Author's response to reviews

Title: Community-based mental health treatments for survivors of torture and militant attacks in Southern Iraq: a randomized control trial

Authors:

William M Weiss (wweiss1@jhu.edu)
Laura K Murray (lmurra15@jhu.edu)
Goran A Zangana (gsabir@heartlandalliance.org)
Zayan Mahmooth (zayanmahmooth@gmail.com)
Debra Kaysen (dkaysen@u.washington.edu)
Shannon Dorsey (dorsey2@u.washington.edu)
Kristen Lindgren (kpl9716@uw.edu)
Alden Gross (agross14@jhu.edu)
Sarah M Murray (smurray9@jhu.edu)
Judith K Bass (jbass1@jhu.edu)
Paul Bolton (pbolton1@jhu.edu)

Version: 4  Date: 29 August 2015

Author's response to reviews: see over
RESPONSE TO REVIEWERS

[Please see Track Changes version of Manuscript in Additional Material Files]

REVIEWER #1

Minor essential revisions

1. It would have been helpful to include more information on the locally developed study instrument in terms of number of items and example items.

   More information has been provided in the text including number and example items (see pages 7 and 8).

2. When reporting test-retest reliability, it would be useful to provide information on time between testing sessions.

   We have added information on the timing between the testing and retest is this information is provided in the text on page 8. The re-interview was within three weeks of the first interview (the average time was seven days after the first interview). The time between the interview and re-interview was longer than usual because of disruptions caused by insecurity and holidays (Ramadan).

3. It wasn’t clear to me what “A median difference in scale scores between those diagnosed with and without a condition (on a scale of 0 to 3)” referred to. Were participants’ diagnoses rated on a scale of 0 to 3?

   We have reviewed this section and agree that it was confusing. We have edited the paragraph to make it clear that comparison was done between the mean of the total scale score for (for example) depression among those diagnosed with depression by a psychiatrist and the mean total depression score of those said by the psychiatrist to not have depression. The same was done for anxiety and PTS.

4. On page 8, the authors refer to ROC analyses. Are these published elsewhere? If so, the reference should be provided. If not, more detail should be provided.

   See Page 9. The ROC analysis was based on diagnosis of PTS by the psychiatrists (dichotomous variable) and the HTQ scale score (continuous variable). Area under the curve for this analysis was 0.75 (significance=0.000) suggesting a level of accuracy that was fair. Lending equal weight to sensitivity and specificity, 35.5 was the score at which sensitivity and 1-specificity were maximized, which we rounded up to 36.

5. Could the authors provide the rate of missing data in this study?

   (See pages 13 and 14). There were two clients in the CPT trial who had more than 40% of the items missing in their baseline assessment of anxiety. There were also two clients in the CPT trial who had more than 40% of items missing in their baseline assessment of
depression. When looking at individual items, there were no items with more than 5% of total responses missing in the baseline data.

Nine CPT clients and 3 CETA clients had no follow up scores due to not receiving follow up or lost records. Among those who had follow up scores recorded, there was 1 client in the CPT trial who had more than 40% of the items missing in the function scale and 1 client in the CETA trial who had more than 40% of the items missing in the trauma scale. When looking at individual items, there were no items with more than 5% of total responses missing in the follow up data among those with recorded follow up.

6. It wasn’t clear to me what the authors meant by “the adjusted-model results were almost identical to the results of the unadjusted model for both CETA and CPT”. Could the authors clarify?

(See Page 15). In the primary analysis, the regression models were not adjusted with added covariates such as age, gender, educational level. The dependent variable (mean scale score) was modeled against only two independent variables: intervention status (intervention or control) and time (time 1 or time 2). We assumed that the randomization process was sufficient to make the intervention and control groups equal for the main analysis (the unadjusted model). As a check, we did a sensitivity analysis that adjusted the regression model with additional independent variables (the adjusted model) such as age, gender, education status, working status. If the findings are similar, this provides more confidence that the randomization process sufficiently equalized the intervention and control groups.

