Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: *Psychological and Pharmacological Treatments for Adults With Posttraumatic Stress Disorder (PTSD): A Systematic Review Update*

Draft review available for public comment from November 15, 2017 to December 29, 2017.

Research Review Citation: Hoffman V, Middleton JC, Feltner C, Gaynes BN, Weber RP, Bann C, Viswanathan M, Lohr KN, Baker C, Green J. Psychological and Pharmacological Treatments for Adults With Posttraumatic Stress Disorder: A Systematic Review Update. Comparative Effectiveness Review No. 207. (Prepared by the RTI International–University of North Carolina at Chapel Hill Evidence-based Practice Center under Contract No. 290-2015-00011-I for AHRQ and PCORI.) AHRQ Publication No. 18-EHC011-EF. PCORI Publication No. 2018-SR-01. Rockville, MD: Agency for Healthcare Research and Quality; May 2018. Posted final reports are located on the Effective Health Care Program search page. DOI: https://doi.org/10.23970/AHRQEPCCER207.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. This research review is posted to the EHC Program Web site or AHRQ Web site in draft form for public comment for a 45-day period. Comments can be submitted via the Web site, mail or E-mail. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Introduction	The Executive Summary is an excellent synopsis of the findings in the SR. The bulleted key findings and Tables ES-2 and ES-3 are effective summaries of the much larger report. The footnotes to Table ES-2 and Table ES-3 are, to those who are familiar with the 2013 SR, the most useful part of the Executive Summary, since they summarize changes in strength of evidence (SOE) between the 2013 SR and the current SR.	N/A
Peer Reviewer 2	Introduction	A very minor suggestion: Figure ES-1. Could you complete the arrow leading from "Patient Characteristics...) to "Psychological or ...."? The patient characteristics of primary interest should be those known or knowable BEFORE choice of treatment (in a randomized clinical trial (RCT)) before randomization, for only then are such characteristics of use in choosing which treatment might be best for which patient.	We have completed the arrow as suggested.
Peer Reviewer 2	Introduction	The definition of SOE should be presented before it is in text.	SOE is first described in its own section in the Methods section. The abbreviation is spelled out and the text that follows defines the grades.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Introduction	However, I think your SOE may confound issues related to the quality of the study, i.e., decisions made by the researchers), with quality of the treatment/control comparison being evaluated (effect size). Prof. Olkin used to describe the "rotten apples" phenomenon. Clearly having completely rotten apples in a barrel will spoil the whole barrel. However, leaving partially rotten apples will end the same way. If a study does not have an acceptable control group, if the subjects are not randomized to the treatment/control groups, if the analysis is not done "by intention to treat", or if the analytic procedure is simply wrong, if there is not a clear 'a priori' acceptable outcome measure(s), if any of these are not there, the study is a "rotten apple", and should be omitted from consideration in any meta-analysis. If I've read this correctly, you did require RCTs for treatment efficacy/effectiveness, but allows observational studies for adverse effects. I think this is a mistake. Adverse effects are seen even with inert placebos, and are often related to the disorder being treated and not to the treatment being evaluated. However, for example, drop out after randomization may	SOE would not provide evidence of a “bad apple” in the way that a high risk of bias rating could. The strength of evidence grade is given after carefully considering more than just the risk of bias of a study (which is one domain considered), but also (mainly) the consistency, directness, precision of the evidence, and reporting bias. The SOE grade reflects the strength of the body of evidence to answer KQs on the comparative effectiveness, efficacy, and harms of the interventions in this review. Although we did “allow” observational studies to answer KQ4, we actually did not find any observational studies that otherwise fit our inclusion/exclusion criteria. All KQ4 studies are RCTs with efficacy or effectiveness findings reported in either KQ1, KQ2, or KQ3. We did not include any high risk of bias studies in our meta analyses or SOE gradings. We also did not pool studies for which there was a high degree of heterogeneity. We agree that doing so would compromise the findings.

Commentator & Affiliation	Section	Comment	Response
		be in response to treatment. No otherwise valid study should be dropped simply because there are dropouts (no matter how many) as long as a valid analysis is done "by intention to treat". No study should be dropped because of a judgment about its outcome. Clearly one laudable goal here is to provide the evidence-base for clinical decision-making in this context. However, the studies you cite also serve as models for acceptable research in future studies. I am concerned about studies you label "Moderate" or "Low" SOE, as to whether these are 'partially rotten apples" that compromise the first goal, but also encourage poorly designed future studies. It is very common for a statistician to hear a researcher excuse poor design/analysis decisions saying that, while this study will likely not prove definitive, it will make a contribution via meta-analysis. Rather, such studies compromise meta-analysis.Any study that does not satisfy minimal criteria, should be cited, along with its deficiencies in the Appendix (as you did), but excluded from the meta-analysis, and thus not in text. I realize that many studies have been excluded (Appendix), but am not sure of those labelled "Moderate" or "Low" SOE that appear in text. Finally, an otherwise valid study would need to be excluded if it did not report a valid effect size and its confidence interval or standard error. This is something the journals should be watching for.	
Peer Reviewer 3	Introduction	p. 14: In Table ES-1, group therapy is listed as a therapy, but it is actually a therapy modality. Also, I expected to see something later on about group therapy but could not find where this was addressed, even using the search function in Acrobat. I can see that group studies were included with the relevant type, e.g., Group CPT with CPT, but I think readers would have expected something to be said about group therapy given this table.	We agree and have deleted it from the table. You are correct in that group therapy is a treatment modality and we describe it in the context of other different treatments that were used if a group format was used (e.g., group CBT, etc.).
Peer Reviewer 3	Introduction	In addition, this section correctly lists hypnosis and neurofeedback as “other” therapies, whereas they are treated as CBT coping skills therapy in the review.	We agree and have moved hypnosis, mindfulness based stress reduction (MBSR) and neurofeedback to the Other Psychological interventions section.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Introduction	Neither of these interventions is CBT, nor is relaxation (see comments on Methods).	We agree, although relaxation is one component of CBT-coping. We checked each of the studies that test a relaxation intervention included in this review and agreed that they should be kept with the stress inoculation therapy studies in the CBT-coping skills section. The prior review categorized relaxation as CBT-coping skills and, in part because this is an update review, we agree that relaxation studies should remain in the CBT-coping skills section.
Peer Reviewer 3	Introduction	p. 32: The prevalence data are old and outdated. It would be more useful to cite data from NESARC-III.	We have added NESARC-III prevalence information to the revised draft. NESARC-III used DSM-5 criteria which is important to cite as you suggest, however, because the bulk of the studies included in this review are based on DSM-IV criteria, we thought it would be helpful to include both findings from both studies.
Peer Reviewer 3	Introduction	In the section on Burden, the Kessler study on utilization (ref. 9) is not cited in the text where its data are presented.	We have added this reference.
Peer Reviewer 3	Introduction	These data, which were collected in 2000, also are outdated.	To our knowledge, there has not be a similar paper published on burden using more recent data (and the only more recent such epidemiologic data are, as you suggested, NESARC-III)
Peer Reviewer 3	Introduction	p. 33: In the section on Psychological Interventions, the PTSD clusters listed are DSM-IV and not DSM-5.	Yes, that is true, but as mentioned, most studies included in this review were based on DSM-IV criteria.
Peer Reviewer 3	Introduction	p. 34: In the section on existing CPGs, is it really fair to say that the existing guidelines are contradictory? There is more similarity than difference, and given the wide range in years when many of been published, they converge more than not around trauma-focused psychotherapy and selected medications.	We agree and have removed this language.
