PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Randomized controlled trial comparing hypnotherapy versus gabapentin for the treatment of hot flashes in breast cancer survivors: a pilot study
AUTHORS	MacLaughlan David, Shannon; Salzillo, Sandra; Bowe, Patrick; Scuncio, Sandra; Malit, Bridget; Raker, Christina; Gass, Jennifer; Granai, Cornelius; Dizon, Don

VERSION 1 - REVIEW

REVIEWER	Gary Elkins, Ph.D., ABPP Director, Clinical Psychology Program Director, Mind-Body Medicine Research Laboratory Department of Psychology and Neuroscience Baylor University
REVIEW RETURNED	15-May-2013

GENERAL COMMENTS	This pilot is a timely investigation of hypnotherapy compared to a well-established antidepressant in the treatment for hot flashes in breast-cancer survivors. Overall, this paper is well organized, logical, and well written. Though this study is admittedly underpowered and accrual required broad inclusionary criteria, this is well explained in discussion. A few issues that should be clearly addressed: *Introduction:* There is an error in the report of previous research of hypnotherapy for the treatment of hot flashes in breast cancer survivors (Elkins et al., 2008). The authors reported a 56% reduction in hot flash frequency whereas the study reports a 68% decrease in hot flash score. Additionally, the fact that this study had a sample of 60 participants and was well powered to detect

	significant differences should be mentioned. This paper does not include the most recently published large-scale randomized controlled trial of clinical hypnosis for treating hot flashes (*Elkins, G. R., Fisher, W. I., Johnson, A. K., Carpenter, J. S., & Keith, T. Z. (2013). Clinical hypnosis in the treatment of postmenopausal hot flashes: a randomized controlled trial. Menopause, 20(3), 291-298*.). In this randomized, single-blind, controlled clinical trial, 187 post-menopausal women were treated to 5 sessions of clinical hypnosis compared to a structured-attention control, finding a mean reduction of 74% of subjective hot flash frequency compared to a 17% reduction in control. Though this study included a post-menopausal sample that was not stratified for cancer-status, it provides effect size data that is very relevant to this study. To date, there has been no study which suggests treatment efficacy for hot flashes differs in regards to cancer-status. On page 5, lines 3-25, the paragraph begins suggesting that non-hormonal drug therapies are associated with side effects, but only the interference with the metabolism of tamoxifen is mentioned. In my experience, participants in hypnosis studies have suggested that they chose not to use antidepressant for other side-effects as well (e.g. sleeplessness, mood changes, etc.). There are a number of reasons why cancer-survivors might want to avoid pharmaceutical therapies and a mention of this seems appropriate here. There is no mention of the large placebo effect found in many hot flash trials and this seems an important oversight, given the effect sizes found in most pharmaceutical trials. *Patients and methods:* Page 7, line 46 'Computer-generated randomization' cannot be

	replicated without further detail. Page 7-8 There is no mention as to how the number of sessions was selected for this study. Previous studies have used 5 sessions (Elkins et al., 2007), finding long-term benefit in that number of sessions. Page 9 – Power analysis: If you are suggesting that the power is underpowered for a definitive study, why mention a goal of 60, and an actual accrual of 27? *Results:* It would seem, given the limited number of participants, that intention-to-treat analyses would be a logical addition as 3 (20%) of the gabapentin group withdrew from side-effects and 2 hypnosis participants had no end-point diaries. *Discussion:* Page 14, Line 20 – The Elkins et al. study is complete and published (see above). This study is recommended for publication with minor edits.

