Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: Child Exposure to Trauma: Comparative Effectiveness of Interventions Addressing Maltreatment

Draft review available for public comment from May 02, 2012 to May 30, 2012

Research Review Citation: Goldman Fraser J, Lloyd SW, Murphy RA, Crowson MM, Casanueva C, Zolotor A, Coker-Schwimmer M, Letourneau K, Gilbert A, Swinson Evans T, Crotty K, Viswanathan M. Child Exposure to Trauma: Comparative Effectiveness of Interventions Addressing Maltreatment. Comparative Effectiveness Review No. 89. (Prepared by the RTIUNC Evidence-based Practice Center under Contract No. 290-2007-10056-I.) AHRQ Publication No. 13-EHC002-EF. Rockville, MD: Agency for Healthcare Research and Quality. April 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program 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.

Source: http://effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1422
Published Online: April 15, 2013

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	General	Yes. The challenge is that the rigor that the authors to identify studies left a very small pool of studies for final analysis; this reflects where the field is, though I would guess that clinicians might also get useful information from trials that were, for example, conducted with mixed populations.	We agree with the reviewer and more clearly call attention to this issue in the report (see Scope and Key Questions, 2nd and 3rd paragraph, Executive Summary; Key Findings and Strength of Evidence, Overview, Discussion, 2nd paragraph). We acknowledge our exclusion decisions may have resulted in the exclusion of trials that might bolster evidence for included interventions or support inclusion of other interventions with at least low strength of evidence. Our intent was three-fold: 1) to reduce the noise of clinical heterogeneity that currently undermines the extant evidence base, 2) to maintain the rigorous approach for study inclusion that has been employed across AHRQ CERs, and 3) to avert yet more heterogeneity due to inconsistent, vague, or absent definitions of samples of children defined as 'at risk' or an admixture of 'at risk' and maltreated.
Peer Reviewer #1	General	The decision not to include trials with children over 14 is defended in the paper, though the ultimate reason this age group was not included (p Executive Summary-23: "in recognition of how maltreatment and its sequelae evolve across the development continuum") seems rather vague...	A clear rationale for the age cut-off decision is now provided in the Methods section of the Executive Summary and in the Limitations of the Comparative Effectiveness Review (1st paragraph) in both the Executive Summary and the main Report.
Peer Reviewer #1	Introduction	The authors are to be commended for an organized and readable review. I think that splitting off the executive summary from the longer more detailed review was sensible.	We thank the reviewer for this positive comment and affirmation that the Executive Summary is useful.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	Overall, the authors' criteria make sense--better to select studies that are more rigorous when guiding the field. In some instances, however, the criteria seemed to limit the capacity to guide clinical intervention. For example on Executive Summary-15 at line 55, that authors note that interventions like PCIT have "idiosyncratic and eclectic approaches" from a theoretical perspective because the interventions are based on two theoretical perspectives (in this case, attachment and social learning theory) rather than a single theoretical perspective. The focus on having a "unifying theory" is unjustified, in my opinion and it would be useful to include PCIT and Child-Parent Psychotherapy alongside the ABC and TF-CBT studies. To discount these intervention trials as some how on shaky theoretical footing seems unjust.	Our intent with a priori focus on theoretical orientation was to examine the literature comparing interventions that clearly ascribed to a particular orientation. This focus was not to elevate treatments with a unifying theory over multiply determined approaches but to facilitate meaningful comparison. However, in our review we found the literature paid scant attention to and provided limited description about theoretical orientation. Thus, KQ3 includes only two interventions (three trials). We note that Child-Parent Psychotherapy (CPP) was not included in KQ3 because the theoretical orientation of the comparator was not clearly described in either CPP trial. We also note that PCIT was not included in KQ3 because there was no head-to-head trial comparing PCIT with an alternative approach as an active control (the comparator was either an enhanced version of the PCIT condition or usual care). KQ3 (see Results section in both the Executive Summary and main Report) has been revised to more clearly describe our approach and the challenges we encountered investigating theoretical orientation as well as the other prespecifed intervention features (modality and setting).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	The studies are completely described with clear and focused Tables and text--all of which is well organized and detailed. I did not review the Tables in the Appendices in detail, as these seemed to be longer versions of those presented in the main section of the review.	No action needed.