Peer Reviewer 4	Introduction	In the Executive Summary, page ES-7, the report makes reference to an APA report on treatments for PTSD. The noted reference does not appear to be a systematic review. I could not find this reference either in the Executive Summary or in the main report.	These references have been corrected, thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Introduction	The introduction was well laid out. It included the definition of PTSD at the outset (first sentence), causes of PTSD and how diagnosed according to DSM. It then lays out the prevalence, burden and treatment relevant to PTSD in the target population to include an overview of psychological and pharmacological interventions and finally outcomes.	N/A
Peer Reviewer 1	Methods	The methods are done well but I have a substantial number of questions / concerns / comments about the methods.	We have responded to each, below.
Peer Reviewer 1	Methods	For bodies of evidence in which all of the included RCTs used the CAPS as an outcome measure for assessing PTSD symptom reduction, only WMDs were reported. Can you also report SMDs as well (i.e., in addition to, not instead of, WMDs)? Readers want to be able to compare effect sizes across interventions when the comparators were the same (e.g., for psychological treatments, when the comparators were wait-list controls). For a guideline panel, it is much easier to have SMDs for all bodies of evidence and to be able to define magnitude of effect on the basis of SMDs rather than to have to define a clinically significant effect on the basis of each individual outcome measure. I recognize that SMDs are controversial (e.g., Greenland's criticism (Greenland S, et al, Am. J Epidemiol 1986; 23(2):203-8) of standardized regression coefficients applies to SMDs as well). However, the SR already uses SMDs. Why not just add them to those places where you report the findings solely in terms of WMDs? I understand that it may appear to be difficult to come up with a theoretically-driven, rational scheme for categorizing CBT studies, each of which used a treatment that included multiple different elements.	In response to this review and those of other reviewers, we have reported SMDs throughout. We debated retaining WMD and adding SMDs (as you suggest) or simply replacing with SMDs (which 2 other reviewers suggested) and decided that for ease of presentation and interpretation, we would only present SMDs. We do, in some instances, calculated pooled mean change in raw scale change scores between groups (WMDs) in the text.

Peer Reviewer 1	Methods	Why did the RTI-UNC EPC (hereafter referred to as "the EPC") continue to use the category "CBT-mixed", as in the 2013 SR? That category was developed on an ad hoc basis in the 2013 SR to describe a wide variety of interventions that seemed to fall under the broad rubric of CBT but that had diverse intervention elements. The category of CBT-M was not developed on the basis of theory or review of psychotherapy literature or PTSD literature. That category has been criticized by a wide swath of clinicians who provide care to persons with PTSD, by persons who have developed psychological treatments for PTSD and by researchers who have studied psychological treatments for PTSD. That category was a lightning rod for criticism in the 2013 report and will likely be a lightning rod for criticism in the current report by clinicians and researchers in the PTSD world. It served as a justification for dismissing the 2013 SR as being out of touch with the world of PTSD treatment. However, I think that there is one: trauma-focused CBT vs non-trauma-focused CBT. Most of the other systematic reviews of psychological treatment for PTSD (such as Bisson's Cochrane SR in 2013 of psychological therapies for chronic PTSD) have classified CBT treatments (that are not other specific therapies, such as cognitive therapy, cognitive processing therapy, etc.) as trauma-focused and non-trauma-focused. Bisson and colleagues defined trauma focused CBT in the following way: ""TFCBT is a variant of cognitive behavioural therapy (CBT), which includes a number of techniques to help a person overcome a traumatic event. It is a combination of cognitive therapy aimed at changing the way a person thinks, and behavioural therapy, which aims to change the way a person acts. TFCBT helps an individual come to terms with a trauma through exposure to memories of the event." This way of categorizing CBT is widely accepted within the PTSD field and corresponds to a theoretically driven scheme for categorizing CBT. It can be operationalized relatively easy (as indicated by the multiple SRs that have used that scheme). Why did the EPC not use this scheme? Can you re-do the section on CBT, eliminating the CBT-M category and instead using the TF-CBT and non-TF-CBT categories? It is a valid approach to categorizing psychological treatments for PTSD and it will help those who are most likely to be users of the report accept its findings. Why persist in using a categorization scheme that was developed ad hoc	We agree that there is a large amount of heterogeneity in the CBT-mixed group. We had extensive discussions about categorization but ultimately decided to retain the CBT-M category for a couple of reasons. One is that this was meant to be an update of the prior review. Shifting around categories would not permit a comparison of findings from the prior report with the updated one. More importantly, however, is that we considered the suggestion about recategorizing the CBT-M category into categories of TF-CBT and non-trauma focused CBT but decided against it because 1) many of the interventions included in other sections (CBT-exposure, EMDR) could be considered to be TF-CBT interventions, too and, furthermore, the definition of TF-CBT requires subjectivity as well (we re-reviewed Bisson's review and their categorizations were very different, included CPT in TF-CBT as well as Exposure and didn't call out individual types like our study did; 2) as you rightfully mention later, many of the interventions contain multiple components, some of which are trauma-focused and others which are not; and 3) there are only 2 treatments in the current CBT-M categorization that would have been (subjectively) classified as non-trauma focused CBT, so the bulk of the interventions included would be the same in the current CBT-M and newly created TF-CBT groups

Commentator & Affiliation	Section	Comment	Response
		and has been widely disparaged by those who actually treat people with PTSD or do research on PTSD treatment?	
Peer Reviewer 1	Methods	Why were quantitative meta-analyses done only for bodies of evidence with five or more studies? I saw the citation to Owens (which I was told by AHRQ actually should have been a reference to Cornell et al (2014). Random-effects meta-analysis of inconsistent effects: a time for change. Annals of internal medicine, 160(4), 267-270.) to justify that decision but I question the validity or utility of that decision. The Cornell paper demonstrates that the confidence intervals using the D-L method may be too narrow when the results of the constituent studies are heterogeneous. I accept that conclusion as demonstrated. The paper also states: "Large variation in study design, conduct, population, measurements, and analyses suggests that it may be unwise to estimate an average effect. When the number of studies is sufficiently large, organizing analyses around clinically or methodologically important study-level characteristics through stratification or metaregression may be more informative than a single summary estimate. When there are too few studies to stratify by study-level characteristics, whether pooling is reasonable must be addressed." It is hard for me to see how it is reasonable to operationalize that recommendation into the rigid procedure used in the current systematic review that quantitative meta-analyses would not be done when there were fewer than five studies in a body of evidence, without even assessing the degree of heterogeneity in methods (i.e., the characteristics in the quote above) or heterogeneity in results. Using this rigid procedure, a body of evidence with, say, four studies, and little heterogeneity in methods or results would still not merit a quantitative meta-analysis.	We now include quantitative MA for bodies of evidence with 3 or more studies when there is limited heterogeneity across studies OR when there are 5 or more studies with moderate heterogeneity. For bodies of evidence with 3 or 4 studies and large heterogeneity, we comment on why findings were not pooled with respect to the source(s) of heterogeneity across the 3-4 studies.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Methods	Figure 19 (page 70) reports the findings for change in depression symptoms for EMDR, compared to inactive control, based on four studies. I2 = 0%, visual inspection of the Forest plot demonstrates clear homogeneity of the point estimates of the study-specific SMDs. The pooled SMD estimate and its 95% CI are shown. If, however, the current rules of not doing quantitative meta-analysis for bodies of evidence with less than five studies were followed, none of those findings would be shown. Instead, the qualitative summary would report the individual point estimates and report that three studies reported statistically significant results and one did not.	We agree that the calculated I-squared indicated homogeneity of the estimates, but we do not believe that homogeneity can be assessed by this statistic, alone. We also considered variation in sample characteristics, study characteristics (length and components of interventions, when outcomes were assessed, etc.) and outcome measures used to make this determination.
Peer Reviewer 1	Methods	How is this considered an advance? Moreover, the current approach of doing a qualitative summary of the evidence when a body of evidence has fewer than five constituent studies seems to me to be a big step backwards.	We have modified our "rules" of pooling studies to include bodies of evidence with at least 3 studies when the degree of heterogeneity was low.

