Title: Preliminary Comments on the 2017 APA Clinical Practice Guideline for the Treatment of Posttraumatic Stress Disorder (PTSD) in Adults

Authors: Sarah K Dominguez and Christopher W Lee

Correspondence to: Associate Professor Christopher Lee
University of Western Australia
chris.lee@uwa.edu.au

Overview

The APA Practice Guideline Development Panel for the Treatment of Posttraumatic Stress Disorder (PTSD) concluded that there was strong evidence for cognitive behavioural therapy, cognitive processing therapy, cognitive therapy and exposure therapy yet weak evidence for EMDR. This is in spite of the findings of a systematic review of the evidence for treatment conducted by the Research Triangle Institute-University of North Carolina Evidence-Based Practice Center (RTI-UNC EBPC) which concluded that EMDR leads to significant loss of PTSD diagnosis and reduction in symptoms and greater reduction than controls on depression scores. In that review EMDR was marked down on strength of evidence (SOE) for symptom reduction for PTSD. However there were several problems with the conclusion of that review. Firstly, in assessing the evidence in one of the studies the reviewers chose an incorrect measure that skewed the data. We recalculated a meta-analysis with a more appropriate measure and found the SOE to improve. Secondly even if the original measure was chosen we highlight inconsistencies with the way SOE was assessed for EMDR and cognitive therapy and cognitive processing therapy. Thirdly we highlight two papers that were omitted from the analysis. We again recalculated the meta-analysis and found that this also had an effect on SOE. The resultant standard mean difference and confidence intervals were actually better for EMDR (more precision) than for cognitive processing therapy or cognitive therapy. Therefore, the SOE should have been rated as medium in that Agency for Healthcare Research and Quality report and rated as strong in the APA guidelines given it was at least equivalent to these CBT approaches. This would bring the APA guidelines in line with other recent practice guidelines from other countries. Less critical but also important, were several inaccuracies in the risk of bias in some studies and the failure to consider studies supporting strong gains of EMDR at follow-up.
1. **Using a more appropriate measure in an included study[1]**

The RTI-UNC review (Figure 17) referred to mean changes in PTSD symptoms for EMDR versus control comparisons. There are 4 studies listed and changes were assessed in each of the studies on identified primary measures. For example in the Rothbaum study [2] this was the Clinician Administered PTSD Scale (CAPS). The primary outcome measure for the Carlson study [1] was also the CAPS and this is reported in the original article for pre and follow-up data. The effect size is large (Cohen’s $d=1.8$). However, CAPS scores were not collected at post-treatment. A battery of self-report measures were collected at post-treatment including the Mississippi Scale for Combat Related PTSD (M-PTSD) and the Impact of Events Scale (IES). However, in the RTI-UNC analysis the IES was chosen above the M-PTSD. Why is difficult to fathom. The M-PTSD is more comprehensive than the IES, and similar to the CAPS it is based on the DSM. A review article at the time recommended the M-PTSD above all other self-report measures for assessing PTSD [3]. Furthermore, when the RTI-UNC reviewers were describing which studies were included in their analysis and wanted to compare the severity of PTSD symptoms at baseline for each study, they chose the M-PTSD over the IES (see Table 9 and 18). Also later in the report when assessing the effectiveness of relaxation they again use the M-PTSD (p70). So why they reverted to the IES in the middle of the report when assessing change in the PTSD symptom level for this study is perplexing.

Changing the outcome measure from the IES to the M-PTSD significantly changes the results with regards to PTSD symptom reduction following EMDR. If this correction is utilized the effect size, precision, and consistency are all improved [SMD, -1.28 (-1.81 to -0.74); $I^2=43\%$].

RTI-UNC guidelines define precision as the width of the confidence interval and consistency is the number of studies in the same direction and appears to take into account the heterogeneity (The RTI-UNC quote heterogeneity when discussing consistency in appendix 1). Therefore, heterogeneity at 43% for EMDR is better than mixed cognitive behavioural therapy (CBT), cognitive therapy (CT), and cognitive processing therapy (CPT) where heterogeneity was significant and ranged between 80 and 87%. In addition to EMDR being more consistent the precision improves to 1.07 (difference between lower and upper end of the confidence interval), which is better than both CPT (1.1) and CT (1.38). Therefore, there is no basis to argue SOE is better for these CBT therapies.
Finally changing the outcome measure to the more comprehensive measure of the M-PTSD provides a result more consistent with the rest of the data from the study. The effect size for the CAPS at follow-up was 1.82 for the EMDR treatment and there were large effect sizes for both depression and anxiety measures post-treatment in comparison to control making the IES result an anomaly.