Peer Reviewer #1	Discussion/Conclusion	I think the future research section flowed nicely from the limitations of the evidence section. The authors build a strong case across the manuscript for the limitations of the data (from small sample sizes to lack of power calculations to limitations in the validity of instruments to challenges in applicability and so on)...the review is exhaustive and well constructed. The challenge to me is that that p.140-150 is really the section most close to a "future research" section. This section is also well written but what is NOT addressed in this section are recommendations for things like ways to fund cross-site studies (which would increase power and move interventions beyond the originator). Can the authors advise agencies about how they might take networks like the NCTSN and try to create collaboratives that conduct cross site studies? Are there examples of this kind of work in other areas of health care or mental health care? What the author suggest is needed are studies with larger N's that look at particular subpopulations, document formally maltreatment types (which is actually quite challenging to do) and so on, but the sheer volume of factors that would make for a stronger research base suggests that changes in funding wish move the field away from single site, single intervention research studies is really what is needed. This is a paradigm shift in my view and this shift is only hinted at.	The report now provides a clear set of specific priorities areas for policy and research, resonating with the suggestions of the reviewer (see Implications for Policy and Future Research Needs sections, Discussion, Executive Summary; Implications for Research and Implications for Policy sections, Discussion, main Report). The topic of how to fund and create the infrastructure for collaborative clinical trials is outside the scope of this review and its recommendations.
Peer Reviewer #1	General	Usability may suffer some from the overall length of the report; having the appendices available separately (online, perhaps) so that the Executive Summary and main report represented the body of the report makes sense to me. Most clinicians would gasp at the full beast.	We agree with the reviewer, that the full report and appendices is overwhelming for clinical providers. Our goal is to that the Executive Summary stands alone stand and that the Executive Summary and main report are highly usable.
Peer Reviewer #2	General	On page 9 of the report, under the Fostering Healthy Futures section, it states that the program was “designed to foster resilience through the promotion of adaptive functioning in emotionally maltreated children.” The program was not designed for any one type of maltreatment, rather it designed for maltreated children placed in out-of-home care.	This has been corrected in the final report.
Peer Reviewer #2	General	On pg 57 of the report, under the Fostering Healthy Futures section, it states that, “Children in the Fostering Healthy Futures group showed greater improvements in trauma symptoms and behavior problems…” The program has not yet demonstrated improvements in externalizing behavior problems, although it has demonstrated improvements in internalizing problems. Just wanted to clarify.	This has been corrected in the final report.
Peer Reviewer #2	General	On page 62 of the report, “behavior problems” are listed as outcomes in three places. Please refer to note in #2.	This has been corrected in the final report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	General	On page 62 of the report, under Study Design and Duration, it states that the post intervention timeframe was "11 to 13 months." The post intervention timepoint was actually immediately following the intervention, or 11-13 months post baseline.	This has been corrected in the final report.
Peer Reviewer #2	General	On page 62 of the report, under Improvements in trauma symptoms, it reports that G1>G2, and that d=.30. Trauma symptoms were lower in the intervention group at the 6-month follow up, so this is incorrect.	The conventions for reporting (i.e., G1 > G2 means G1 did better than G2) have been clarified in the final report
Peer Reviewer #2	General	On page D-3 of the Appendix, on the Taussig, 2010 line, the table indicates that intent-to-treat analyses were not conducted, when indeed ITT analyses were conducted in this trial.	This has been corrected in the final report.
Peer Reviewer #2	General	On page E-41 of the Appendix, under Research Objective, it states that the objective was "improved quality of Life in 6 mos." The actual 6 mos objectives were broader and are listed in the paper.	This has been corrected in the final report.
Peer Reviewer #2	General	On page E-41 of the Appendix, under Study Duration, it states that the intervention lasted for 11-13 months. The intervention was actually implemented over a 30-week (or approximately 9-month period).	This has been corrected in the final report.
