Author's response to reviews

Title: Randomized Trial of Tapas Acupressure Technique for Weight Loss Maintenance

Authors:

Charles R Elder (Charles.Elder@kpchr.org)
Christina M Gullion (Christina.Gullion@kpchr.org)
Lynn L DeBar (Lynn.DeBar@kpchr.org)
Kristine L Funk (Kristine.Funk@kpchr.org)
Nangel M Lindberg (Nangel.Lindberg@kpchr.org)
Cheryl Ritenbaugh (Ritenbau@email.arizona.edu)
Gayle Meltesen (Gayle.Meltesen@kpchr.org)
Cherri Gallison (Cherri.Healingtouch@gmail.com)
Victor J Stevens (Victor.Stevens@kpchr.org)

Version: 2 Date: 20 January 2012

Author's response to reviews: see over
Dear Editors:

Thank you for the opportunity to resubmit a revision of our paper, “Randomized Trial of Tapas Acupressure Technique for Weight Loss Maintenance,” to BMC CAM. We appreciate the time, attention, and effort of the reviewers, as well as their constructive criticism. The review team has made a number of excellent suggestions, enabling us to significantly improve the manuscript.

In the revised manuscript, we respond to the reviewers’ suggestions, taking several approaches. The thrust of reviewer comments was, as stated by Reviewer #1, that “the interpretation of the findings and the conclusion seems wrong. The primary analysis seems to generate a negative result which the authors manage to present as though it was positive.” In response, we have emphasized and prominently highlighted in the abstract, discussion, and conclusion sections that our primary analysis gave a negative result, while clarifying the exploratory nature of the secondary and post hoc analyses. We have also added several paragraphs to the discussion section emphasizing the limitations of our analysis and conclusions, as guided and specified by reviewer comments. In addition, we have made multiple changes, both substantive and stylistic, in the abstract, conclusions, and indeed throughout the text toward setting a more “tentative” tone to the report, guided by reviewer comments, as suggested by Reviewer #2.

In addition, reviewer comments highlight that we need to do more to help readers better identify the background and rationale for our secondary and post hoc analyses. Our primary analysis approach was analysis of covariance (ANCOVA). The use of a single estimate of covariate effect in both treatment groups (TAT vs. SS) is warranted only when there is no interaction between treatment arm and that covariate [1,2] When a significant interaction is present, i.e., the covariate has different effects in the two treatment groups, the implication is that the treatment effect varies over the range of the covariate, and the test on the main effect term cannot be interpreted correctly without accounting for the difference in the covariate. Instead, the significant interaction (in our case, treatment group assignment by pre-randomization weight loss) suggests a treatment (TAT) effect that is dependent upon the value of the interacting covariate (initial weight loss), and post hoc testing is indicated toward identifying what the treatment (TAT - SS) effect is at various levels of that variable (initial weight loss).

We have revised and added text throughout to help readers distinguish that our post hoc analysis represents not a subset analysis but rather the predicted treatment effect calculated using the parameter estimates in the final model at selected values of the covariate. We agree with the reviewers that these analyses are exploratory, and must be interpreted with caution.

Thank you again for considering this revised manuscript, and we look forward to your review of and response to our revised submission.

Sincerely yours,

Charles Elder MD MPH FACP
RESPONSE TO REVIEWERS

Below we address each of the reviewers’ comments, point by point. Revisions within the original manuscript are highlighted in yellow.

Reviewer # 1

“The interpretation of the findings and the conclusion seems wrong. The primary analysis seems to generate a negative result which the authors manage to present as though it was positive.”

Response: We have emphasized and prominently highlighted in the abstract, discussion, and conclusion sections that our primary analysis gave a negative result, while clarifying the exploratory nature of the secondary and post hoc analyses.

“The discussion needs more self critical input”

Response: The discussion section is substantially modified to address limitations of the analysis and findings.

“The trial should be reported according to consort guidelines.”