2. **Strength of evidence using only the data supplied in the RTI-UNC report.**

There appears to be differences in how the consistency domain was rated with respect to SOE for PTSD symptom reduction. This section only refers to the analysis on the 4 included studies in the RTI-UNC report not an analysis that should have included the 6 studies described later in this report. With regards to PTSD Symptom Reduction the RTI-UNC study rated EMDR as *Inconsistent*. This is based on the heterogeneity of the related studies, the direction of the effects and the magnitude of these effects. Examination of the impact of CT on PTSD symptom reduction suggests that there is high heterogeneity in the evidence for both scales ($I^2 = 79.6\%$), as shown on Table G-2. However, rather than *Inconsistent*, this category has been labelled *Some Inconsistency*. The annotation of this table indicates that the ‘Direction of effects were consistent; magnitude of effects ranged from very large to small’ (page G-4). Similar annotations have been made on Table G-1 and G-13 resulting in studies with high heterogeneity obtaining ratings of *Consistent* or *Some Inconsistency*.

These annotations have not been applied to the analysis of EMDR. With regards to impact on PTSD Symptom Reduction, while the heterogeneity of EMDR results is high ($I^2=70\%$), this is lower than the same measure for CT mentioned above. Further, the direction of the effects from EMDR studies is consistent and the magnitude of these effects ranged from ‘almost small to very large’, which is similar to related results for CT. This suggests that the Consistency domain for EMDR on PTSD Symptom Reduction should have been moved from *Inconsistent* to *Some Inconsistency*, to ensure uniformity in rating across therapies.

A change of the Consistency domain would mean that the domains for PTSD Symptom Reduction following EMDR would be comparable to that for CT across all measures. Following this would move the SOE for EMDR for PTSD symptom reduction from Low to Moderate.
It may have been argued that this annotation may not apply to the EMDR results with regards to Symptom Reduction as one of the studies [1] had a confidence interval where the lower point falls below zero. However two of the studies in CBT-Mixed Interventions [4, 5] have their confidence intervals falling below zero, and this intervention is still rated as consistent. Further, if the outcome measure analysed for this study was altered, as suggested from the IES to M-PTSD, this would no longer be the case.

3. **Omissions of RCT studies relevant to the research questions.**

An additional error in the analysis occurs in the RTI-UNC report is the failure to include two studies relevant to the issue of whether EMDR leads to more symptom reduction than a control condition. The report purports to assess, as its first research question, the effectiveness of psychological treatments “compared with wait list, usual care (as defined by the study), no intervention, placebo,” (pES-5). Yet a significant study was excluded from the analysis represented in Figure 17. The study omitted is van der Kolk et al. [6]. This study is reported in Table 18, however it is inexplicably missing from the meta-analysis. The study by van der Kolk [6] had three arms to it. Participants were randomised to either EMDR or SSRI treatment condition, or a placebo control. In the RTI-UNC review, Figure 17 refers to mean changes in PTSD symptoms for EMDR vs control comparisons. As placebo is clearly a control condition Figure 17 should include the data from this study.

This omission cannot be justified on a basis of methodological procedures because other studies that included multiple arms were utilised in more than one place in order to answer key questions. For example, Marks et al. [7] appears in Table 9 when discussing coping skills trials, and again in Table 13, looking at the efficacy of exposure trials [8]. This suggests that there is no methodological issue that would result in the exclusion of van der Kolk et al. [8] data. The inclusion of this study into the analysis, with regards to PTSD symptom reduction for EMDR, would change the conclusions of SOE in the report. When we calculated the new confidence interval it was from -1.56 to -0.37 which is better precision than CPT and heterogeneity improved from just the 4 studies to remain better than CPT or CT.

