org/10.1016/j.maturitas.2013.02.005> (2013).
49. World Health Organization. *Adherence to Long-Term Therapies. Evidence for Action*. 194 (WHO, 2003).
50. Buhling, K. J., Daniels, B. V., Studnitz, F. S., Eulenburg, C. & Mueck, A. O. The use of complementary and alternative medicine by women transitioning through menopause in Germany: Results of a survey of women aged 45–60 years. *Complement Ther. Med.* **22**, 94–98. <https://doi.org/10.1016/j.ctim.2013.12.004> (2014).
51. Suarez-Almazor, M. E. *et al.* A randomized controlled trial of acupuncture for osteoarthritis of the knee: Effects of patient-provider communication. *Arthritis Care Res. (Hoboken)* **62**, 1229–1236. <https://doi.org/10.1002/acr.20225> (2010).
52. Pan, Y. *et al.* Non-concealed placebo treatment for menopausal hot flashes: Study protocol of a randomized-controlled trial. *Trials* **20**, 508. <https://doi.org/10.1186/s13063-019-3575-1> (2019).
53. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA* **310**, 2191–2194. <https://doi.org/10.1001/jama.2013.281053> (2013).
54. Hunter, M. S. & Liao, K. L. A psychological analysis of menopausal hot flashes. *Br. J. Clin. Psychol.* **34**, 589–599. <https://doi.org/10.1111/j.2044-8260.1995.tb01493.x> (1995).
55. Rumpf, H. J., Hapke, U., Meyer, C. & John, U. Screening for alcohol use disorders and at-risk drinking in the general population: Psychometric performance of three questionnaires. *Alcohol. Alcohol.* **37**, 261–268. <https://doi.org/10.1093/alcalc/37.3.261> (2002).
56. Lowe, B. *et al.* A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *J. Affect. Disord.* **122**, 86–95. <https://doi.org/10.1016/j.jad.2009.06.019> (2010).
57. Iliodromiti, S. *et al.* Variation in menopausal vasomotor symptoms outcomes in clinical trials: A systematic review. *BJOG* **127**, 320–333. <https://doi.org/10.1111/1471-0528.15990> (2020).
58. Sloan, J. A. *et al.* Methodologic lessons learned from hot flash studies. *J. Clin. Oncol.* **19**, 4280–4290. <https://doi.org/10.1200/JCO.2001.19.23.4280> (2001).
59. Guttuso, T. Gabapentin’s effects on hot flashes in postmenopausal women: A randomized controlled trial. *Obstet. Gynecol.* **101**, 337–345. [https://doi.org/10.1016/s0029-7844\(02\)02712-6](https://doi.org/10.1016/s0029-7844(02)02712-6) (2003).
60. Heinemann, K. *et al.* The Menopause Rating Scale (MRS) scale: A methodological review. *Health Qual. Life Outcomes* **2**, 45. <https://doi.org/10.1186/1477-7525-2-45> (2004).
61. Hauser, G. A., Schneider, H. P., Rosemeier, P. J. & Potthof, P. Die Selbstbeurteilungsskala für klimakterische Beschwerden (Menopause Rating Scale-II). *J. Menopause* **6**, 13–17 (1999).
62. Girod, I., de la Loge, C., Keininger, D. & Hunter, M. S. Development of a revised version of the Women’s Health Questionnaire. *Climacteric* **9**, 4–12. <https://doi.org/10.1080/13697130500487372> (2006).
63. Hunter, M. S. The Women’s Health Questionnaire (WHQ): Frequently asked questions (FAQ). *Health Qual. Life Outcomes* **1**, 41. <https://doi.org/10.1186/1477-7525-1-41> (2003).
64. Haanstra, T. M. *et al.* How do low back pain patients conceptualize their expectations regarding treatment? Content analysis of interviews. *Eur. Spine J.* **22**, 1986–1995. <https://doi.org/10.1007/s00586-013-2803-8> (2013).
65. Klein, E. M. *et al.* The German version of the Perceived Stress Scale—Psychometric characteristics in a representative German community sample. *BMC Psychiatry* **16**, 159. <https://doi.org/10.1186/s12888-016-0875-9> (2016).
66. Glaesmer, H., Hoyer, J., Klotsche, J. & Herzberg, P. Y. Die deutsche Version des Life-Orientations-Tests (LOT-R) zum dispositionellen Optimismus und Pessimismus. *Zeitschrift für Gesundheitspsychologie* **16**, 26–31. <https://doi.org/10.1026/0943-8149.16.1.26> (2008).
67. Hull, S. C. *et al.* Patients’ attitudes about the use of placebo treatments: Telephone survey. *BMJ* **347**, f3757. <https://doi.org/10.1136/bmj.f3757> (2013).
68. Ziller, V. *et al.* Adherence to adjuvant endocrine therapy in postmenopausal women with breast cancer. *Ann. Oncol.* **20**, 431–436. <https://doi.org/10.1093/annonc/mdn646> (2009).
69. Rief, W. *et al.* Assessing general side effects in clinical trials: Reference data from the general population. *Pharmacoeconom. Drug Saf.* **20**, 405–415. <https://doi.org/10.1002/pds.2067> (2011).
70. Stearns, V., Beebe, K. L., Iyengar, M. & Dube, E. Paroxetine controlled release in the treatment of menopausal hot flashes: A randomized controlled trial. *JAMA* **289**, 2827–2834. <https://doi.org/10.1001/jama.289.21.2827> (2003).
71. Girod, I., Abetz, L., de la Loge, C., Fayol-Paget, C. & Hunter, M. S. Women’s Health Questionnaire User Manual. 55 (MAPI Research Institute, Lyon, 2004).
72. Grady, D. *et al.* Is a shorter hot flash diary just as good as a 7-day diary?. *Menopause* **16**, 932–936. <https://doi.org/10.1097/gme.0b013e3181a4f558> (2009).
73. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* 2nd edn. (Lawrence Erlbaum Associates, New Jersey, 1988).

## Acknowledgements

This study was conducted with board funds from Yvonne Nestoriuc’s lab. Additionally, the study was supported by a grant to YN by the Foundation for the Science of the Therapeutic Encounter (F-STE). The funders played no role in the design of the study, analysis, and interpretation of data or in the decision to submit the manuscript for publication. The authors thank Victoria Sump, Twyla Michnevich, Miriam Lea Frank, Anna Zaplatnikova, Maria Lara Mehlhorn, Antonia Loeck, Anne Winkelmann and Maja Glahn for conducting assessments and helping with recruitment and administrative tasks.

## Author contributions

All authors made substantial contributions to the conception of the study, Y.P. acquired, analysed and interpreted the data, and drafted the manuscript. All authors revised and approved the final manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-77255-z>.

**Correspondence** and requests for materials should be addressed to Y.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020