Peer Reviewer #2	General	On page E-42 of the Appendix, under Child Clinical Presentation, it states that scores on outcome scales at baseline are NR, but these numbers are provided in the article.	We have added a note in this section that tells the reader to look at the results tables for this information.
Peer Reviewer #2	General	On pages E44-46 of the Appendix, under Mental Health and Behavior (Part 3), there are several numbers that are missing, incorrectly placed, or have the wrong sign. On page E44 of the Appendix, in Mental Health and Behavior (Part 4), under "Recent MH therapy, adjusted follow up %," G2 is reported as 10, when it should be 71.	This table was reviewed and corrected for the final report.
Peer Reviewer #3	General	I think the authors did an excellent job in following the protocol for a Comparative Effectiveness Review, including accessing relevant research studies, carefully citing and applying inclusion and exclusion criteria, and describing individual studies that they included in the review. Most of my criticisms of the resulting document have to do with the methodology of the effectiveness review and interpretation of results of the summarization of included studies. Many of these criticisms the authors are aware of as they state them, particularly in the concluding sections of the documents on the limitation of the methods and the database. I will illustrate some of my criticisms with quotes from the literature to indicate I am not the only one with a certain perspective. I would strongly agree with their overall findings that based on the review that the "scientific evidence" for the effectiveness of intervention for "maltreatment" is thin and weak. I would also apply this conclusion to almost all interventions in child mental health. "Evidence" reduces uncertainty, but seldom eliminates it, comes in many different forms, and can vary greatly in quality: "In many respects, the field of child/adolescent psychosocial treatment of mental health problems is at too early a stage to make many evidenced-based	We did not intend to overstate the strength of the evidence. We have revised the report to emphasize that the vast majority of studies yielded a low strength of evidence. This point is made several times in the Discussion (see esp. Limitations of the Evidence Base and Future Research Needs) and Conclusions sections of both the Executive Summary and the Main Report. We are careful not to recommend one particular treatment over another, as the comparative evidence is as yet too limited to do so. The Abstract has

Commentator & Affiliation	Section	Comment	Response
		recommendations about which treatments show the most benefit for which psychiatric disorders. Conclusions regarding efficacy are limited by the fact that there are not enough large-scale and methodologically rigorous studies for many combinations of treatment and disorder." p. 424 Target, M. and Fonagy, P. The psychological treatment of child and adolescent psychiatric disorders. In A. Roth and P. Fonagy (2005) What Works for Whom: Second Edition. New York: Guilford, 385-424.) I myself tend to reduce the persuasiveness of intervention restudies for the following reasons: small sample size, research conducted by intervention developers ("allegiance effect"), high rate and differential attrition (despite the author's claim ITT does not make up for this type of estimation bias), numerous rather than one critical outcome measure (lends itself to "outcome shopping"), post hoc analyses, etc. The result is that any single intervention study does not merit any degree of certainty about the effective of an intervention. Only if large numbers of recipients have received an intervention as implemented by independent implementer can we begin to have a sense of the effectiveness of an intervention: Our findings suggest that the magnitude of the effect size of therapeutic and preventive interventions in mental health has changed considerably with the accumulation of new information…With 500 randomized subjects, it is not uncommon to see 1.5 fold changes in the OR <odds ratio> when more data appear. At least 2,000 randomized subjects would be required to diminish this uncertainty to a change of less than 25% in either direction…To conclude, our results indicate that evidence of effectiveness derived from only a few hundred participants should be appraised critically and carefully…" p. 1127,1129. Trikalinos, T. A., et al. (2004) Effect size in cumulative meta-analyses of mental health randomized trials evolved over time. Journal of Clinical Epidemiology, 57, 1124-1130.	also been revised accordingly. See also response to Reviewer #1