Peer Reviewer 1	Methods	The Cornell paper suggests that when a quantitative meta-analysis is not done: "A critical synthesis that highlights the variations in the evidence and describes the possible sources of variation will almost always be more useful than one that averages over these dimensions and can point the way toward improvement of future studies." However, that is not what the qualitative summaries of evidence in the current report actually are. For instance, in multiple places in the report, the qualitative summary seems to resort to the old (and invalid) approach of "vote-counting", i.e., stating how many studies found "positive" (i.e., statistically significant) results and how many found "negative" results. For example, on page 39, lines 11-16, describing loss of PTSD diagnosis of for exposure therapy versus cognitive interventions, the report states, "Three trials reported data on loss of PTSD diagnosis between exposure and cognitive intervention groups. Two studies favored exposure (RD range 0.08 to 0.16), but differences were not significant; one found a zero risk difference between groups (insufficient SOE)." There are many examples like this throughout the text. How can this possibly be better than presenting the results of a quantitative meta-analysis (with appropriate caveats and with use of appropriate methods, like empirical Bayes, that generate confidence intervals that capture the appropriate amount of underlying uncertainty due to heterogeneity)? This is definitely not the "critical synthesis that highlights the variations in the evidence and describes the possible sources of variation" recommended by Cornell. While there may be problems in doing quantitative meta-analyses when there are less than five studies, the current approach (qualitative review of evidence) seems to me to be much worse and potentially much more misleading. Have other organizations that do systematic reviews adopted a similar rigid cutoff for doing quantitative meta-analyses? Has the Cochrane Collaboration made a similar decision to not do quantitative meta-analysis when a body of evidence has less than five studies? I looked at the Cochrane Methods manual and the Cochrane methodological standards and I didn't see anything like that. Has GRADE recommended this?	We have modified our MA requirements (we now do them when we have at least 3 studies with low heterogeneity), but we argue that we do point the way "forward" in how we can use the information gleaned from 1 or 2 studies (or 3 or 4 studies with high levels of heterogeneity) where we do not report pooled estimates but instead give an overview of what was found in the evidence base to point the way "forward" in our Discussion section. We also give an explanation for the few instances where we had 3 or 4 studies in an evidence pool that were too heterogeneous to combine. We note the source(s) of heterogeneity in these instances in the text. Furthermore, these decisions are supported by the new report that recommends considering NOT pooling studies for various reasons (heterogeneity being one of them) https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/methods-guide-quantitative-synthesis-update.pdf We have cited this newly published report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Methods	Finally, particularly since this is an update to a previous SR, the reader is left in the awkward position of having meta-analytic summary estimates from the 2013 report for some treatments for some outcomes but having no summary estimates for even more data from more studies for some treatments from the current report. Why limit the usability of the report or make it difficult for readers to compare the findings in the 2013 report and the current report? I strongly urge you to reconsider the rigid decision to not do quantitative meta-analyses when a body of evidence has less than five studies. If a body of evidence has less than five studies and there is strong heterogeneity in methods or results, then it may be reasonable to not do a quantitative meta-analysis. If, however, a body of evidence does not show a large amount of heterogeneity, I think it would be much wiser to do a random-effects meta-analysis and incorporate one of the methods, like empirical Bayes, that incorporates appropriate uncertainty due to heterogeneity.	We relaxed the rigidity in this revision to include bodies of evidence with at least 3 studies of low heterogeneity. We argue that the synthesis in the update is more thoughtful than the prior, which pooled 3 or more studies without as much consideration of the characteristics of the individual studies being pooled. We have modified to presenting MA pooled estimates when the evidence included at least 3 studies of low heterogeneity.
Peer Reviewer 1	Methods	Why did this SR not include sensitivity analyses like the 2013 SR? The current report states that Appendix F (actually that is an error—it should be cited as Appendix G) "lists each study rated high risk of bias that met the inclusion criteria, along with details on the consistency between the findings from the studies rated as having high risk of bias with those from studies rated as having low or medium risk of bias for each intervention, comparator, and outcome combination reported." However, I reviewed Appendix G and see that it does not include sensitivity analyses at all. Appendix G does provide a qualitative assessment of what the high ROB studies show, compared to what the low and medium ROB show (in terms of effect magnitude and precision). But if you're going to summarize the data like that, why not just do a quantitative synthesis and sensitivity analysis? Sensitivity analyses that assess the	We have corrected this error in appendix citation. The decision to not include a quantitative sensitivity analysis was made because we did not believe in the utility of pooling findings from poor quality/high risk of bias studies with those with better study designs. We elected NOT to conduct quantitative meta-analysis when quality is heterogeneous across studies. We instead present a thoughtful qualitative synthesis that comments on the consistency of the evidence base of the studies with high risk of bias as compared to those with low or medium risk of bias. These decisions were made in line with the recently published guidance on pooling studies from AHRQ: https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/methods-guide-quantitative-synthesis-update.pdf

Commentator & Affiliation	Section	Comment	Response
		impact of including high ROB studies are incredibly important and I'm disappointed that they are not included. For the 2013 SR, many naysayers criticized the ROB ratings of one study or another and challenged the validity of the findings in the SR on those grounds. The inclusion of sensitivity analyses (and the finding that, in the vast majority of cases, the substantive conclusions were unchanged by inclusion of high ROB studies) was the best evidence to support the findings of the SR. Critics can and often do challenge the ROB ratings that were done by the group that did a SR. If sensitivity analyses are not included in the current SR, it will make it easier for someone who disagrees with the ratings of even a single study that was rated high ROB (and therefore excluded) to question the validity of the findings for a given intervention. Yes, the qualitative summaries of the findings in the high ROB trials reported in Appendix G will help address concerns of critics. But not as much as a sensitivity analysis would.	
Peer Reviewer 1	Methods	I agree with your decision to not conduct meta-analysis to obtain pooled estimates of subgroup effects by patient characteristics or type of trauma exposure.	N/A
Peer Reviewer 1	Methods	Is there a reason that you didn't consider doing a network meta-analysis for psychological interventions that used waitlist controls as comparators? I understand that a network meta-analysis would not be appropriate for those studies that used treatment as usual (TAU) as the comparator because TAU can differ dramatically from one study to the next (thus undermining one of the required assumptions that the comparators in a network meta-analysis be the same). However, waitlist controls are the same from one study to the next.	Yes, we considered it but determined that the interventions themselves (and often times, the comparators) were so heterogeneous that a network meta-analysis would not be warranted.
Peer Reviewer 1	Methods	Comments on specific sections in the methods: Page 12, line 14: I believe that "Appendix F" should be changed to "Appendix G"	We agree and have changed.
Peer Reviewer 1	Methods	Page 12, line 42: "rate difference" should be "risk difference"	We agree and have changed.
Peer Reviewer 2	Methods	Inclusion/exclusion criteria justifiable and clear.	N/A
Peer Reviewer 2	Methods	Research strategies well and clearly stated.	N/A
Peer Reviewer 2	Methods	I am not a clinician, and don't feel comfortable commenting on the choice of outcome measures from that perspective.	N/A

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Methods	However, the effect size in any RCT is strongly affected by the quality of the outcome measure.	N/A
Peer Reviewer 2	Methods	Yes, certainly its test-retest reliability in the population sampled, and reliability is among your criteria (although it is not clear that this is test-retest reliability, not, for example Cronbach's alpha which is often erroneously called "reliability").	We did not include Cronbach's alpha as a measure of reliability.
Peer Reviewer 2	Methods	However, also important is the sensitivity to individual differences among the patients Thus an ordinal measure of symptoms repeated m>2 times over the treatment course is better than the same ordinal measure done only pre and post, and a pre-post difference may be better or worse than simply the post measure depending on its test-retest reliability. Dichotomizing that measure will always decrease the effect size and attenuate power. Where one dichotomizes may have very different effects. Thus I am concerned about what seems equal emphasis on measures of symptom reduction (ordinal) and loss of diagnosis (dichotomized), as I am about multiple outcomes from a single study.	We did not attempt to give "equal" meaning to the two measures of symptom reduction and loss of diagnosis included in our review. We simply reported what the trial reported in its text.