REVIEWER	Dr Reza Oskrochi Medical Statistics coordinator Oxford Brookes University Oxford UK
REVIEW RETURNED	24-May-2013

THE STUDY	The study design seems appropriate, but there are some problems as follow: 1- the dropout in control group is related to the treatment (side-effect) and hence is not independent from the research question of interest. 2- Why use Wilcoxon sign-rank test where they have compared

	independent samples. This test is neither appropriate nor powerful. 3- In the table one descriptive is given for all patients. I think it is more important to give this descriptive for those who actually participated.
RESULTS & CONCLUSIONS	It is a good research question but I suggest to consult a statistician to be able to analyse the data properly and show significant improvement which I believe exist. The tables need more explanation for reader, for example I believe the confidence interval is given in the brackets but no mention of it in the foot note

VERSION 1 – AUTHOR RESPONSE

Reviewer: Dr. Elkins,
Thank you for pointing out the error in the reported results from your trial. We have corrected this in the manuscript and have included the data from the most recent clinical trial.

“On page 5, lines 3-25, the paragraph begins suggesting that non-hormonal drug therapies are associated with side effects, but only the interference with the metabolism of tamoxifen is mentioned. In my experience, participants in hypnosis studies have suggested that they chose not to use antidepressant for other side-effects as well (e.g. sleeplessness, mood changes, etc.). There are a number of reasons why cancer-survivors might want to avoid pharmaceutical therapies and a mention of this seems appropriate here. “
We have made changes to the introduction to point out the issue of side effects associated with drugs. In addition, there is a thorough discussion of this issue as it relates to patients’ concerns over treatment options and clinical trial participation in the discussion section.

"There is no mention of the large placebo effect found in many hot flash trials and this seems an important oversight, given the effect sizes found in most pharmaceutical trials. "
We agree that the significant placebo effect reported in clinical trials is an issue unique to evaluating hot flashes as an endpoint. We report the data in the introduction, but were trying to avoid comparing treatment effects across clinical trials. The impact of the placebo effect is further addressed in the discussion section.

"Page 7, line 46 ‘Computer-generated randomization’ cannot be replicated without further detail. "
We have clarified in the manuscript.

"Page 7-8 There is no mention as to how the number of sessions was selected for this study. Previous studies have used 5 sessions (Elkins et al., 2007), finding long-term benefit in that number of sessions. "

It was our subjective experience prior to developing the study that our patients were seeing immediate benefit after a single hypnosis session. In addition, most drug studies that have shown clinical benefit have seen such benefit at the four week mark. We developed the standardized hypnotherapy protocol such that the therapist-directed intervention could be completed in the first four weeks. The protocol was based upon the practice of one of our authors, who had been providing successful therapy to patients in the Breast Health Center. In the interest of brevity, this information was not included in the manuscript.

"Page 9 – Power analysis: If you are suggesting that the power is underpowered for a definitive study, why mention a goal of 60, and an actual accrual of 27? "

One of our goals in this study was not only to produce pilot data to inform a larger study, but also to evaluate the feasibility of a trial comparing a CAM treatment to a standard/traditional treatment. Including the disparity between our goal and our reality highlights one of the inherent challenges in this kind of trial design.

"It would seem, given the limited number of participants, that intention-to-treat analyses would be a logical addition as 3 (20%) of the gabapentin group withdrew from side-effects and 2 hypnosis participants had no end-point diaries. "
Thank you for pointing this out. We did complete a modified intention to treat analysis with the data that were provided by patients, though this was not clear in our original manuscript. We have edited the paper accordingly. (Further discussion below addresses Dr. Oskrochi's similar concern).

Reviewer: Dr Oskrochi ,

"1- the dropout in control group is related to the treatment (side-effect) and hence is not independent from the research question of interest. "

We acknowledge that differential dropout raises the possibility of biased treatment effect estimates. The pilot trial is not large enough to allow for modeling attrition patterns or using multiple imputation to account for missing outcomes.

The primary analysis was a modified intention-to-treat limited to patients with primary outcome data. These evaluable patients were analyzed as randomized, regardless of treatment adherence and dose. We have added a sentence to the methods section to clarify this approach. The lack of primary outcome information for the remaining eligible patients precluded the use of a full intention-to-treat analysis. The small sample size limited the application of standard methods to handle attrition and intermittent missing responses, such as maximum likelihood-based regression, multiple imputation, and attrition modeling.