Peer Reviewer #3	Introduction	The very definition of maltreatment itself presents a challenge to researchers. Many of the included studies define maltreatment in terms of a child's involvement with child protective services—a criterion affected by state-level differences in how maltreatment is defined." p.147 There are 2 primary ways that children are identified as "maltreated" they are reported to the CPS systems and processed in a some manner, e.g., a determination of substantiation made, or they are identified through screening and/or assessment about their traumatic experiences, particularly, in CPS, CW, or other child-serving systems, e.g., schools. Mixing up these 2 types of maltreatment identification often leads to confusion and inaccuracies. The reasons children are substantiated in the CPS system vary widely in CPS systems, even within units within the same CPS system, and often are not dependent on the actual experience of abuse or neglect. In the CPS system interventions for "maltreatment" would primarily be interventions to prevent the occurrence of abuse and neglect and interventions to improve outcomes	We agree with the reviewer regarding the complexity of the population we undertook to study in this review. Defining the population to reflect the real-world while at the same time addressing clinical heterogeneity was perhaps the most difficult aspect of the review. We were surprised by the limitations in the literature in terms of specificity regarding the population. We recognize this limitation in large part reflects the challenges of collecting accurate

Commentator & Affiliation	Section	Comment	Response
		of children in the CPS system should primarily focus on system level interventions to reduce negative stress from investigation, finding, involvement of birth parents, and out-of-home placement. Unfortunately, the review excludes service system interventions from the review, which does not provide an adequate description of the types of interventions that should be implemented in the CPS system although review of such interventions might not be available as comparative RCTs. For children who are identified as having maltreated experiences (often just on the report of the child) it is often found that these children have multiple traumatic exposures and a major clinical problem is implementing clinical treatment that is responsive to their trauma histories but cognizant of the fact that they likely have other mental health issues that affect their mental health issues. Interventions for reported "maltreatment"-related mental health problems are covered in the effectiveness review in the description of interventions that target child "well-being." Sometimes these 2 different "maltreatment" approaches are crossed in that CPS might contract with a mh provider to do screening and assessment of children who are substantiated and/or placed in out-of-home settings , but these are overlapping but different child "maltreatment" populations.	and complete background information, particularly for children in the child welfare system. We have revised the report to more strongly emphasize the complexity of identifying maltreated children (see Definitions, 2nd paragraph, under Introduction in the main Report; Research Gaps, 2nd paragraph, Discussion, in both the Executive Summary and main Report). Specifically, we strengthen language regarding the considerable heterogeneity in the way study populations are defined in the literature and clarify our intention to reduce heterogeneity by excluding studies with widely heterogeneous populations. Additionally, we clarify that the population included children involved with child protective services (i.e., not necessarily substantiated). See PICOTS table, Methods, in both the Executive Summary and the main Report. We recognize that system-level interventions (i.e., strategies at the service-delivery or organizational level) leaves out a number approaches that are highly relevant to the CW system. We note their relevance in the Limitations of the Comparative Effectiveness Review section, 2nd paragraph, in the Discussion (Executive Summary and main Report). We also provide readers with a definition of these

Commentator & Affiliation	Section	Comment	Response
			approaches in the Intervention section, Methods chapter, main Report. Based on our review of the literature, we had serious concerns about the generalizability of these approaches because they are so diverse (and, indeed, diffuse) in terms of strategies and also population - typically targeting families with children birth through adolescence, with minimal specificity about the study population.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Methods	The authors use standard narrowly-conceived statistical methods and a "classical/frequentist" approach(as (mis)interpreted by behavioral researchers) to summarize and interpret results of their analysis of the studies. Thus, one of their major criteria for inclusion of studies and "strength of evidence" is whether or not the studies have weaknesses that would "bias" the estimation of the "true effect" of an intervention. The notion that an intervention has a "true" effect across a range of populations, contexts, implementations, etc. has never made much sense to me: Finally, Sinn refers to 'the true effect of treatment.' We are not sure what he means by this. We discussed the effect of treatment on a particular measure that almost surely, given that we are in the context of RCTs, does not have perfect reliability and validity. Thus, we cannot know 'the true effect of treatment.' We can only estimate the effect of treatment in a certain population (represented