Response: A completed consort checklist is included with the re-submission, and appropriate additions to the paper are inserted.

Reviewer #2

Reviewer's report:

Minor essential Revisions:

1. This article has a negative result relative to its original hypothesis. The positive aspects of this study were found in exploratory secondary analyses confined to 20% of the original cohort (28 in the TAT group and 28 in the SS group). In view of this, I think it is important that the conclusions be stated in more tentative terms such as the secondary analysis suggests the need for further studies looking at those that lose above a certain percent of baseline body weight for confirmation that TAT maintains weight loss better than social support in that subgroup.

Response: The conclusion has been revised in keeping with the reviewer’s suggestions.
2. This reader questions the significance of a 1.25kg difference in weight maintenance over the course of a year. If this amount of weight difference has clinical relevance, it would be helpful for the authors to elaborate on that relevance.

**Response:** We have added to the discussion section an assessment of the clinical significance of these weight loss differences.

Level of interest: An article of limited interest

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I have consulted for companies making pharmaceuticals, dietary herbal supplements and medical foods, but none of these relationships represent a competing interest in reviewing this manuscript.

Reviewer # 3

This rigorously designed, well written paper presents findings from a relatively large 2-arm RCT that compared a professionally facilitated energy psychology group-based intervention (Tapas Acupuressure Technique, TAT) against a professionally facilitated support group with the goal of maintaining weight loss following a 6-month behavioral weight loss program. Whereas the study design and size of the trial were strengths, 11 of the 19 items on the CONSORT checklist for non-pharmacological (behavioral) trials were either missing and/or unclear, thus compromising the quality of the report as currently written.

**Response:** These items have been added, and an updated CONSORT checklist provided.

As detailed below, my major concerns are that the primary conclusions do not follow from the data, and there are several alternative explanations and confounding factors that merit discussion in interpreting these results, including clinical implications for health care (if any).

**Response:** We have emphasized and clarified in the abstract, discussion, and conclusion sections that our primary analysis gave a negative result. We have introduced edits emphasizing the exploratory nature of our secondary and post hoc analyses, and have added text to both the discussion and conclusion sections emphasizing study limitations.
Discretionary Revisions

1. References: #6-9 are not current. For example, see Tapper et al. 2009 Appetite, and other RCTs published since 2005.

   **Response:** The references have been updated as suggested.

Minor Essential Revisions

2. Title: To help identify the specific objective of the trial, use the term “TAT” rather than the more vague “Mind and Body Technique” in the title.

   **Response:** This has been changed as suggested.


   **Response:** This has been corrected as suggested, thank you.

4. Table 2: Change title from Regression Models to ANCOVA Models. Add effect sizes and confidence intervals.

   **Response:** We have added standard errors of estimate and confidence limits for the parameter estimates in the final model.

5. Table 3: Add effect sizes and confidence intervals.

   **Response:** This has been changed as suggested.

6. Background: In my view, the Background section lacks a compelling theoretical or empirical basis for the TAT intervention. I read the authors’ prior publications on this topic (2007 and 2009, J Altern Complement Ther), and I encourage elaborating on the theoretical mechanisms of how TAT may work in this paper, as well. This can be a couple of additional sentences in 4th paragraph.

   **Response:** This has been added as suggested.

7. Background: My review of the authors’ prior publications also revealed that there are numerous potentially active ingredients of TAT, few of which are discussed in interpreting the results of this trial. For example, TAT not only involves acupressure (which could ostensibly influence or unblock chi, should chi
exist), but it also involves several other therapeutic elements that could conceivably help reduce stress, negative mood, sleep disturbance, quality of life, and ultimately, weight, in their own right. These techniques include visualization, asking for forgiveness, thanking God (or some other higher power) for the requested/expected healing response (a call for religious or spiritual healing of sorts), and drinking 6-8 glasses of water on the days TAT is practiced (which could help maintain a sense of fullness during the day, thereby mitigating excessive eating, and in turn helping to maintain weight loss). Although I can respectfully consider the Chinese medicine/chi/meridian theory of how TAT purportedly works, any or all of these constituent processes could very well explain any potential therapeutic effects of TAT on stress-related mental or physical health conditions. The contribution of the paper would be strengthened (more balanced and even handed, and less potentially biased and driven by non-significant results) by presenting this issue – and these alternative explanations or theories – in both the Background and Discussion.