Discretionary revisions

1. The authors also stated that “Criterion validity was supported for all scales except for depression among women”. Could some supporting evidence be provided for this statement?

The supporting evidence is provided immediately after the statement quoted above, but may not have been clear given the questions the reviewer had above (#3) about diagnoses and scale scores. We have altered the text slightly to make it clearer that the results after the statement refer to criterion validity.

2. On page 8, the authors refer to a locally validated HTQ. More information could be provided about this scale, including how many items, on which scale it was scored etc.

(See pages 7 and 8). Our locally validated HTQ included 29 questions, each about symptoms of trauma. The response to each question as scored on a scale from 0 to 3 reflecting how often the participant reported experiencing each symptom: No or never (scale score = 0); sometimes (scale score = 1); often (scale score = 2); or, always (scale score = 3).
3. What comprised the screening instrument referred to on page 8?

(See Page 9) The screening instrument was the same instrument used to measure the severity of symptoms experienced by participants (the dependent variable of the study). We also use the instrument to screen for eligibility for the study. The instrument had a section of function symptoms, a section on depression and anxiety symptoms, a section on trauma symptoms, a section on problems of torture survivors (identified during a qualitative study before the trial), and a section with demographic questions. A score of 36 or higher on the 29-question trauma section was the cutoff for eligibility into the study based on the ability of that cutoff to discriminate those diagnosed with PTSD from those without PTSD, as determined during the earlier validity study.

4. I found the Common Treatment Elements Approach interesting and compelling. However, it seemed to me that selecting and implementing multiple modules in a tailored fashion would require a considerable degree of clinical skill – perhaps more than implementing a standardized intervention approach such as CPT. Could the authors comment on the implications of this for disseminating this intervention in contexts where limited training and supervision was available?

We are glad reviewers found it interesting and compelling. The reviewer is correct -- having a flexible intervention that includes decision making could introduce some increased complexity as we train providers/supervisors in the decision making. However, we still provided the training in essentially the same amount of time as when training providers in standardized intervention approaches like CPT (compare pages 10 & 11), due to simplification of all components for lay providers. We clarified that the CETA training was 10 days and the CPT training was 7 days. Thus there was an increase in training days but it was fairly negligible. The finding that paraprofessionals were able to learn CETA including the decision rules is one major outcome of this trial (as well as our other trial of CETA on the Thailand border). Part of the CETA process was to simplify the decision rules and create default flows, as compared to how common elements interventions have been developed in high income countries. Access to supervision to help counselors also was essential in utilizing either intervention.

Interestingly, the state of the science on dissemination and implementation in the US and globally suggests that without sufficient training and supervision, neither standard interventions (CPT) nor modularized interventions (CETA) would “work.” Each of these interventions require sufficient training and supervision (what is “sufficient” for both, with lay counselors, is currently an unanswered empirical question). At the current time, we feel the potential tradeoff of greater applicability of CETA’s reach (e.g., can be used for multiple presenting problems and comorbid presentations) and some increased complexity for training and supervision is worthwhile. Nonetheless, this reviewer highlights an important limitation of this work (the high levels of training and supervision) and a future direction for research. Effective and efficient methods of training and supervising lay providers to deliver interventions are needed.
5. In the Sample Size section, it wasn’t clear to me what “A standard design effect of 1.5 given a lack of other studies” was referring to. Could the authors clarify?

The design effect is the multiple that a sample size—calculated using a formula for a simple random sample—needs to be increased by to compensate for our study design that did not use a simple random sample of eligible participants. Not all persons in the study area were not eligible to enter the trial: only clients of counselors trained by the study were eligible. We expect that limiting who was eligible for the study reduced the variation of characteristics of persons in the study, relative to if participants had been selected by simple random sample. Increasing the sample size helps compensate for this reduced variation. For this study (see protocol also), we selected a design effect of 1.5, given limited information about design effects in the region for this topic, based on the values for a medium effect size per the following reference:


6. Did the authors look at therapist effects? How many therapists implemented treatment in each arm?

Therapist effects were looked at and compensated for by using a multi-level regression model with therapist specified as a level in the model. 12 counselors provided CETA, and 17 counselors provided CPT (Page 5).