Another important study omitted from the meta-analysis was published in 2015 [9]. A problem with the APA guidelines is that, although they are based on the review by RTI-UNC published in 2013, they won’t be made available until 2017. This means that while readers
may believe they are reading 2017 guidelines, they are actually 4 years out of date. Three recent randomised control trials [9-11] that support EMDR as evidence based are not considered in these conclusions. One study in particular, by van den Berg and colleagues [9] meets a high methodological standard. Indeed, in the RTI-UNC appendices this study is highlighted. The APA committee in reviewing the RTI-UNC findings acknowledged that the addition of this study to the analysis was likely to narrow the confidence interval and therefore impact on precision and would also improve consistency. “If a new meta-analysis were to be done... the confidence interval would be narrower and it is possible that the SOE might be upgraded from low to medium as a result.” (p. 622). This would have a significant bearing on SOE. However, seemingly paradoxically, after highlighting the impact of the addition of this study, they then conclude that there is insufficient evidence to determine whether the study would change the recommendation for EMDR. Moreover, they acknowledge that if the effect size stayed at medium/large and given the increased sample size that the SOE would be changed.

Testing the idea if whether precision improves and the effect size stays the same is a science issue. It is not difficult or time consuming to do the analysis. We used Comprehensive Meta-Analysis software (CMA) and input the same effect sizes reported in Figure 17 but added CAP scores and confidence intervals from the studies of van der Kolk and van den Berg. The results are presented in Table 1. The effect size remained large SMD =-.89 (-1.34, -0.44). The precision improved to a confidence interval difference of just .9 (see Table 1). Using the RTI-UNC own guidelines of assessing SOE, EMDR is doing better than both CPT and CT in both consistency and precision. In fact, it is closer to mixed CBT in precision than CPT or CT. Even more compelling is the heterogeneity which at 66% is better than mixed CBT, CT and CPT (heterogeneity was used in the RTI-UNC as an indication of consistency, see appendix 1). The total N is also substantial at 284. It is not possible from a science point of view to rate CPT and CT higher in SOE than EMDR.

Finally redoing the analysis for all 6 studies that compared EMDR to a control condition and using the more appropriate M-PTSD measure for the Carlson study (see Figure 1 below) the SMD is -0.99 and the confidence interval is from -1.41 to -0.58 (I²=57%). This is the best reflection of the state of the literature today. This is the result that should have been used by the APA. This data means that consistency for EMDR is better than CT, CPT and mixed CBT and EMDR has more precision than CT or CPT.
Table 1

Comparative statistics on precision and consistency analysis including changes when all relevant EMDR studies are included with appropriate comprehensive measures

<table>
<thead>
<tr>
<th>Treatment</th>
<th>PTSD symptom reduction</th>
<th>Difference</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive processing therapy</td>
<td>SMD -1.40 (-1.95, -0.85)</td>
<td>1.10</td>
<td>87%</td>
</tr>
<tr>
<td>Cognitive therapy</td>
<td>SMD -1.22 (-1.91, -0.53)</td>
<td>1.38</td>
<td>80%</td>
</tr>
<tr>
<td>CBT-mixed</td>
<td>SMD -1.09 (-1.4, -0.78)</td>
<td>0.62</td>
<td>87%</td>
</tr>
<tr>
<td>EMDR (original report using IES for Carlson)</td>
<td>SMD -1.08 (-1.83, -0.33)</td>
<td>1.50</td>
<td>70%</td>
</tr>
<tr>
<td>EMDR (using M-PTSD for Carlson)</td>
<td>SMD -1.28 (-1.81, 0.74)</td>
<td>1.07</td>
<td>48%</td>
</tr>
<tr>
<td>EMDR with van der Kolk and van der Berg and using IES for Carlson</td>
<td>SMD -.89 (-1.34, -0.44)</td>
<td>0.90</td>
<td>66%</td>
</tr>
<tr>
<td>EMDR with van der Kolk and van der Berg and using M-PTSD for Carlson</td>
<td>SMD -.99 (-1.41,-0.58)</td>
<td>0.93</td>
<td>57%</td>
</tr>
</tbody>
</table>

SMD, standard mean difference; IES, Impact of Events Scale; M-PTSD, Mississippi Scale for Combat Related PTSD