Response: Review of this issue has been added to the background and discussion sections.

Major Compulsory Revisions

1. Abstract, Results: Primary statistical test of efficacy is misinterpreted. There is no significant difference between groups. Please state as such. Do not say “arms differed by 1.24 kg” and “more weight loss was maintained in the TAT group” when these primary statements about efficacy are not supported by the data (p=.097). The way the primary result is currently written is fundamentally misleading and misrepresents the major trial outcome.

Response: We have changed the description of the results from the primary outcome model in keeping with the reviewer’s suggestion.

2. Abstract, Conclusions: The main conclusion does not accurately reflect the test of the primary or secondary hypotheses, but rather focuses on the results of a secondary, subgroup analysis based on questionable methodology. An equally plausible interpretation of the trial data, from both the R21 and the R01 grants, is that TAT is simply no more effective in maintaining weight loss than facilitated peer group support. In my view, the findings, and health care implications, would be more accurately presented as such. See additional comments/concerns below about multiple confounding factors that could alternatively explain TAT effect.

Response: Abstract and conclusions have been substantially revised. We have emphasized and highlighted in the abstract, discussion, and conclusion sections that our primary analysis gave a negative result. Also we clarify that the post hoc analysis of the significant interaction was not a subgroup analysis but estimation of least squares means conditional on selected values of the covariate in the significant interaction.
3. Methods, Weight-loss maintenance interventions: The issue of clustering (i.e., patients are clustered within group facilitators) needs to be addressed, methodologically and statistically, per CONSORT guidelines. The current statistical model is misspecified without taking clustering [aka nesting, or hierarchical data] into account, which can produce biased estimates of treatment effects.

Response: We did not account for clustering in our analysis, as this was not part of our pre-specified analysis plan, and we recognize now that the study was not adequately powered to do so. While technically the reviewer is correct, this was not state of the art for weight management trials when we wrote our grant proposal 7 years ago. Accounting for clustering in NIH funded trials may not be feasible due to the very large increases in sample size and trial costs that would be required. We were neither powered for, nor were asked to account for, clustering in any of the other landmark weight management studies published from our institution, including WLM[3] and PREMIER.[4]

We have added text addressing this limitation in the discussion section of the manuscript.

4. Methods, Weight-loss interventions, TAT: Briefly describe what is involved in TAT certification (# of hours of training, # of hours of supervision, # of TAT sessions performed, and verified outcomes obtained [if possible]).

Response: This has been added as suggested.

5. Methods, Weight-loss interventions, SS: Briefly describe degree of experience, training, and qualifications of SS group facilitators.

Response: This has been added as suggested.

6. Methods: Please add section on how fidelity to TAT protocol was assessed during the trial. Method of TAT quality assurance (adherence to protocol)? Consistent quality or delivery across interventionists?

Response: This has been added as suggested.

7. Methods: The 2009 JACM paper mentions that an expectancy to benefit scale would be administered to both the TAT and the SS groups, however, this is not described in the current manuscript. Please add and discuss relevance, since non-specific effects (like expectancy) are known to largely explain many CAM intervention effects, including acupuncture and hypnosis among others.

Response: There is no mention of an expectancy to benefit scale in our 2010 design paper. We did in fact include a measure of expectancy in our 2007 publication describing results of the pilot study. However, as published in that paper, we found that expectancy did not significantly moderate weight change in the CAM as compared to the self
directed support interventions. We thus did not measure expectancy in this study, but acknowledge that such phenomena can play a role. Text commenting on this limitation has been added to the discussion section.