Peer Reviewer 2	Methods	One way others have dealt with this is to choose no more than one outcome per RCT, with a priority defined. Thus if symptom reduction is used, loss of PTSD diagnosis will not be. I do recognize the arguments against this.	We agree this could be a good solution if we had many trials that reported multiple PTSD-related outcomes, however, very few actually reported symptom reduction or loss of diagnosis in addition to other continuous-type measures.
Peer Reviewer 2	Methods	I have a problem with mixing "statistical significance" with "clinical significance". p-values should not be reported, and "vote-counting" (how many studies reported p<.05?) should not be reported. A "statistically significant" result, p<.05, merely means that the sample size was large enough to detect some deviation from the null hypothesis. That deviation may be clinically trivial, or huge. No matter how small the deviation, there is always a sample size large enough to give 80% power of getting p<.05. One of the advantages of meta-analysis, which moved the field away from "vote counting", is the emphasis on effect sizes. If a RCT does not give the information necessary to compute an effect size, it should be excluded from consideration. With binary (dichotomized) outcome, Risk Difference (RD) is a fine effect size. NNT=1/RD is a very useful effect size to report for clinical consideration.	We have focused primarily on effect sizes in this revision. We only include studies that included information to calculate an effect size. We also do not count the number of studies with statistically significant findings in this review.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Methods	With normally distributed outcomes, SMD (Cohen's d) is a fine effect size, as long as the variances in the two groups are near equal, or the average variance in the two groups is used to standardize the mean difference. Then NNT=2normsdist(d/sqrt(2))-1, where normsdist() is the standard normal distribution function and sqrt() is the square root. However, mean difference (weighted or not) is not an interpretable effect size for clinical use. Suppose the mean difference is 5. If the within-group standard deviation is 1, this an enormous effect, with almost no overlap between the two response distributions, and NNT essentially 1.0, as good as it could get. Almost every patient given treatment will be better off than every patient given Control. However, if the within-group standard deviation is 50, then SMD=.1 and NNT is about 18, i.e., roughly speaking 17 or every 18 subjects given Treatment would do just as well with the Control treatment. Yet you would be reporting these two as giving the same result if you report MD or WMD. I would urge that every MD and WMD be converted to a SMD, and if the RCT does not report sufficient information to compute the SMD from the MD or WMD, it should be excluded.	Thank you. We now report SMD for continuous outcomes. We also have modified WMD to SMD throughout the report.
Peer Reviewer 2	Methods	I'm a little worried too (from a few comments in the Appendix) that some of the MD or WMD came from studies in which some baseline variable was "controlled for" or "adjusted for". The inclusion of a covariate changes the research question, from the effect size in the total population sampled to the effect size for some subpopulation matched on that covariate. If different RCTs used different covariates, they are reporting effect sizes on different subpopulations. Such effect sizes should not be mixed together.	We used changes in raw scale scores whenever the author reported them to inform the meta analyses.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Methods	p. 38: In the section on inclusion criteria, I strongly recommend reconsideration of the decision to include studies in which not all participants have PTSD. If we had good evidence that individuals with full vs. partial PTSD responded to treatment in a similar way, this decision would be justifiable. We do not have this kind of evidence, and the vast majority of studies (e.g., 170 of the 184 included in the review) include only people who met full criteria. One might say the small number of mixed studies could therefore be ignored, which would be a reasonable argument if there were not one treatment for which both of the available studies included partial PTSD: TARGET (references 59 and 60). As a result, the data for this modality are of unknown comparability to the data for other modalities. An additional reason for revising this inclusion criterion is to enhance comparability and relevance to existing guidelines. I recommend that the authors remove the 14 studies in question. If the authors want to include something about the partial PTSD studies, perhaps adding an Appendix that shows how estimates for the categories in which the studies would appear would have changed if the studies had been included.	We have retained these studies after careful consideration. All studies included a majority of participants with clinical PTSD and the portion who did not had subthreshold levels (therefore, were experiencing some level of symptoms, which presumably could be improved with treatment). In our revision, we do include information about the proportion of the sample with clinical PTSD in the sample characteristic table for each of our 13 studies that included a portion of participants without clinical PTSD (one study included in our last version has been excluded due to a retraction).
Peer Reviewer 3	Methods	p. 38: In the section on inclusion criteria, it is surprising that relaxation is included as an "intervention" given that it is only used as a comparison treatment, and that present-centered therapy was not included in this way. PCT is included in the review as an active comparator, but it is not treated as a bona fide treatment, even though it is recognized as such in the APA's Division 12 list of evidenced-based treatments for PTSD and in the VA/DoD PTSD guideline. Th ere are 2 studies showing that PCT is effective relative to WL (Classen et al., 2011; McDonagh et al., 2005), as well as a 3rd that included full and partial PTSD (Ford et al. 2011; ref 59 above). There are also more studies comparing it to a variety of other therapies. Even if relaxation were not treated as an intervention, PCT should be included as an intervention in this review.	We defined relaxation training as an intervention of interest a priori; it was in our list of interventions (and comparators) of interest. PCT was not on our inclusion list in this updated review.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Methods	p. 40: Risk of bias coding is very challenging, and is something that the field could better standardize to enhance validity as well as within and between-review reliability. One recommendation I urge you to consider is downgrading the use of LOCF as a method of performing ITT analysis. The appendix on this by Goodman in the 2007 IOM report on PTSD treatment shows how much bias can result from this practice when missing data are > 10% of a sample, which is almost always the case. I realize that the use of more robust techniques such as hierarchical models and MI is not always justified or well done, but LOCF is pretty much always going to lead to problems.	We looked over the included studies and determined that even if we considered LOCF as a reason to "downgrade", it would not have moved the risk of bias rating from medium to high for any of the studies included in the review.
Peer Reviewer 3	Methods	Also related to RoB coding, in Table E-25 in Appendix E, the entry for Schnurr et al. 2007 reflects another difficult issue. The coding of attrition and differential attrition is intended to reflect loss of data and differential loss that could produce bias. In many studies, data capture ends following treatment dropout, but studies have increasingly been following a better practice of continuing to collect data from treatment dropouts. Treatment dropout is therefore not the same thing as attrition from measurement. In the Schnurr et al. study, treatment dropout was 29% (38% in PE and 21% in PCT), but the actual attrition from measurement was only 13%, with no differential between arms.	We agree that the concepts are not the same (e.g., treatment dropout versus loss to follow-up). In general, we erred on the side of caution by taking the highest rates of loss to follow-up or treatment dropout to (conversely) indicate the proportion of individuals who actually received the treatment and had outcomes measured afterwards.
Peer Reviewer 4	Methods	The report indicates the use of the random effects model. The report should indicate either the method for estimating the variance component for the random effect model (for example, DerSimonian & Laird estimate or other method) or the computer program used (for example, STATA, metafor in R).	We have added this information to the revision.
Peer Reviewer 4	Methods	The report indicates that STATA was used for estimating the network meta-analysis, and similar information should be provided for the random effects model.	We have now indicated that STATA was used for all meta analyses.
Peer Reviewer 4	Methods	In the discussion of the network meta-analysis, the report indicates that transitivity was examined but no details are provided.	We added that we evaluated studies for potential inclusion in NMA prior to including them, and did not find any concerning differences in the studies we decided to ultimately include.