4. Papers inappropriately included in the analysis of the RTI-UNC review.

In examining the papers included from the analysis in the RTI-UNC review [8] there appear to be errors made in the inclusion of certain studies from the analysis of evidence. An example of this is the inclusion of Taylor et al. [12], despite several significant validity concerns and concerns regarding the interpretation regarding psychometric properties.
In Table E1 of the RTI-UNC paper there is a category that examines whether the participant groups in the study were equivalent at baseline [8]. On page E-21, this category for the study by Taylor et al., [12] was rated as yes [8]. However, no pre-treatment test scores analysis for treatment conditions is reported. The only pre-treatment analysis reported is that there are no significant differences between dropouts and completers—regarding demographics and primary measures of interest. Furthermore, Figure 2 indicates that the participants in the exposure group reported less symptoms than those in the EMDR group at pre-treatment [12]. The confidence intervals on the bar graph show the mean score for the exposure group was outside the standard error of the EMDR group at pre-treatment for Hyperarousal, Re-experiencing, and Avoidance symptoms.

The bias in the Taylor [12] study is further inflated as it relied on a treatment completer analysis rather than an intent-to-treat analysis. This is critical in that given EMDR was more severe to begin with but the PE had a higher dropout rate (11% greater), the chance of systematic bias is elevated.

An additional error in the rater’s assessment of this study was the judgment that the providers of the therapy were masked. However logic asserts that this assessment is not possible in a design comparing two psychological treatments. Given these errors in the risk of bias the Taylor et al. [12], study should be reclassified from a medium to high risk of bias.

Thus the information on Table E-21 regarding Taylor et al. [12] is wrong in two categories. The study clearly has a high risk of bias not medium. Therefore, it should not have been used in decisions regarding CBT—Coping Skills (Table 9). Nor should it be included in any discussion on Change in CAPS for exposure therapy compared with relaxation therapy (Figure F-36). In this analysis the failure to find a significant difference between relaxation and exposure was at odds with the more methodological sound study by Marks et al [7].

With respect to findings related to EMDR, removing this study changes the interpretation of the RTI-UNC report with regards to EMDR and PTSD Symptom Change. The conclusion that all studies found a greater reduction in PTSD symptom scores in favour of EMDR over comparators (p. 67) still stands. However, the qualifying statement that not all differences reached statistical significance would now misrepresent the data. A more accurate statement is that in all bar one study the differences reached statistical significance. Clearly this results in a different sense of SOE with respect to consistency of results.
Another conclusion in the RTI-UNC report that needs adjusting by the removal of Taylor study is ‘Percentage of subjects achieving loss of diagnosis for exposure compared with EMDR’ (p53) as Taylor should not be included here which increases the effect size in favour of EMDR. Also it should be removed from the discussion of ‘PTSD symptom reduction for EMDR compared with relaxation’ (p F-73) and ‘Loss of PTSD diagnosis at 3 month follow-up for EMDR compared with relaxation’ (p. F-74). Similarly to Taylor being at odds with other better studies on loss of diagnosis comparing exposure therapy and relaxation the same is true that is at odds when comparing relaxation therapy with EMDR or exposure.

5. Papers inappropriately excluded from the analysis of the RTI-UNC review.

In examining the papers excluded from the analysis in the RTI-UNC report [8] there appears to be errors made in the exclusion of some studies from the analysis of evidence. Research by Lee and colleagues [13] was assessed as a high risk of bias however as explained below there appear to be errors in the examination of the results of this study.

Firstly, in Table E1 on the RTI-UNC paper there is a category that examines whether the participant groups in the study were equivalent at baseline. On page E-13, this category for the study by Lee et al. [13] was rated as unclear. However page 1077 of the Lee et al. [13] article reports,

“Independent t-tests were used to investigate differences between the groups on pre-treatment measures. No differences were found for the IES (t(22)=.11, p=.91), BDI (t(22)=1.05, p=.31), SI-PTSD (t(22)=1.63, p=.12), or MMPI-K (t(22)=1.31, p=.21). Therefore, the groups appeared to be equivalent on major variables.”

Therefore, the raters made an error in asserting that the paper was not clear on whether there were differences at baseline. This is in sharp contrast to the Taylor et al. [12] study where no baseline comparison data was analysed.