8. Methods, Statistical methods: No statistical power analysis or sample size determination is presented in either the 2009 design/rationale paper, or in this manuscript. Please add, and frame in the context of your specified models. In this description, please address statistical power (what it was & how it was calculated) for the augmented model, which includes interaction terms that require greater power than main effects.

Response: This has been added as suggested.

9. Methods, statistical methods: Alpha levels are described for main effects (.05) and interactions (.10), however, many statistical tests were performed and several data-driven model modifications were made without discussing the issue of multiple testing. Please address the multiple testing issue, including the role of taking a data-driven versus a theory-driven approach to testing your augmented and final models. How does multiple testing and data-driven analytic approach influence the reliability (or spuriousness) of the trial findings? This is an important topic to address in both the Methods and Discussion/Limitations.

Response: For the primary analysis, there is a single main outcome test, i.e., TAT vs. SS effect on weight change at 12 months, adjusted with four planned covariates.

Because ANCOVA strongly assumes no interaction between treatment and covariates used to adjust the treatment effect, in our planned secondary analysis, before adding additional hypothesized covariates to the “initial” outcome model, we tested the four terms for treatment-by-covariate interaction to the model. This check on validity of assumptions has the same essential standing before interpreting results as checking whether regressions are linear and within-group variances are homogeneous. Moreover, because ANCOVA strongly assumes no interactions, the main focus in the case of assessing for interactions was in avoiding a type II, as opposed to type I, error. Finally, in our initial regression model, pre-randomization weight loss was not a significant predictor of weight regain (p=.46, see Table 2). This is a seemingly surprising finding, since initial weight loss has been a strong predictor of weight regain in previous studies. The finding is a consequence of averaging parameters with opposite slopes, as shown by the description of the treatment assignment by pre-randomization weight loss interaction in the post hoc analysis (greater initial weight loss associated with more weight regain for SS but less weight regain for TAT, so the 2 effects wash out).

The significant treatment assignment by pre-randomization weight loss interaction suggests that the TAT effect depends upon the level of initial weight loss. In our post hoc analyses as shown in Table 3, we describe what the TAT effect is at various levels of initial weight loss. We are not performing multiple sub-group tests, nor claiming an effect for a sub-group, but rather the analysis is intended to illustrate the varying
treatment effects over the selected values of the covariate. We have extensively edited

text in the methods, results, and discussion sections toward clarifying that what is
different about the two groups is the trajectory of weight results across pre-randomization
weight loss values.

I agree that our procedure for adding covariates (education, age at entry, 0-M score on
PHQ-8, ISI, and PSS, and percent attendance at sessions during the active 6-month
weight loss-maintenance intervention) to the augmented model toward arriving at the
final model involves multiple tests. We have clarified in the text that these analyses are
exploratory and not confirmatory, and have added comment to the discussion section
specifying this limitation.

10. Methods, statistical methods: Per CONSORT guidelines, please add
estimates of effect size and confidence intervals for ANCOVA and regression
derived parameter estimates.

Response: These are added as suggested.

11. Results, Change in weight: The description of results must be consistent with
statistical tests. Please correct misleading use of language here when describing
tests of main effects. It is not accurate to say “…a difference favoring the TAT
arm…” when the p-value is not significant. It is accurate to simply say “change in
weight did not significantly differ between the TAT and SS groups (p=.098).” The
first line of the Discussion, for example, does a good job in this regard.

Response: We have implemented the change as suggested.

12. Results, Secondary outcomes: First, please show results of secondary
outcomes in a table, even though they are not statistically significant. Second,
given numerous dramatic case reports and anecdotes, and the purported theory
of how TAT unblocks chi, why were none of the secondary outcomes significantly
improved in the TAT group? (Topic for Discussion)

Response: The secondary outcomes table, as well as additional relevant discussion, have
been added as suggested.