7. The discussion was generally well-written and provided a thoughtful consideration of the study. At times, more context could be provided. For example, in the first paragraph, the authors state that “Rahman and colleagues found effect sizes for CBT of 0.70 – 0.80 for depression, disability and functioning”, without stating where the study was conducted or with which population.

The context provided in the discussion has been reviewed. Note that the Rahman study was carried out in rural Pakistan. The study population included married women (aged 16-45 years) in their third trimester of pregnancy with perinatal depression. The providers were primary health workers were trained to provide CBT (intervention) or not trained (control).

8. At times, the distinction between the result and discussion are not entirely clear. For example, the authors include the “implementation” section in the results, without providing data. Perhaps this would be better placed in the discussion. Further, in the discussion the authors state that the investigated various factors including the effect sizes of the therapy, differences in context and the possibility of misconduct. It is not always clear whether quantitative or qualitative methods were used to investigate these factors; or whether the authors are simply considering them theoretically. This could be clarified.
The implementation section has been moved to the Discussion as per the reviewer’s suggestion. The investigation of effect sizes, differences in context and the possibility of misconduct was done using quantitative methods with the existing data collected from participants in the study.

REVIEWER #2

Major compulsory revisions:

The trial is described as a 3-arm trial yet it appears to comprise two separate parallel arm trials. Although it is discussed that the aim was to combine the two control groups, it is not clear why the randomisation didn’t take place overall, assigning each individual to either of the interventions or to the control group. The randomisation also seems to have taken place at CMHW level and overall randomisation may have resulted in more balanced groups at baseline in regard to the demographics. Although tables 1a and 1b indicate that the control and intervention arms are similar for the mental health outcomes, there are some demographic differences. But these may be unavoidable due to the low numbers in the trial.

The trial protocol is available at: http://tinyurl.com/CETA-Iraq-Protocol. The study objectives were to assess the effectiveness of CETA (called CBI at the time the protocol was written) and the effectiveness of CPT. The study was not intended to compare CETA to CPT (see Point 12 in protocol) so the reviewer is correct that the study is better described as two separate parallel trials arms rather than a 3-arm trial. The text has been corrected to reflect this on Pages 5 and 6.

Randomization was done at the CMHW level because we fixed the treatment provided by CMHWs. We did not want to open the possibility that CMHWs would mix the components of the two different treatments. In addition, the CMHWs were not mental health professionals and we did not believe they could effectively learn to provide two different therapies at the same time.

Another decision was to provide the treatment by geographic region. In this, the CMHWs in one geographic region would provide CETA, while CMHWs in the 2nd geographic region provided CPT. This helped prevent unplanned cross-over of therapies. This decision also allowed the study team to train a team of supervisors to learn and supervise one type of therapy only. One team of supervisors learned CETA, while the other team learned CPT. Each supervision team lived and worked in the area where their assigned intervention was provided, allowing them to have face-to-face meetings with CMHWs and review cases.

There are issues with the sample size calculation as this is for a 3-arm trial and it is also not accounting for clustering my CMHW. Additionally it is not clear why a medium effect size was chosen for this calculation.
Please see Point 12 of the protocol and the Sample Size section of the manuscript. The design effect of 1.5 was the inflation factor (a multiple) for the sample size to account for clustering by CMHW. Each arm (CETA, CPT, Control when combined) had a sample size that accounted for clustering by CMHW. We realize that when we decided not to combine the controls in the analysis, that we lost power, increasing the chance of a beta error. However, given the differences in the settings and experiences between the two areas, we believe it is appropriate not to combine the controls from each area, even though this results in a loss of power.