by the sample) for a certain outcome (the one selected). We hope that Senn is not suggesting that since we can never know the 'true effect of treatment', we should not be evaluating treatments at all. Acion, L. Author's Reply . Statististics in Medicine, 2006, 25, 3944–3948 They often seem to accept the "significant results" provided by an individual study represent the literal "truth" about the effectiveness of an intervention: Perhaps worst is the practice that is most common: accepting at face value the significance verdict as a binary indicator of whether or not a relation is real. What drives all of these practices is a perceived need to make it appear that conclusions are being drawn directly from the data, without any external influence, because direct inference from data to hypothesis is thought to result in mistaken conclusions only rarely and is therefore regarded as 'scientific.' This idea is reinforced by a methodology that puts numbers-a stamp of legitimacy-on that misguided approach." p. 1002 Goodman, S.N. (1999) Toward evidence-based medical statistics. 1: The p value fallacy. Annals of Internal Medicine, 130, No .12, 995-1004. The authors could benefit from a good dose of Bayesian thinking: With Bayesian analysis, assertions about unknown model parameters are not expressed in the conventional way as single point estimates along with associated reliability assessed through the standard null hypothesis significance test. Instead the emphasis is on making probabilistic statements using distributions." p. 4 Gill, Jeff (2008) Bayesian Methods: Second Edition. Boca Raton, Fl.: Chapman and Hall.	We agree with the issues raised by the reviewer regarding estimation of the "true effect" of an intervention. We would note that this issue is addressed, in part, by the "low" strength of evidence grade (see again criteria for evaluating the strength of the evidence in the Methods section of both the Executive Summary and main Report). We have added language calling attention to the limitations of frequentist data (note that that we did not perform a quantitative meta-analysis). See Future Research Needs, Statistical Considerations in the Discussion of the Executive Summary and the main Report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Results	The major weakness in most behavioral intervention research and in most of the studies reviewed in this effectiveness review is the small sample sizes: Psychologically, there is a tendency to underestimate the degree to which means (or other calculated statistics) may vary from one sample to another. We are inclined to endow quantities we calculate with an aura of exactitude and finality, not worrying enough about the results than might obtain were other samples to be drawn. This proclivity is particularly misleading when the size of the sample is small: Twersky and Kahneman(1971) coined the phrase, 'The law of small numbers,' to refer to the tendency to impute too much stability to small-sample results." p. 27 Abelson, R. P. (1995) Statistics as Principled Argument. Hillsdale, N.J.: Erlbaum The authors do report on this limitation numerous times in the report, but I would give it much greater weight in evaluating intervention studies. The authors do identify one of the major limitation of small sample sizes in that it is difficult to statistically identify moderate effects in small samples because of "low power." particularly in comparative studies (the "dodo bird effect."), but the The Cohens in an early study identified the "charybdis" of exaggerated effects to the "scylla" of low power to detect effects in that if you do find a significant effect in a small sample it is likely to be much larger than the effect in the population because of being an unrepresentative sample. (Cohen, P., Cohen, J., and Brook, J.S. (1995) Bringing in the sheaves, or just gleaning? A methodological warning. International Journal of Methods in Psychiatric Research, 5, 263-266, p. 266.) The authors on several occasions regret the absence of subgroup analyses. Again this is a major limitation of small sample research as subgroups (crossed with different interventions) represent interaction terms in statistical models and estimating these interaction terms require very large samples. There is little hope that the small scale RCTs in child mental health will legitimately have the number of participants to investigate subgroup effects (e.g., effects of mh severity, client characteristics, provider variation) that are critical in interpreting whether an intervention is appropriate for different types of clients.	The Strength of Evidence Grading section in the Methods and the Limitations of the Evidence Base section in the Discussion (Executive Summary and main Report) address the limitations of single studies with small sample sizes. We explicitly discuss the need for high quality, large-scale, multisite trials that can assess treatment moderators in the Implications for Policy and Implications for Research sections of the Discussion in the Executive Summary and the Report. This point is raised again in the Conclusions section (Executive Summary and main Report) and the Abstract.