To address this issue, we have performed a sensitivity analysis assuming improvement for patients missing data and then assuming no improvement for these patients. The sample of patients with complete outcome data was divided at a median percent change at week 8 vs. baseline. This created two samples of "high" responders who showed a greater decrease in symptoms and severity and low responders who showed a lower decrease or no change. From each of these groups, a random sample with replacement was draw to serve as imputed outcome values for all eligible patients missing all outcome data. The data were reanalyzed by randomized group using the imputed "high" responder outcomes and the imputed "low" responder outcomes. None of the difference by time point were statistically significant. Under the assumption of higher response, the hypnotherapy group still showed greater improvement for number and severity. Only a minimal difference in quality of life was seen. Under the assumption of lower response, the hypnotherapy group had a greater response for number and quality of life, but a lesser response to severity. Although this analysis is based on few values and does not account for imputation variability, it suggests that the trend seen in the primary analysis may hold. However, looking across the tables at the extremes (gabapentin dropouts had greater improvement, hypnotherapy dropouts had less) indicates it could go either way.

Data from the sensitivity analyses have been uploaded as a supplemental file.

"2- Why use Wilcoxon sign-rank test where they have compared independent samples. This test is neither appropriate nor powerful. "

We used the Wilcoxon rank-sum test (Mann-Whitney-U test) for testing for location shift between two independent samples. We did not use the sign-rank test for paired data. To clarify, we have added

“Wilcoxon rank-sum/Mann-Whitney-U test” in the methods section and table footnotes. We selected to use this nonparametric test instead of a two-sample T-test or parametric regression for longitudinal data for several reasons. First, the small sample size in each study group and missing response trajectories for the primary outcome limited the application of maximum likelihood-based approaches that would take advantage of the longitudinal data and interpolate missing response values. Second, graphical inspection indicated deviation from normality for the outcome variables: frequency, severity, and quality-of-life scores. With small sample sizes, the application of the Central Limit Theorem was questionable. In addition, the observed skewness of the data suggested that the Wilcoxon rank-sum test may be just as powerful as a parametric test.

"3- In the table one descriptive is given for all patients. I think it is more important to give this descriptive for those who actually participated. "

Thank you for pointing out this issue. The descriptive data reported actually are for the women who participated, and we have edited Table 1 to make this more clear.

"It is a good research question but I suggest to consult a statistician to be able to analyse the data properly and show significant improvement which I believe exist. "

We also believe that a difference exists, but unfortunately we do not have the power to prove it. One of our co-authors, Dr. Christina Raker, is a statistician. She performed the analyses reported and assisted with our revisions and clarifications outlined here.

"The tables need more explanation for reader, for example I believe the confidence interval is given in the brackets but no mention of it in the foot note."

We have edited each of the tables for consistency and clarification.

VERSION 2 – REVIEW

REVIEWER	Dr. G. Reza Oskrochi Senior Lecturer in Statistics MSc Medical Statistics Programme Coordinator Faculty of Technology, Design and Environment Oxford Brookes University
REVIEW RETURNED	02-Aug-2013

THE STUDY	This is a pilot study and nothing more than that; therefore the small number of patients cannot be representative of actual patients. Due to very small sample size the statistical analysis have no power and all are inconclusive. The statistical methods are named but not described. There is a potential allocation bias, but again not verifiable due to small sample size.
RESULTS & CONCLUSIONS	The result did not answer the research question due to small sample size. The study failed to show any significant differences between treatments. I think the employed method is OK but inconclusive, however this might be acceptable as a pilot study only to encourage a proper investigation for this interesting research question.
GENERAL COMMENTS	This is an interesting research question. The study might be acceptable as a pilot study only to encourage a proper investigation for this interesting research question.