13. Results, Exploratory analyses: Why wasn’t TAT practice significantly
associated with weight change outcomes? (Again, good topic of Discussion,
given theory and presumed necessity of self-practice to elicit a healing response
with TAT.)

Response: Practitioners of energy psychology interventions maintain that symptoms or illness can
sometimes be caused by suppressed emotional trauma, and that through practice of the
energy psychology technique, such trauma can be healed, or definitively resolved. Viewed through this lens, more practice is not necessarily better.

This issue is now reviewed in additional text added to the Discussion section.

14. Discussion: The result of the secondary, subgroup analysis – despite questionable subgroup splitting by 20/80%ile – is interesting from the perspective of “what works best for whom” with CAM therapies. I would strongly suggest doing a sensitivity analysis to more closely examine the robustness of this potential secondary/exploratory finding. For instance, I would suggest repeating the analysis you did but using different sets of %iles to compare (e.g., top 10% vs. bottom 90%; top 25% vs. bottom 75%; top 33% vs. bottom 67%, top 50% vs. bottom 50%). If you can create a sensitivity analysis table to substantiate and build a greater degree of confidence around the 20/80%ile analysis, it would strengthen the contribution of this paper.

We did not do a subgroup analysis, and the 20/80% “subgroup splitting” actually represents 2 of the series of values on the covariate that we arbitrarily selected for calculating the model-predicted treatment effect in the final model. We have edited and revised the text throughout toward clarifying that we are not performing multiple sub-group tests, nor claiming an effect for a sub-group, but rather describing the differing trajectories of weight results across pre-randomization weight loss values for the two arms. In the revised text the 20th percentile of initial weight loss is mentioned only once (Line 398) and is clearly only an example to guide interpretation of Table 3.

15. Discussion, 2nd paragraph: Please do not present new data in the Discussion. Figure 4 data should be moved to Results, and only interpretation discussed in Discussion section.

Response: Figure 4 data are in fact presented in the results section, and then expanded upon in the discussion section.

16. Discussion, 3rd paragraph: Greater initial weight loss is confounded by younger age, fewer depressive symptoms, and less perceived stress. These factors alone could explain greater weight loss, and subsequent between-group differences in weight regain, irrespective of TAT. Please add this point to a general, more even handed discussion of alternative explanations for this trial’s null effects and secondary/exploratory findings.

Response: This has been addressed in paragraphs added to the discussion section.

17. Discussion, 3rd paragraph, last sentence: Please give at least one specific example of how a future trial could be designed to capitalize on the lessons learned from this trial.
Response: The paragraph in question has been eliminated altogether. We felt that this paragraph was distracting, and was strongly contributing to a misconception that we are focused on a subset analysis, and/or on demonstrating an effect in a subset.

18. Conclusions: Again, please lead with statement of primary (null) trial finding, then present secondary/exploratory findings. Please emphasize how your conclusions may be confounded/biased/spurious/unreliable given (a) the number of statistical tests performed (b) unclear statistical power, (c) a multitude of alternative explanations for how TAT could exert a therapeutic effect, and (d) the fact that two NCCAM-funded RCTs have now documented that TAT is not superior to group support in maintaining weight loss.

Response: The conclusion has been revised as suggested.

Reviewer #4

Reviewer's report:

This study investigated the effects of two weight loss maintenance (WLM) interventions following a weight loss trial. It is a substantial trial in terms of numbers of participants, is well designed and executed.

The primary concern is the reluctance of the investigators to accept that there is really no statistical or meaningful difference between the interventions.

Response: We have emphasized and prominently highlighted in the abstract, discussion, and conclusion sections that our primary analysis gave a negative result.

It is also not clear whether either of the programs should be regarded as successful in terms of maintenance of lost weight.

Response: Both groups did well, with results favorable compared to interventions in other weight loss maintenance studies. Text has been added to the discussion section to clarify this point.