Peer Reviewer 4	Methods	In addition, the report does not provide information about the consistency of the network.	We now give details on consistency and report findings in a new table.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Methods	Search strategies were clearly stated, and inclusion criteria were justifiable.	N/A
Peer Reviewer 5	Methods	Also, additional concepts from other standard (e.g., RCTs) criteria were used.	N/A
Peer Reviewer 5	Methods	So too were AHRQ methods, ROBINS-1 tool.	N/A
Peer Reviewer 5	Methods	Statistical methods were not used, but the methods to abstract, code and determine quality and bias of the individual studies was appropriate.	N/A
Peer Reviewer 1	Results	On page 11, you report that you accepted the ROB ratings for studies included in the 2013 SR. However, I identified some instances of studies that were included in the 2013 SR (and therefore not rated high ROB) that are not included in similar analyses in the current report. For instance, Neuner1 was included in the calculation of WMD for PTSD symptom reduction for narrative exposure therapy compared with inactive controls in the 2013 SR but does not appear in Figure 14 of the current report showing SMDs for PTSD symptom reduction for narrative exposure therapy compared with inactive controls. Why? There are other examples like that in the results. 1Neuner F, Schauer M, Klaschik C, et al. A comparison of narrative exposure therapy, supportive counseling, and psychoeducation for treating posttraumatic stress disorder in an African refugee settlement. J Consult Clin Psychol. 2004 Aug;72(4):579-87. PMID: 15301642.	That is true. We evaluated the full body of evidence for each treatment of interest and, in the update, do not include active comparators in any of the efficacy synthesis. The comparators include in the Neuner 2004 study were trauma counseling and psychological education, both considered active and therefore not pooled with findings of studies that used inactive comparators (waitlist, placebo, treatment as usual, usual care).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Results	KQ1a: On page 66, lines 11 – 16, there is a description of the study comparing efficacy and comparative effectiveness of EMDR, fluoxetine and place between those with childhood onset versus adult-onset trauma. The paragraph reports no group (child or adult onset) by treatment interaction in the study. However, the 2013 SR reports, for the same study, "At 6-month follow-up, more than twice the percentage of participants with adult-onset trauma than with child-onset trauma achieved asymptomatic functioning (75% versus 33%, respectively) in the EMDR group. No participants achieved this level of relief in the fluoxetine or placebo group." The description in the 2013 that I just quoted sounds like a group x treatment interaction. Was it not included in the current SR because they didn't do a formal group x treatment interaction test or for some other reason? It would be useful to explain why this was included in the 2013 SR but not in the current SR.	We note that we only include findings when differences in efficacy or effectiveness are reported via interaction analyses. This study did not report findings from such an analyses. No formal comparisons between subgroups were done.
Peer Reviewer 2	Results	I would have appreciated having every title of a table specifying not only the treatment (e.g., CBT) but which Control was being used. Generally I could find this in text, but had to look for it.	Various controls were used in some tables, but, generally speaking, efficacy in the table title means that only studies with inactive comparators are presented (e.g., waitlist, treatment as usual, usual care, placebo, etc.) and effectiveness in the table title means that active interventions of interest were compared.
Peer Reviewer 2	Results	When the total number of patients from 5 studies was 399, it makes a difference whether 1 study had 300 and the other 4 totaled 99, or each study have about 80, and then whether the treatment and control samples were balanced or not. Too much detail is not needed, but could you report instead the range of samples per study, e.g., 399 (20 to 300) or (79 to 80). That would give warning when some studies are abysmally underpowered, as they are here.	We include this information in the study characteristic tables about sample sizes for each arm of each study. We did not report this in the summary tables because we thought it would detract from what we were trying to make a simple, easily understandable summary of the findings for readers.
Peer Reviewer 2	Results	As noted above, I'd like to see all MD and WMD replaced by SMD and its confidence interval, and accompanied by NNT which can be computed from SMD and RD. No confidence interval for NNT because of its peculiar wrap-around scale.	We have replaced all WMD with SMD throughout the report.
Peer Reviewer 2	Results	I'd prefer that all p-values or "vote-counting" be omitted.	We have instead reported effect size ranges in most cases throughout the report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Results	I suspect that I'd prefer to see many of the studies reported as "Moderate" or "Low" excluded.	As described earlier, moderate or low SOE grading is very different from a high RoB rating. We evaluated high ROB studies separately.
Peer Reviewer 3	Results	In general, the reporting in this section is clear and accurate. My concerns are largely related to the consequences of methodological decisions p. 46: In Table 5, neurofeedback, relaxation, and hypnosis are incorrectly classified as CBT coping skills. None of these are CBT or are considered to be coping skills training—especially hypnosis.	We have moved neurofeedback and hypnosis to the Other Psychological Intervention section but kept relaxation in the CBT-Coping Skills section because it is a component of CBT and the prior review, to which this is an update, included relaxation studies in that section.
Peer Reviewer 3	Results	Also note that this classification is inconsistent with the classification in Table ES-1. Something like this can cause readers to dismiss many good aspects of the report because it appears that the raters doing the coding did not understand what they were coding. Come to think of it, I doubt the developer of MBSR would consider it to be CBT coping skills training either. I cannot think of another review that has classified these interventions in this way. This comment is also relevant to other parts of the text where these interventions are discussed in greater detail, e.g., p. 57.	We have revised to have the example of treatments shown in ES-1 reflect the types of interventions for which studies that met the review inclusion criteria are presented in the report.
Peer Reviewer 3	Results	p. 48: Table 6, ref. 44 for cognitive therapy is listed in the CPT row for depression symptoms.	We have corrected our error, thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Results	p. 71: Table 12, Engel et al., 2015 is not a study of psychotherapy. Participants were randomized to a multi-component collaborative care intervention and not to psychotherapy per se. They took part in varying aspects of the intervention, which even could have included referral to actual psychotherapy. The fact that there was a computerized CBT intervention available as part of the collaborative care package is not the same as if participants were randomized to definitely receive psychotherapy. And even if everyone received the intervention, the fact that it was delivered as part of a set of interventions would preclude drawing inferences about any one piece.	We agree that particular aspects of the study included different forms of delivery of the CBT but that, if applying our inclusion and exclusion criteria, it does indeed meet criteria for this review. All patients received optimized usual primary care (OUC) and one group additionally received the DESTRESS-PC intervention described as nurse-assisted, online self-management CBT. This treatment is cited by the authors as a validated treatment for PTSD. The self-management portion of the intervention has to do with how the intervention was delivered. Our criteria did not specify mode of delivery. The CBT-M interventions were, indeed, quite heterogeneous, with each intervention containing different components (in this study, the CBT-M intervention included educational info, strategies to manage anger and promote better sleep hygiene, info on how to perform and practice stress management strategies, and cognitive reframing activities, which were similar to components included in other CBT-M interventions. It is true that patients in the treatment group could have been referred to "actual psychotherapy", but so could the comparator group since members received "optimized usual primary care"
Peer Reviewer 3	Results	p. 89: for IRT, the Cook study (ref 128) is not included in the PTSD information with ref. 89 even though Cook reported CAPS scores. (This has implications for Table ES-1 on p. 17 and the main Table 28 on p. 129.)	The Cook study is not included in the references section because it used a psychological education comparator, which we categorized as an active treatment.
Peer Reviewer 3	Results	The sentence in the section on comorbid conditions indicating that the Cook trial used the Ham-D is incorrect, as is the reference to the HAM-D (ref. 89 is the other IRT study); Cook et al reported the BDI as a depression outcome.	Thank you. We have corrected this error.
Peer Reviewer 3	Results	Also, why aren't the depression data being reported in the summary tables?	We do report information on depression differences in summary tables.
Peer Reviewer 3	Results	p. 93: I believe that the paper by Wolfe et al. on how the patients with and without the dissociative subtype of PTSD respond to PE is relevant in the section on KQ 1a (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4830387/.	Thank you for your comment. We did consider this study but excluded it for wrong/no comparator. The analyses performed in the paper compared those with PTSD subtypes and those without with respect to outcomes across both treatment types (PE and PCT), not whether those with versus without dissociative types of PTSD differed with respect to treatment comparisons.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Results	The results section and Appendix H provide a number of Forest plots, but the decision process for presenting a Forest plot is not provided.	We have added an explanation in the Methods section detailing when we include forest plots.