Peer Reviewer #3	Discussion/Conclusion	The authors in their final section have good discussions of the limitations of the research they investigated. However, they do not offer any discussion of the limitations of RCTs in that they are: • Hard to do well • Require large samples to be done well in particular they do not lend themselves to subgroup analysis • They only estimate average effects which does not tell us much about who interventions work for and who they don't work for nor how they work Results of RCT usually do not generalize to any population because they typically are implemented with convenience sample in a specific restricted service setting (e.g., an academic clinic)	We strongly acknowledge the challenges in conducting high quality RCTs with this particularly vulnerable population. This issue is raised in the Research Gaps section, 1st paragraph, Future Research Needs section, 1st paragraph, and Conclusions section, 1st paragraph in the Executive Summary and Main Report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	General	I do not think the results of the review is very informative to the practice community because it only focuses (and in excessive detail) on results of research studies. The practice community has very limited understanding of the technical aspects of behavioral research, particularly its limitation, and of "statistical analysis" (as if this were one thing). A concern given the weak and thin evidence reported in this review is that if policy makers or payer also with limited understanding pick one or two of the interventions that received some favorable comments as "evidenced based" they will attempt to require use of these intervention when they are not appropriate and or effective.	We agree with the reviewer that we must be careful not to suggest that there is strong enough comparative evidence to make specific treatment recommendations. This point is made in the Conclusions and Abstract. Additionally, the report now makes clear that the goals of this AHRQ-commissioned review are to provide stakeholders with a synthesis of the best evidence in the field of child maltreatment and to identify critical areas to address in future intervention research. See Background: Condition and Therapeutic Strategies, 2nd paragraph in Executive Summary, and Introduction, 2nd paragraph, main Report). We also revised the Implications for Clinical Practice section in the Discussion (Executive Summary and main Report) to suggest ways that the report may factor into clinical decisions regarding treatment selection. An important part of the AHRQ review process for many CERs is that the Eisenberg Center often translates full reports into separate guides – designed to brief and highly usable – for the practice/clinical, policy, and consumer audiences. Such corollary documents may be made available for this report (these products are typically ready for posting on the AHRQ website 3-6 months after the full report is published on-line).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	General	This report provides a masterful summary of published articles that describe rigorous studies of interventions for maltreated children. A large body of literature has been identified and the fraction of that literature meeting standards of rigor has been carefully reviewed, and presented without redundancy. That this is a fraction of published literature makes the contribution clinically meaningful on the one hand, but on the other hand, suggests that work with maltreated children may be occurring in situations that make rigorous research difficult or impossible, so that some clinical realities may not be captured by the studies that met criteria for review. A last careful reading might be needed to identify typos that are not mis-spellings (e.g. "sever' instead of "severe" - Executive Summary4 line 46)	As noted above, we have added language acknowledging that our exclusion criteria may have resulted in the exclusion of trials that might bolster evidence for included interventions or support inclusion of other interventions with at least low strength of evidence. We also emphasize in the report our intent (1) to reduce the noise of clinical heterogeneity that currently undermines the extant evidence base, (2) to maintain the rigorous approach for study inclusion that has been employed across AHRQ CERs, and (3) to avert yet more heterogeneity due to inconsistent, vague, or absent definitions of samples of children defined as 'at risk' or an admixture of 'at risk' and maltreated.
Peer Reviewer #4	Introduction	The Background summarizes current demographics and social costs of maltreatment, winnowing unwieldy statistics with reference to solid citations. Key questions and the analytic framework are well set forth. Fig A (Analytic Framework) is very useful.	No action needed
Peer Reviewer #4	Methods	Inclusion and exclusion criteria are well-justified. Search strategies clearly stated and logical. Definitions and criteria for the outcome measures are appropriate, although terms such as "secure attachment" are not diagnoses. Regarding statistical methods, the Method consisted of data synthesis and stratification by evidence grading, rather than statistical analysis, which is appropriate. Table A presents a detailed list of inclusion and exclusion criteria.	Secure attachment is an outcome included under healthy caregiver-child relationship domain. None of the healthy caregiver child relationship outcomes are diagnostic in nature.
