Author’s response to reviews

Title: Does different information disclosure on placebo control affect blinding and trial outcomes? A case study of participant information leaflets of randomized placebo-controlled trials of acupuncture

Authors:

Soyeon Cheon (lizwarrr@gmail.com)

Hi-Joon Park (acufind@khu.ac.kr)

Younbyoung Chae (ybchae@khu.ac.kr)

HYANGSOOK LEE (erc633@khu.ac.kr)

Version: 1 Date: 30 Nov 2017

Author’s response to reviews:

Reviewer #1 (Heejung Bang):

1. What is the authors' final message? I personally think some trick is acceptable or even called for as long as meeting requirements on ethics.

ANSWER> We agree on the reviewer’s point that some tricks may be necessary as far as such practice is ethically legitimate. While our study taking an example of placebo-controlled trials of acupuncture suggests that information on placebo in PILs may affect blinding and outcomes thus routinely accepted placebo information disclosure in PILs should be revisited, we chose to cautiously call for more research on this issue in other medical disciplines before voicing a rather hasty conclusion based on our single study results. Nevertheless, to be more specific in our final message in the abstract (P. 3) and conclusion part (P. 17), we revised our conclusion as follows:

Original: (P. 3) [As we have few empirical findings on this issue, future studies are needed to determine whether the present findings are relevant to other medical disciplines.]

Revised: As we have few empirical findings on this issue, future studies are needed to determine whether the present findings are relevant to other medical disciplines and at the same time a routine practice of fully disclosing placebo information in PILs calls for a reevaluation.

Original: (P.17) [Placebo information disclosed to trial participants in PILs was found to possibly affect blinding and trial outcomes. How participants are told about placebos may be another potential factor that may influence participant blinding and the study outcome by possibly modulating patient expectation. As we have few empirical findings on this issue, future studies are needed to determine whether the present findings are relevant to other medical disciplines.]
Revised: How participants are told about placebos may be another potential factor that may influence participant blinding and the study outcome. As we have few empirical findings on this issue, future studies are needed to determine whether the present findings are relevant to other medical disciplines and at the same time a routine practice of fully disclosing placebo information in PILs calls for a reevaluation.

2. One author generously provided 2 PILs. Are they for different studies, not just different versions for the same study? Also, these studies are indicated in Supplement?

ANSWER> One author generously provided two extra PILs from two different studies that were not in our original invitation list (160 studies invited). Hence, we indicated “2 studies added by contacted author” in Figure 1, P. 24. Both studies are also included in Table 1 and Appendix 3.

3. Editorial questions:

1) In Abstract, "..PILs which…” is grammatically correct?

ANSWER> We have deleted ‘which’ and corrected the sentence.

2) P. 8: Authors may want to state how "test for subgroup differences" was done?

ANSWER> To be clear about the subgroup analysis, we revised our description as follows:

Original: [The I2 test was used to examine heterogeneity across the studies, and different PIL categories were subject to the subgroup analysis. In addition, a post hoc analysis grouped studies by its originating region (Asia and Non-Asia) in each PIL category and tested for subgroup differences applying the same method.]

Revised: The I2 test was used to examine heterogeneity across the studies, and different PIL categories were subject to the subgroup analysis to investigate whether the estimates in the primary outcome varied between subgroups (FD or DD). In this meta-analysis using a Review Manager software, this test was undertaken using Cochran’s Q test and Higgins’ I2 by comparing the subtotal estimates between the subgroups. In addition, a post hoc analysis grouped studies by its originating region (Asia and non-Asia) in each PIL category and tested in the same manner above whether the estimates in the outcome of interest varied in different regions.

We also added two forest plots in the Appendix 6 to help understand additional subgroup analysis results based on the study origin (Asia and non-Asian countries) for each category of FD and DD.
3) P. 10, middle: "When the number….. 1 to 1 and 2 to 1…" is confusing; authors may clarify.

**ANSWER**> For clarity, we revised this sentence as follows:

Original: [When the number of studies in ideal or possibly problematic blinding scenarios was compared within each category, the ratio was 1 to 1 in FD category, and 2 to 1 in DD/MI category, revealing a higher rate of ideal blinding in the latter (Table 1).]

Revised: When the number of studies that belonged to ideal or possibly problematic blinding scenarios was compared within each category, DD/MI category had a higher rate of studies with ideal blinding than FD category (Table 1). The number of studies in S1 and S5, the ideal scenarios, was more than double the number of studies in S4 and S6 in DD/MI category (7 studies vs. 3 studies), whereas the numbers were the same in FD category (5 studies vs. 5 studies).