The raters of Lee’s study also marked it down saying that that the differential attrition data was unclear. However, the study clearly indicates that 24 participants entered the study, 12 were assigned EMDR and 12 were assigned to CBT, with three people dropping out, leaving 21 completers [13]. On page 1075 it is stated that 21 participants completed the study, 11 for
stress inoculation with prolonged exposure and 10 from EMDR. The article then describes how one of the EMDR non-completer was sent to prison. It does not make sense that the raters can claim that the attrition is not clear.

Given the above two errors the risk of bias in the study deserves to be reclassified from high risk of bias to moderate. Following, it should be moved from Table 19 to Table 18 and included in the analysis. This inclusion strengthens the evidence base for a reduction in PTSD symptoms and for loss of diagnosis for EMDR.

The inclusion of Lee et al. [13] and the exclusion of Taylor et al. [12] together would have a very substantial effect on conclusions of the SOE. However, the removal of just Taylor makes a substantial difference.

So the final number of randomized control trials that the APA should considered when assessing for evidence for EMDR is 8 if using the RTI-UNC criteria. Six of these [1, 2, 6, 9, 14, 15] included a waitlist or other minimal intervention control and 5 of these 6 compared EMDR to another manualised active treatment. Two trials [13, 16] compared EMDR to another manualised treatment only.

6. Lack of attention to follow up data

In the RTI-UNC analysis it states “Our meta-analysis (Figure 17) found greater reduction in PTSD symptoms for EMDR than for controls…. Treatment gains were maintained for studies reporting follow up at 3, 6, or 9 months (p. 67).

This statement ignores the considerable data that EMDR treatment gains are maintained far beyond end of treatment time points. At the very least the follow up study on the Högberg data [17] which reported treatment gains for EMDR were maintained at 37 months should have been mentioned. However there is a case that other data such as that presented in Wilson’s study [18, 19] should have been included. In this study EMDR achieved large treatment effect sizes for both participants who all met the criteria for PTSD and for those that didn’t quite make the diagnosis. In fact, whether participants met full diagnostic criteria or not did not have any significant effect on treatment outcome. In a follow-up article they give separate figures for both groups of participants showing that the treatment gains were maintained at follow-up (15 months) with large effect sizes.
Conclusion

The APA guidelines are utilised worldwide and the accuracy of the document and the data it contains is crucial. This review highlights some inaccuracies regarding the way studies were handled in the statistical review of papers particularly with respect to evidence concerning EMDR. Therefore, the subsequent conclusions of the draft guidelines are flawed. These flaws explain why the proposed 2017 guidelines are at odds with other best practise guidelines from other countries and international based guidelines such as the World Health Organisation in 2013 [20]
References.


### Meta Analysis

<table>
<thead>
<tr>
<th>Study name</th>
<th>Std diff in means</th>
<th>Standard error</th>
<th>Variance</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Z value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hogberg 2007</td>
<td>-0.680</td>
<td>0.403</td>
<td>0.160</td>
<td>-1.484</td>
<td>0.104</td>
<td>-1.699</td>
<td>0.098</td>
</tr>
<tr>
<td>Rothbaum 2005</td>
<td>-1.610</td>
<td>0.318</td>
<td>0.101</td>
<td>-2.234</td>
<td>-0.986</td>
<td>-3.000</td>
<td>0.002</td>
</tr>
<tr>
<td>Rothbaum 1997</td>
<td>-1.890</td>
<td>0.499</td>
<td>0.278</td>
<td>-2.859</td>
<td>-0.931</td>
<td>-3.765</td>
<td>0.000</td>
</tr>
<tr>
<td>van der Kolk 2007</td>
<td>-0.596</td>
<td>0.269</td>
<td>0.084</td>
<td>-1.103</td>
<td>-0.029</td>
<td>-2.050</td>
<td>0.039</td>
</tr>
<tr>
<td>van den Berg 2015</td>
<td>-0.650</td>
<td>0.204</td>
<td>0.042</td>
<td>-1.049</td>
<td>-0.251</td>
<td>-3.126</td>
<td>0.001</td>
</tr>
<tr>
<td>Carson 1950 Mississippi</td>
<td>-0.944</td>
<td>0.451</td>
<td>0.204</td>
<td>-1.828</td>
<td>-0.059</td>
<td>-2.082</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Figure 1. Results of the meta-analysis using all appropriate studies and measures.