Peer Reviewer #4	Results	the detail presented in the results section is appropriate, and masterfully condensed. Characteristics of the studies are described sufficiently to enable the reader to make meaningful comparisons. Table B clearly shows how selected studies were analyzed. The narrative conveys the key messages clearly and succinctly. Figures, tables, and appendices comprehensive and descriptive. Table C is useful, to anchor the reviewed interventions with respect to the populations for which they were applied. The investigators have include all studies that fit their criteria.	No action needed

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Discussion	The summary of Key Findings and Strength of Evidence maps the terrain covered by the studies reviewed, and by contrast, the areas (e.g. type of maltreatment, amount of intervention) and methodologies (e.g. head-to-head comparison of interventions) that lacked quality publications. Limitations were well- stated. Particularly important is the trade-off of study rigor versus applicability in a somewhat chaotic and under-funded field of work. Research and Substantive Gaps summarizes limitation of the evidence-based studies described but may understate the difficulties of the social context for achieving full fidelity, adequate sample sizes, control groups etc, such as the need to partner with state or local agencies that have variably low staffing and resources or standards for good quality improvement methodology, or even data sharing. Future research section points the way to good research but may not make that research possible given the constraints mentioned above.	As noted above, we strongly acknowledge the challenges in conducting high quality RCTs with this particularly vulnerable population. This issue is raised in the Research Gaps section, 1st paragraph; Future Research Needs section, 1st paragraph; and Conclusions section, 1st paragraph in both the Executive Summary and Main Report. Also, as noted previously, the report now explicitly emphasizes the need for high quality, large-scale, multisite trials that can assess treatment moderators. This recommendation is made in the Implications for Policy and Implications for Research sections of the Discussion in the Executive Summary and the Report. This point is raised again in the Conclusions section (Executive Summary and main Report) and the Abstract.
Peer Reviewer #4	General	The report is well-structured and well-organized. Main points are clearly presented. The extent to which conclusions can be used to inform practice and policy decisions will depend on political will and funding.	No action needed
Peer Reviewer #5	General	The report identifies interventions that may be useful to clinicians and shows that the evidence supporting these interventions is quite limited but sufficient to identify promising interventions for both mental health and safety/placement outcomes.	No action needed
Peer Reviewer #5	Introduction	Thorough and clear. Some description of the core elements of each class of interventions would be better placed here than in the Results.	Detail about the core elements of interventions has been excised from the Results. The core elements of the interventions (e.g., whether parent, child, and/or parent-child focused; core components; format and intervention strategies) are succinctly presented in Table 1-3 in the Introduction.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	Methods	Well justified i/e and outcome criteria, search strategies, and reporting of statistical findings.	No action needed
Peer Reviewer #5	Results	The studies are very clearly and thoroughly described. There is a great deal of repetition when studies are referred to multiple times for the different key questions; this is difficult to follow until the summary of key question findings, which are excellent except that I did not see a summary of all key findings for each intervention in one place. For example, one more summary table might be helpful: a matrix that shows interventions in rows and their status on each key finding in columns, perhaps organized so that readers can quickly identify interventions that addressed several key questions. Ultimately many readers will want to know which interventions addressed several key questions successfully, as well as which interventions have the best evidence for each of the key outcomes.	We considered different formats for the report but decided to maintain the structure in the interest of consistency across EPC reports. We did, however, revise the Results to streamline and reduce repetition. A summary of key findings for all interventions for child well-being and child welfare outcomes is provided in Table B (see Results section of the Executive Summary and Key Findings section of the Discussion in the main report). The findings for the other KQs did not lend themselves to integration in a single table.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	Discussion	Discussion/ Conclusion: Clinical implications are a detailed but largely repetitious summary of the results, with some research caveats that belong in the research implications section. Three key things are missing: 1. a synthesis for the clinician concerning what reasonable choices they have for addressing each of the key outcomes, rather than a listing of which interventions showed evidence of effectiveness; 2. relatedly, there is no distinction between evidence of efficacy and effectiveness--this is alluded to in earlier sections describing some interventions, but ultimately it is not clear whether the strongest findings reflect effects obtained under controlled conditions with major exclusions to rule out complex cases (efficacy) or in intervention implementations that are faithful to the population and demands/constraints of real-world practice; 3. policy implications (of which there could be many) are entirely lacking, except to encourage programs to collect evaluation data.	The reviewer raises important issues. The Clinical and Policy Implications sections in the Executive Summary and the Discussion in the main report have been considerably revised, addressing concerns 1 and 3. Regarding the issue of efficacy vs. effectiveness studies: The vast majority of included studies were efficacy trials, and the Results chapter now clearly identifies which trials effectiveness trials (see first paragraph for each intervention). We also address the issue of efficacy and effectiveness in the Research Gaps and Future Research Needs sections of the Executive Summary Discussion and Discussion in the main Report. We differentiated efficacy from effectiveness trials based on guidance provided in Gartlehner, 2006.