4) P. 11: t-test for real and MW test for placebo?

**ANSWER**> Each test was selected based on the data distribution by Shapiro-Wilk test, as we have explained in P. 7, the methods section.

To elaborate, in the placebo acupuncture group, the BI values of studies from FD group did not meet the assumption of normality (p = 0.042, Shapiro-Wilk test), whereas BI values of real acupuncture in both FD and DD/MI groups did follow normal distribution (p = 0.512 and p = 0.940, respectively, Shapiro-Wilk test). Therefore, we ran Mann-Whitney U test for placebo acupuncture group and independent t-test for real acupuncture group.

5) P. 12: In "When BI values… significance", English is long and hard to follow.

**ANSWER**> We revised this sentence as follows:

Original: [When BI values of the included studies were compared between FD and no disclosure (DD/MI) categories, placebo groups in the studies with no disclosure were more likely to make opposite guesses on the intervention which may reflect more wishful thinking than those with FD, but without statistical significance.]

Revised: When BI values of the included studies were compared between FD and no disclosure (DD/MI) categories, placebo groups in the latter were more likely than those in FD category to make opposite guesses on the intervention, i.e., guessed that they have received real acupuncture.
Although this difference did not reach statistical significance, it may reflect more wishful thinking of placebo groups in no disclosure category than those in FD category.

6) P. 13: "in large"?

ANSWER> We revised our sentence as follows:

Original: […, information disclosure in PIL may be taken as an issue in clinical research in large.]

Revised: …, information disclosure in PIL may be taken as an issue in clinical research across the board.

7) P. 14: line -4: "limits"?

ANSWER> For clarity, we revised the sentence as follows:

Original: [Nonetheless, the fact that only small number of studies were included here and that this is merely observational in nature limits a broader interpretation.]

Revised: Nonetheless, the fact that only small number of studies were included here and that this is merely observational in nature allows for an interpretation in a narrow context.

8) P. 18: What is "reasonable" request?

ANSWER> By “reasonable” request, we meant cases such as (collaborative) research purpose. Any request can be acceptable except one where anonymity of the included studies cannot be preserved.

9) In table 1's footnote, it may be indicated: "Unblinded" generally means "More correct guess" as unblinded is a simple but strong term.

ANSWER> Thank you for your suggestion. Incorporating your suggestion, we added a sentence in the footnote as follows: Here, the term “Unblinded” generally means "More correct guess," not broken blinding literally.

Additionally, in the Methods, we added another sentence to clarify this notion throughout the paper:
It should be noted that throughout this paper, the term “unblinded” generally means "more correct guess," not broken blinding literally.

We appreciate reviewer #1’s valuable comments and suggestions which helped us deliver our key message in our manuscript more clearly.

Reviewer #2 (Valerie Durkalski-Mauldin):

1. The authors should consider different terminology than 'deceitful disclosure'; it should not be implied that investigators intended to be 'deceitful' in the PIL.

ANSWER> We fully understand the reviewer’s concern which we ourselves also had at the start of the analysis process. We would like to explain how and why we adopted this term. Although terms such as ‘incomplete’, ‘partial’, ‘disingenuous’, ‘inexplicit’ or ‘dishonest’ were suggested within our team, we finally decided on ‘deceitful disclosure’ because the term ‘deceptive disclosure’ for PILs where a placebo acupuncture is introduced as a different type of acupuncture treatment, was actually used in Miller’s articles which are directly relevant to our study , . Our PubMed search could not locate a study that provided a specific term for this ‘DD’ category - we may have missed relevant studies, though. It also shows that few studies investigated this issue, i.e., disclosure of information on placebo in PILs.

As the reviewer pointed out, we also do not believe that investigators intended to be ‘deceitful’ in the PIL. However, even if the trialists of the studies in DD category never intended to be deceitful and did not use specific terms such as sham, placebo, or fake due to different reasons and rationale, degree of information disclosure is clearly different from that of full disclosure category; after all, the participants were not explicitly informed of the existence of a placebo control in the studies that belonged to DD category.

To address reviewer’s concern that we too perceived, the following was added in P. 6; PILs that used neutral or deceitful words like ‘group two’, or simply the ‘control group’ were put into DD category even though we trusted that the study authors never intended to be ‘deceitful’ on purpose in their PILs.