Major Compulsory Revisions

Abstract

P2, line 52 – there is some confusion here and in the main document regarding the time frames of the WL and WLM phases. In the abstract, recommend the following: ‘lost weight in a prior behavioural weight loss program’.

Response: We have made this change as suggested.
P2, line 60 – use of the term randomization as a time point – it would be clearer to state this as the start of WLM; also state that the WLM interventions had a 6 month active stage and a further 6 month follow up.

Response: We have made this change as suggested.

P2, lines 64-67 – report mean ±SD for age, BMI and weight loss. I disagree with the statement ‘arms differed by 1.24 kg’ – rather the statement should read ‘the arms did not differ (p<0.097)’.

Response: We have made these changes as suggested.

P2, line 68-69 – there is no sense of how many participants are in the group with the most initial weight loss or the extent of that weight loss. Is this an artefact of multiple comparisons?

Response: For the revised manuscript the description of results in the abstract for the post hoc tests is now limited to “showing that greater initial weight loss was associated with more weight regain for SS but less weight regain for TAT.” This will hopefully help in clarifying that what of interest is the different trajectories of the weight regain results across initial weight loss values, as opposed to a magnitude of weight change difference in a particular subset.

The issue of multiple comparisons is addressed previously in the response to Reviewer #3.

P2, lines 72-74 – a conclusion about the effectiveness of either WLM interventions is required

Response: We report in the abstract both initial weight loss, as well as weight regain for both groups at 12 months post randomization. A paragraph discussing the overall weight change for participants in our study compared to others in the literature has been added to the discussion section, but is difficult to succinctly summarize for the abstract.

Methods

P5, lines 134-147 – recommend calling this section Study Design. Start with the participants in the WL phase and use terms like phase to distinguish the WL and WLM phases. Recommend referring to the start of WLM as WLM-0 or something rather than RAND; this would make it clear what that the 6 and 12 months assessments reported relate to this time 0.

Response: We have changed RAND to “0-M” as suggested throughout.
P7, lines 185-189 – more details regarding the randomization process are required.

**Response:** This information has been added.

Also, what WL stratification categories were used? Was there no consideration of sex or age?

**Response:** We balanced entry into the two treatment arms by stratifying on weight loss program (WLP) group, with permuted blocks of varying size. There was no consideration of sex or age at randomization, though these parameters were adjusted for in the analyses, as described in the text.

P9, lines 244 – statistical methods – Was a sample size/power calculation undertaken? Was the study sufficiently powered?

**Response:** This information has been added to the text.

P9, line 255 – the critical values for statistical significance were not adhered to in the reporting of results.

**Response:** We have rephrased our reporting as suggested.

Results

Table 1 – BMI category – states that the units are kg but presume are %

**Response:** Columns 2-4 are in percent, as indicated above those columns. The units of kg mentioned in column one refer to the delineation of BMI categories in that column.

Weight regain – the mean weight regain in each WLM group is modest, suggesting that both might be regarded as effective?

**Response:** Correct, both groups did well. Text has been added in the Discussion to this effect.

The wording needs to make it clear that there were not significant differences when this was the case.

**Response:** We have rephrased our reporting as suggested.

Better ways to present the more meaningful findings of the study should be sought.
Response: We have added text help readers better identify the background and rationale for interpreting the data

Discussion

P14 lines 352-359 – the statements here clearly reflect the true outcomes of the study and should be the focus of any revised manuscript.

While it is clearly disappointing to the investigators that the TAT intervention was not more effective, the tendency to overstate a possible random finding (that those who lost more weight initially benefited more from the TAT WLM intervention) suggests serious bias in interpretation.

Response: For the revised manuscript, we have emphasized and prominently highlighted in the abstract, discussion, and conclusion sections that our primary analysis gave a negative result, while clarifying the exploratory nature of the secondary and post hoc analyses.

Conclusions

Overstated. Revise to be consistent with opening paragraph of discussion.

Response: We have introduced additional text toward a more understated tone as suggested.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

Reference List