Peer Reviewer 4	Results	Forest plots are not provided for every meta-analysis conducted in the main report, but there are Forest plots for sets of studies where a meta-analysis was not conducted. For example, Figure 14 appears in the main report even though a meta-analysis was not conducted. In Appendix H, there are some Forest plots for only two studies. The reader will naturally pay more attention to results that appear in a Figure in the main report and in Appendix H. The report should provide a short rationale for when Forest plots appear in the text and when they appear in Appendix H.	Yes, this is true. We determined on a case-by-case basis where displaying the findings pictorially were helpful in understand the findings, even when we did not conduct pooled analyses.
Peer Reviewer 4	Results	In many of the Forest plots, a p-value appears next to the value of I-squared for the mean effect. This p-value is not referenced in the text or table. I assume that this value is related to I-squared. I suggest that the 95% confidence interval for I-squared is provided as indicated in the AHRQ guidance for comparative effectiveness reviews instead of the p-value.	Yes, we now present I squared but not the p value for the I squared as per the guidance.
Peer Reviewer 4	Results	Note that not every Forest plot includes the value of I-squared for the mean effect. For example, Figures 21 and 22 do not have the value of I-squared included.	We have added I squared to all figures.
Peer Reviewer 4	Results	In the section on drug treatments, the meta-analysis pools results across drug types. For example, Figure 18 pools results across two drugs, providing a mean effect across both drugs. As a general reader (not a medical doctor), I would like a rationale for pooling the results across drugs here and in other analyses.	We present all findings by drug type but, in some instances, also pool studies that examined medications in the same drug category, which assumes the mechanism of action is similar (e.g., SSRI, antipsychotics, etc.) as the prior review had done. The pooled drug category data is not presented in the text but only in tables to allow readers to have the information. We agree there may be substantial heterogeneity across studies of different medications in a single category so do not mention the pooled drug category findings in the text.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Results	The amount of detail and characteristics of the study is appropriate. In fact there is more than enough detail to understand what the findings are and how they are relevant to each key question and key points are presented for the different types of interventions and therapies reviewed. That being said, however, there were too many results described and thus they became tedious to wade through and make sense of all the information presented. It was also difficult to find the relevant results/findings that would be needed to make a decision on an intervention or therapy.	We agree and have attempted to simplify, and summarized the findings in in the summary tables.
Peer Reviewer 4	Results: Describe the implications of the network geometry on the validity of the network meta-analysis.:	As acknowledged in the report, the network relies on comparisons between drug treatments and placebo. Few studies report on direct comparisons between drugs. Thus, all estimates of direct comparisons between two drug treatments are based on placebo comparisons. The network meta-analysis results must be interpreted with caution as indicated in the report.	Yes, as stated, we note this in the report.
Peer Reviewer 4	Results: Exploration for consistency/incon sistency Are the results reported clearly and accurately: If you answer no, please elaborate in the Other Comments text box at the end of the form.	No	We have added information about consistency to the revised draft report.
Peer Reviewer 4	Results: Describe the implications of the result of the analysis (or of diagnostics) on the validity of the network meta-analysis.	The report does not discuss diagnostics for the consistency/inconsistency of the network.	We have added this information to the revision.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Results: Describe the implications of the result of the analysis (or of diagnostics) on the validity of the network meta-analysis.	The report does indicate that transitivity of the network was assessed, but does not provide details of the assessment method.	We have added information about how we assessed transitivity and our findings.
Peer Reviewer 4	Results: Describe the implications of the synthesis of results on the validity of the network meta-analysis	As discussed, the evidence for the network meta-analysis relies on indirect comparisons between drug treatments as most studies report only on a single drug treatment compared to placebo.	We agree that the evidence for the network meta-analysis relies on indirect comparisons and we noted this in the report.
Peer Reviewer 1	Discussion	p107, section on "Implications for Clinical and Policy Decisionmaking". I understand why the EPC would recommend CBT-exposure and CBT-M as first-line treatments for PTSD. (The words "first line" are not used, but both therapies are the first ones mentioned (with the rationale being large magnitude benefit and high SOE for both treatments) after the sentence, "Nevertheless, choices must be made for patients in need of treatment".)	Yes, and we were not intending to recommend psychological over pharmacological interventions. We were merely stating the interventions that had high SOE for some key outcomes in support of their efficacy. We have revised the report to make sure this is clearly communicated.
Peer Reviewer 1	Discussion	I have concerns about giving these specific psychological interventions or indeed any psychological interventions preference over medications on the basis of the current evidence. The current SR concludes, rightly, that the SOE for KQ3 (psychotherapy vs pharmacotherapy) was insufficient, since there was only one medium ROB trial that includes a head-to-head comparison of any psychological treatment to any medication treatment.	We were not intending to recommend psychological over pharmacological interventions. We have revised the report to make sure this is clearly communicated. In addition, we have added "psychological" before treatments to say "helpful psychological treatments" for clarification.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Discussion	The current recommendation for psychological interventions over medications is therefore based not on head-to-head (direct) evidence but on the comparative magnitude of the beneficial effect (i.e., the size of the effect estimates) and the strength of evidence. The problem in using those two factors as the yardsticks for determining which treatment to recommend is that they depend on characteristics of the studies that are different for psychological interventions and medication interventions. The first way in which they differ is masking (blinding). None of the participants in the RCTs of psychological interventions were blinded while all of the participants in the medication trials were. As Huhn2 reports, three meta-analyses of treatments for mental disorders demonstrated larger effect sizes for RCTs without blinding than for RCTs with blinding. The second way in which they differ is in the use of contemporaneous versus historical controls. By definition, a wait-list control is not a contemporaneous control. All of the medication trials used contemporaneous placebo controls. Huhn2 reports that eight meta-analyses of psychological and pharmacological treatments for mental disorders showed larger effect sizes for non-contemporaneous controls (wait-list) than for contemporaneous controls (placebo controls, treatment as usual and ineffective treatment). Finally, as reported in Huhn2, in the research literature for treatment of depression, the effect size for psychological treatments compared to inactive comparators (0.67) was more than twice the effect size for medication treatments compared to the inactive comparator of placebo (0.31) but the effect size for the head-to-head comparison of medications to psychological treatments was near null (0.05). For all of these reasons, I believe it is incorrect to recommend psychological treatments over medication treatments. It is unfortunate that there have not been more studies comparing psychological treatments to medication treatments for PTSD but that is the state of the evidence at this time. 2 Huhn, M., Tardy, M., Spineli, L. M., Kissling, W., Förstl, H., Pitschel-Walz, G., Leucht, C., … Leucht, S. (2014). Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: A systematic overview of meta-analyses. JAMA Psychiatry, 71, 706-715.	We were not intending to recommend psychological over pharmacological interventions. We were merely stating the interventions that had high SOE for some key outcomes in support of their efficacy. We have revised the report to make sure this is clearly communicated.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Discussion	The future research section is right on target.	Thank you
Peer Reviewer 1	Discussion	Comments on specific sections: Page 111, lines 46-47: The sentence shown is not grammatical as shown: “The article, authored by “pioneer” creators of created an empirically based psychotherapy to treat PTSD …”	Thanks. We have revised.
Peer Reviewer 2	Discussion	Again, there are two aspects for future research. First, what treatments require more attention, versus what controls. (Personal opinion: I do not favor placebo/waiting list controls when there are effective treatments available, for I consider that withholding treatment from those needing it. In any case, comparing new treatments versus placebo can lead to recommending new treatments that are not as good as existing old treatments.) However, I'd like to see more emphasis on the SOE issue.	We agree. The strength of evidence grades were used to grade the evidence supporting each intervention of interest.