2. Page 10, line 26: You should explain to the audience why Scenario 5 in table 1 is an 'ideal' blinding situation even though the acupuncture group was 'unblinded'.

ANSWER> As is mentioned in response to Reviewer #1’s question no. 9), the term “unblinded" generally means "more correct guess" in our study. That being noted, Scenario 5 is considered ‘ideal’ because acupuncture group guessing they are on real acupuncture (unblinded or more correct guess), and placebo group also guessing they are on real acupuncture (opposite guess)
may reflect participants’ wish to receive active intervention regardless of actual treatment received.

To clarify this, we added the following explanation in Table 1:

S5 Ideal – patients tend to have wishful thinking, strong placebo effect, and any treatment administered is perceived as real treatment.

We also put “Here, the term “Unblinded” generally means “More correct guess,” not broken blinding literally” in the footnote of Table 1.

3. Page 10, line 43: It appears that Figure 2 illustrates the calculated BIs for each included trial. Is the quoted 42% AND 41% in the text, the overall BIs for each group? Or is this simply the proportion of correct guesses. This should be clarified to the reader.

ANSWER> We clarified this by adding another sentence as follows:

Average BI values for real acupuncture groups in FD and DD/MI categories were 0.42 and 0.41, respectively.

4. Page 13, line 4: Of these half, how many were the older studies that were included from the previous review? If you are trying to show a change over time, it would be helpful to know this piece of information.

ANSWER> We agree on this point and thus added the following explanation:

… being more than a half: out of 70 placebo groups from 65 PILs, 40 were classified into FD category, and 12 of those 40 were from the older studies that were included in the previous review.

5. Page 13, line 38-50 and Page 14, line 4-9: I would be hesitant to make this assumption. It is not a given that if you sampled more studies, that you would see a link between the PIL information and direction of blinding. Your results did not support this based on the confidence intervals of the proportions (16% vs 21%).

ANSWER> This is a valid point and we mitigated the sentence as follows:

Original: [Given our BI analysis and the results from other relevant studies, it may be carefully assumed that with a larger number of studies, the present finding could be more solid to show the
link between the information in a PIL and direction of blinding in randomized placebo-controlled trials.]

Revised: Given our BI analysis and the results from other relevant studies, the hypothesis can be generated that full disclosure of placebo in PILs may undermine successful blinding of participants and this should be systematically evaluated in the future studies.

6. Page 14; line 35: The term 'investigator bias' should be expanding for the setting of acupuncture studies; I am assuming the investigator is unblinded in this setting (they know if the patient is receiving placebo or real acupuncture?). If that is the case, do the authors know if the outcomes were completed by a blinded assessor? That would be important to know to help explain the level of investigator bias.

ANSWER> Thank you for bringing up this important point. We added sentences as follows:

Original: [Another explanation is that the treatment effect in DD category is rather overestimated due to investigator’s bias. Not only the patients but also trialists have expectations, wittingly or unwittingly [43]. Considering the fact that more studies in DD/MI category originated from Asia, which is an acupuncture friendly region where researchers are eager to prove its efficacy, than FD category, it may be trialists’ expectation on top of the participants’ expectation that resulted in such higher estimation.]

Revised: … in such higher estimation. Besides, when we evaluated whether outcome assessor was blinded to the intervention in the included studies, the proportion of studies with appropriate blinding for outcome assessor was slightly lower in DD category (33 out for 40, 80%) compared to FD category (20 out of 25, 82.5%). This too supports investigator bias may possibly play a part in overestimation of treatment effect in the studies of DD category.

7. Page 14, line 26: Commenting on the expectation is helpful for the discussion but the study as designed does not appear to measure expectation of treatment. This should be made more clear in the discussion.

ANSWER> We added a sentence to clarify this point at the end of the 1st explanation commenting on the modulation of expectation:

Original: [This is supported by a pooled analysis from four, large population based German RCTs which demonstrated higher expectation was associated with a better outcome regardless of given treatment [42].]

Revised: This is supported by a pooled analysis from four, large population based German RCTs which demonstrated higher expectation was associated with a better outcome regardless of given
treatment [42]. Considering that our study was not designed to directly measure participants’ expectation of treatment, however, we cannot estimate to what extent expectation contributed to greater effect size from the studies in DD category than those in FD category.

We appreciate your comments and suggestions which we believe helped us improve this manuscript substantially.