A moderate effect size was selected over a large or small effect size because we wanted to be able to identify clinically relevant differences with a reasonable amount of resources. We believe that a moderate effect is sufficiently large to be clinically relevant, and that we did not want a lower sample size to limit the study to detecting only large effect sizes. The ability to detect a small effect size would have required a very large sample size and it would not have been of clinical value to detect a small difference.

The CETA intervention is described as a very flexible approach, but it sounds like the delivery could differ quite significantly by CMHW and it may therefore be difficult to ensure intervention fidelity, and this issue is raised in the results. Furthermore, it is not clear if this intervention has been piloted and the description of the key elements of this intervention are limited.

The intervention was piloted (Murray, Dorsey et al., 2013; cited in the text pg. 10). The pilot paper provides a more lengthy description of the key elements in CETA as well as results from the pilot. The reviewer is correct, CETA could differ by client and/or provider by design. Providers were expected to use assessment and observational data to make decisions about differences in CETA delivery; e.g., element selection and delivery, and had a “default flow” based on inclusion criteria. Part of assessing fidelity involved ensuring that the assessment and counselor’s report of observation data matched selection of elements, sequencing and dosing. When the supervisor or trainer did not agree with the counselor, further discussion was initiated (with supervisors and trainers having ultimate control over decisions; see Murray, Dorsey et al., 2011 for more about our Apprenticeship Model of training). Fidelity to CETA was closely tracked— and it included following the decided upon plan (which elements would be provided and in what sequence) and ensuring that delivery of each element had fidelity (e.g., lay counselor completed all expected steps of Relaxation, for example).

Although two different interventions are studied, they are not compared which may have been possible if the randomisation had been conducted as a 3-arm trial. This information would have been valuable in comparing the effectiveness of the CETA to CPT which is a more established intervention.

The reviewer is correct (as also commented on above) that the two interventions are not compared. This was a decision primarily because we had not previously determined the effectiveness of CPT in the Southern Iraq setting. This was the first trial of CPT in Southern
Iraq and we have not established that comparing CETA to CPT is appropriate at this time. We believed it was important to first establish the effectiveness of each intervention independently in that setting.

There are some issues with the statistical analyses which have been used and it is likely that there may not be the statistical power to test the multilevel models which were conducted, given that the power calculation did not account for clustering by CMHW. The Hausman test was used to determine if random effects models were appropriate, but given this was of borderline significance, there is uncertainty as to whether these are the most appropriate models.

As described above, the power calculation prior to the study did include an inflation factor (a design effect of 1.5) to account for clustering. In addition, the standard errors in the analysis have been adjusted for clustering by CMHW and by client (as we have two time points for each client). The Hausman test results were not of borderline significance. Although the decision threshold was set at 0.05, the actual test results (not reported) were closer to 1.0 (a significant Hausman test < .05 would indicate that a fixed effect model is preferable over a multi-level model). Apart from the limitation of not combining the controls in the analysis (reducing power), we believe the sample size calculations used, and the selection of a multi-level model are appropriate.

Finally, I have concerns regarding the implications and interpretation of the findings given that the control group for CETA was selected as a post-hoc decision. The effect sizes found may have been much smaller if the control groups had been combined as originally planned and I think that a better case could be made for the value of this intervention if the comparison was conducted as planned in the protocol.

The reviewer suggests that combining the controls, as in the original analysis plan, would make a better case for the value of CETA. Although we certainly can see the advantages to this suggestion, as can be seen in Table 2 in the manuscript, there were substantive differences between the controls in the CPT geographic area as compared to the controls in the CETA area. For this reason, and given our knowledge now of the differences between the areas, we believe it would not be appropriate to mix these very different samples and provide an average result. Similar to reporting separately the results where an interaction is present, we believe it is more appropriate to provide the results separately for the different regions.