Peer Reviewer 2	Discussion	It should be made clear that studies that do not have a control group, are not randomized, do not have a well-justified 'a priori' outcome and a reliable way of measuring it, do not do analysis "by intention to treat", or use valid analytic methods (e.g., do not "control" or "adjust" unless that is the hypothesis to be tested), should be discouraged. I'd also discourage underpowered studies as wasting time, effort and resources, and imposing a research burden on PTSD patients they do not need.	All of these factors went into our risk of bias ratings. We did not accept any study designs other than RCTs (and large observational studies just for KQ4, adverse events), but we did not find any studies that met our criteria. All studies were required to have a comparator of interest. Reliability and a priori naming of the outcome
Peer Reviewer 3	Discussion	The discussion is reasonable given the review, but there are a few sections that need attention. p. 135: I question the statements on the applicability of the findings in light of the inclusion of studies that did not require participants to meet full PTSD criteria. I could not find where in the report there was evidence presented to support this claim, and as I indicated above, for the therapy TARGET, all of the evidence (2 trials) is based on mixed PTSD groups.	We have added some text to explain our decision to include studies with less than 100% of respondents having clinical PTSD (and the remainder having subthreshold) to this section. We have added some text to explain our decision to include studies with less than 100% of respondents having clinical PTSD (and the remainder having subthreshold) to this section.
Peer Reviewer 3	Discussion	This decision to include partial PTSD is quirky and unnecessary. It does not strengthen the review, and instead weakens its applicability because of the unknown influence of partial PTSD on findings. Saying the results of the trials that included partial PTSD were “generally consistent” with the full PTSD trials does not go far enough.	We have added some text to explain our decision to include studies with less than 100% of respondents having clinical PTSD (and the remainder having subthreshold) to this section.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Discussion	The future research section is reasonable. It would be helpful if the authors were able to prioritize the long list of items, however, because some gaps are more critical than other. For example, given the range of effective treatments for PTSD, is the lack of evidence for some treatments as important as the lack of evidence about comparative effectiveness, or what works for which patients? This may be asking the authors to go beyond the evidence, but even discussion of the fact that not all gaps are equally important could be useful.	We have noted what we see are priority items. These are important questions that the literature has not adequately addressed as we say in the report.
Peer Reviewer 4	Discussion/Conclusion	The discussion and conclusion section are clearly written.	N/A
Peer Reviewer 4	Discussion/Conclusion	The report is careful to make conclusions based on the available evidence and makes important suggestions to strengthen the evidence base.	N/A
Peer Reviewer 4	Discussion/Conclusion	On page 105 of the main report, there is a reference to the recent APA report on PTSD treatments. I could not find this reference as numbered.	We have corrected this reference.
Peer Reviewer 4	Discussion/Conclusion	On page 111, the report references a new report by a "pioneer" creator of a new psychotherapy for PTSD. The reference for this report is not provided.	We have added this citation.
Peer Reviewer 5	Discussion/Conclusion	The Implications are generally clearly stated as are the limitations of the study.	N/A
Peer Reviewer 5	Discussion/Conclusion	Again though the reader either has to wade through the discussion to find what is relevant/meaningful to them because of the level of detail provided; however, headings for relevant sections are helpful for the reader to get a sense of what will be discussed and if focusing within each section independent of others, to the extent necessary, the information is well conveyed.	N/A
Peer Reviewer 5	Discussion/Conclusion	After having presented so much information, the conclusion was very succinct and generally captured the findings.	N/A
Peer Reviewer 1	Clarity and Usability	Clarity is excellent.	N/A
Peer Reviewer 1	Clarity and Usability	The conclusions are relevant to practice decisions but usability would be improved by: 1) including SMDs for all bodies of evidence (in addition to WMDs);	We have replaced WMDs with SMDs as previously noted.
Peer Reviewer 1	Clarity and Usability	2) including sensitivity analyses;	We have not added quantitative sensitivity analysis because we did not feel it helpful to pool poor quality studies with those included in our review of high and medium quality.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Clarity and Usability	3) Using trauma-focused CBT and non-trauma-focused CBT categories instead of the CBT-M category used in the systematic review;	We did not recategorize our CBT-M group to be consistent with the prior report that we built our update upon and because of the difficulties in determining trauma and non-trauma focused interventions within this category and across other categories. We also only had 2 of 31 CBT-M studies that were non-TF CBT.
Peer Reviewer 1	Clarity and Usability	4) including quantitative meta-analyses for bodies of evidence that included fewer than five studies, in the absence of large amounts of heterogeneity.	We have added quantitative meta-analysis for bodies of evidence with at least 3 studies with limited heterogeneity.
Peer Reviewer 2	Clarity and Usability	To the first two questions and the last, yes.	N/A
Peer Reviewer 2	Clarity and Usability	As noted above, MD and WMD are easily misinterpreted for policy or practice decisions, and should be replaced with SMD and its confidence interval, plus NNT.	We have replaced WMDs with SMDs.
Peer Reviewer 3	Clarity and Usability	Overall, the review is helpful and can be made even more with attention to the concerns that I and other reviewers have raised.	N/A
Peer Reviewer 4	Clarity and Usability	As mentioned in other comments, I would like the report to provide a clear rationale for when Forest plots are presented in the text and in Appendix H. Providing the Forest plots signals to the reader that these studies and associated analyses are important, and it is not clear in the report how decisions to present Forest plots were conducted.	We have added a rationale for when we present forest plots to the Methods section.
Peer Reviewer 5	Clarity and Usability	The report is well structure, organized and the continuation of multiple headings through each major section (introduction, methods, results, discussion and conclusion) is helpful for the reader and the reviewer. This structure facilitates the reader in identifying the relevant content area(s) and/or question(s) for which they seek answers to without having to otherwise wade through the major sections to find what if of most interest.	N/A
Peer Reviewer 5	Clarity and Usability	Conclusion are relevant and will likely have policy and practice implications.	N/A
Peer Reviewer 5	Clarity and Usability	There is new information that is contributed to the field and additional understanding especially as to the benefits, lack of benefits. or inconclusive evidence of interventions and/or therapies for treating PTSD.	N/A

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	General	I have substantial concerns about the following methodological decisions used by the RTI-UNC EPC to complete this SR: 1) use of the category CBT-mixed rather than separating CBT into trauma-focused CBT and non-trauma-focused CBT;	We did not recategorize our CBT-M group, for one reason, to maintain consistency with the categorization used by the prior report we were updating and because of the difficulties in determining trauma and non-trauma focused interventions within this category and across other categories.
Peer Reviewer 1	General	2) not doing quantitative meta-analysis unless a body of evidence had at least five studies;	We have added quantitative meta-analysis for bodies of evidence with 3 or more studies that had limited heterogeneity.
Peer Reviewer 1	General	3) not doing quantitative sensitivity analyses. Those methodological decisions will significantly affect the usability of this systematic review by guideline panels and by policy makers. I think that the current work was done carefully and thoughtfully and that the results are likely to be valid. I hope that the RTI-UNC EPC will reconsider those methodological decisions so that this work can be more useful to guideline panels and policy makers.	We have not added quantitative sensitivity analysis because we did not feel it helpful to pool poor quality studies with those included in our review of high and medium quality.
Peer Reviewer 2	General	By current statistical standards, the composition, exposition, completeness, clarity of the report is excellent. If distributed as is, it is valuable. However, such a report may be up to the highest current methodological standards and still potentially misleading to clinicians and patients reading the report. It is this issue that I primarily focused on when reviewing the report, for the preface suggests that clinical impact should be the primary consideration.	N/A

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	General	The report is clinically meaningful, the target population and audience are explicitly defined, and the key questions are appropriate and explicitly stated. However, the impact and relevance would be strengthened by revisions of some methodological decisions, as indicated in the following sections. All of my comments refer to the pdf numbering. The references need careful attention. For example, on p. 16, 5 studies of CPT are mentioned for PTSD symptoms, yet only 1 study is cited (Monson), and then for depression 6 are cited, but ref. 6 is Ehlers et al., which is Cognitive Therapy, not Cognitive Processing Therapy. In the main text on p. 48, there are 5 references cited for CPT for PTSD (39-43), but the incorrect Cognitive Therapy citation of Ehlers is included for depression (39-44), even though the table correctly identifies 5 studies. As another example, on p. 20, the reference 84 to the VA/DoD guideline is actually Reich's risperidone study. Reference 84 in the overall references is to a van der Kolk study. Reference 84 in the overall references is to a van der Kolk study. The references are not correct for the references to the APA guideline either.	We have carefully checked all references in this revision and made sure all references are correct.
Peer Reviewer 4	General	The report is clearly written for a general audience.	N/A
Peer Reviewer 5	General	Overall, this report was relevant and provided additional information that was clinically relevant and that would be meaningful to the field. I do not recollect the audience being explicitly defined in the review. The target population that was included in the review; the articles included as well as the PTSD population was explicitly defined. Key questions however are appropriate and explicitly stated.	N/A

Commentator & Affiliation	Section	Comment	Response
Public Reviewer 1 (Yuval Neria, Columbia University Medical Center, Department of Psychiatry)	General	Hello .Overall, I think it's a decent report. Yet, I don't agree with the decision not to include IPT for PTSD (see attached) since: -IPT is comparable to CBT for PTSD patients, particularly when PTSD is comorbid with MDD (up to 50% of the cases)	We did include IPT for PTSD.
Public Reviewer 1 (Yuval Neria, Columbia University Medical Center, Department of Psychiatry)	General	-Recent JAMA review (Steenkamp et al., (2015). Psychotherapy for military-related PTSD: A review of randomized clinical trials. JAMA, 314(5), 489-500) found CBT and CPT treatments NOT superior to comparison treatments (attached) in veteran settings (VA and DoD), whereas IPT can be particularly useful. In the veterans Clinic I run at Columbia University Medical center, we utilize IPT quite often, and successfully.	This is a systematic review, not a primary study so the JAMA SR is excluded from this report. We did, however, check references listed in the SRs we identified.

Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	We appreciate the opportunity to review and provide comments on this draft report. Please note, these are not official comments from the American Psychological Association. Overall, this report is a useful update to the report published in 2013. The methods and results are clear and understandable. Out biggest concern centers on how psychotherapies are identified and categorized. Table ES-1 (page ES-1) lists the following as Psychological Interventions Cognitive behavioral therapy • Cognitive processing therapy • Cognitive restructuring • Exposure-based therapies • Coping skills therapy • Various "mixed" therapies Eye movement desensitization and reprocessing Other psychological or behavioral therapies • Psychodynamic therapy • Interpersonal therapy • Group therapy • Hypnosis/ hypnotherapy • Eclectic psychotherapy • Brainwave neurofeedback • Energy psychologyThen, on page 17, Table 5, 'Classes and categories of psychological treatments for posttraumatic stress disorder' the following are listed as other categories of psychotherapy • Brief eclectic psychotherapy • Emotional freedom techniques • Interpersonal therapy • Imagery rehearsal therapy • Memory specificity training • Narrative exposure therapy • Present-centered therapy or group present-centered therapy • Psychodynamic therapy • PTSD family education • Seeking Safety • Structured writing therapy • Trauma affect regulation But then on pages 19-20, Table 6, 'Summary of efficacy and strength of evidence of PTSD psychological treatments' provides data on only some of the other psychological therapies (BEP, IRT, NET, SS and TAR).	We have coordinated the interventions we describe in the introduction with those searched for and found for this review.

Commentator & Affiliation	Section	Comment	Response
		Further on, in the Summary on page 97, the report indicates low SOE for BEP, IRT, TAR, and SS and insufficient evidence for MCT, SIT, relaxation, neurofeedback, EFT, MEST, IPT SWT.In this summary, nothing specific is noted about group therapy, psychodynamic therapy or energy psychology from Table ES-1 or EFT, present centered therapy, PTSD family education, structured writing therapy while Meta cognitive therapy is now reported on. These changing groupings as well as changes in which specific types of therapies are reported in summary tables and paragraphs is problematic for the casual reader. Practitioners who provide these specific therapies will eagerly search for information and be frustrated by the changing grouping of psychotherapy and the challenge of readily identifying results for these treatments.	
Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	It would be helpful to have a table indicating more clearly what evidence- including none to be evaluated- for each of the identified therapies. One such place might be a summary table following the narrative on pages 62-63.	We feel that the current summary tables describe our main results and conclusions clearly. We note that interventions that did not have at least one outcome for which low SOE was available were not included in the tables. In the Discussion, we highlight evidence gaps in the current literature.
Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	Furthermore, it would be helpful to have Figures (similar to Figure 2, disposition of articles) that details for every therapy that was searched, how many articles were identified, excluded (and why), assessed for eligibility, included in review and the risk of bias. Those who routinely provide particular psychotherapies in practice are highly invested in reading the results for those treatments. If this information is provided in more detail in the appendices, please reference in the main body of the report.	We did not collect data at the intervention-specific level at all stages of our literature search (title and abstract review, full text review, risk of bias review levels.
Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	Additionally, the footnote for Figure 2 indicates how many new studies were identified but it would be useful to know how many new studies for each intervention were identified for inclusion in the review.	Some interventions had studies fall out of the review because of slightly different inclusion criteria. Some studies include more than one intervention type (potential for double counting).

Commentator & Affiliation	Section	Comment	Response
Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	In addition to clarifying the above, it would be useful to have reports specific to Prolonged Exposure (PE), separate from analyses on all studies evaluating exposure treatments.	We were updating the prior review, which included PE as part of the larger analysis of CBT-exposure therapies. Also, there were several different kinds of PE tested (e.g., COPE and variants of PE), which speaks to the myriad ways that these interventions could be categorized. We thus kept the category at the broadest level.
Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	The majority of the studies analyzed in the exposure category appear to be PE and it would be useful to have a summary of the results specific to PE, distinct from all studies that evaluated exposure treatments.	We do not present findings specific to different types of exposure therapies to be consistent with the original review we were updating. For each outcome of interest, the findings note which specific types of CBT-exposure included that outcome so that readers interested in a specific type of PE could make conclusions as needed.
Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	Table 1, Diagnostic criteria for posttraumatic stress disorder is helpful but it would also be useful to have some discussion of the implications of the changed DSM diagnostic criteria on the evaluated evidence base. Many studies were conducted using DSM-IV diagnostic criteria but some studies may have used DSM-5 criteria. Does this have any impact on the report conclusions?	We note that studies are needed to determine the impact of the changes in criteria as this information has not yet been published, to our knowledge. However, changes between DSM-IV and DSM-5 were minimal for PTSD, so we do not anticipate any appreciable difference in findings. Furthermore, we had several studies where a portion of the respondents had subthreshold levels of PTSD according to DSM-IV criteria. These same individuals, however, may have met new DSM-5 criteria (or vice versa, those with clinical PTSD according to DSM-IV may only have met partial criteria for DSM-5). These instances are likely to be few, however, because of the modest differences in criteria. The spectrum of PTSD symptoms is continuous in nature and not easily black and white with respect to who does and does not meet criteria.
Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	The prevalence data reported on page 3 discusses PTSD among military men and it would be helpful to contrast rates with military women.	We have added this information.
Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	Finally, while a network meta-analysis was conducted on medications, one was not conducted on psychotherapies. The reader may wonder whether network meta-analysis is an appropriate tool for evaluating psychotherapies and if so, why one was not conducted. Please provide at least a brief discussion on this matter.	We have added text to the report that provides a rationale for our decision NOT to conduct a network meta-analysis on psychotherapies to the Data Synthesis section of the Methods.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer 2 (Lynn Bufka, American Psychological Association)	General	One typographical error: ES-7, 6th line from bottom, need a space between “from” and “insufficient”	We have corrected.