Peer Reviewer #5	Clarify and usability	The executive summary and concluding sections are very usable. The results are difficult to follow because they mix descriptions of interventions, the studies testing them and their methods and results, and seem more like an appendix than a summary of the major results.	We have streamlined the results section, particularly by reducing repetition of information provided in Tables 1-3 in the Introduction (i.e., description of the interventions).
Peer Reviewer #6	General	The report is clinically meaningful as the outcomes measures are transparent to problems that contribute to significant emotional, cognitive, and social disturbances in children and are predictors of functional impairment and adverse outcomes in later years. The key questions are appropriate and explicitly stated. The target population and audience are explicitly defined.	No action needed.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Introduction	The introduction provides a clear and focused summary of the purpose and methodology of the project. It identifies where specific information can be found as well as the location and content of all of the tables in the report. The document is almost 400 pages and thus could be off-putting to clinicians and others in the target audience who are looking for very specific information. The clarity of the executive summary and the organization of the contents really eliminated this potential problem.	No action needed.
Peer Reviewer #6	Methods	The inclusion and exclusion criteria are justifiable. The search strategies are explicitly stated and logical. Definitions and diagnostic criteria for the outcome measures are appropriate as are the statistical measures.	No action needed.
Peer Reviewer #6	Results	There is substantial detail but it is appropriate given the various audiences the report has in mind. The characteristics of the studies are clearly described particularly in association with the figures, tables and appendices. I do not see any studies that have been overlooked or conversely any studies included that should not have been. The key messages are explicit and applicable.	No action needed.
Peer Reviewer #6	Discussion	Implications of major findings are clearly stated. Limitations are adequately described. The authors did not omit any important literature. The future research section is clear and easily translated into new research.	No action needed.
Peer Reviewer #6	General	Report is extremely well-structured. A model of organization. Should be used as a prototype for other similarr reports. The conclusions can be used to inform policy and practice decision.	No action needed.
Peer Reviewer #7	General	Yes, the report was quite comprehensive in its approach and provides a meaningful contribution to the field. The target population is clear. The key questions were appropriate and also clearly stated.	No action needed.
Peer Reviewer #7	Introduction	The introduction section provides detailed background information on key definitions, incidence and prevalence, etiology, disease burden, interventions, and the study methodology. On page 3, lines 55-56, the reference for the foster care statistics should be updated (see: http://www.acf.hhs.gov/programs/cb/stats_research/afcars/tar/report18.htm). There are also several studies that may be useful to reference (Longitudinal Studies on Child Abuse and Neglect) in this background etiology section (see: http://www.iprc.unc.edu/longscan/)	We made minor revisions to the Incidence and Prevalence section of the Introduction (main Report) focusing on data from the most recent Child Maltreatment Report (2011). References to several relevant papers reporting findings from the LONGSCAN study are now included (Background, 1st para, Executive Summary and Introduction; Background, 1st para, and Etiology, main Report, 4th para).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Methods	This section is very clear and detailed. However, I just want to add that despite the desire to make the clear delineation between children who have been maltreated and children at-risk of maltreatment, there are actually a few studies now that demonstrate that these populations have very similar risk profiles (i.e., there is very little difference between the in-risk and at-risk populations). I know this may not be something to consider after the review has been completed-- but this may be something you might want to reference in the conclusion section. Here's one reference: http://www.mendeley.com/research/defining-maltreatment-according-to-substantiation-distinction-without-a-difference/	We agree with the reviewer that delineating between children who have been maltreated and those at risk of maltreatment is not a clean divide and may be considered a rarified approach by some. We have revised the report to explicitly call attention to these populations sharing similar risk/clinical profiles, citing the suggested references. We would argue, however, that intervention work with children known to have been maltreated and with parents involved with CPS presents markedly different therapeutic and operational challenges compared with preventive intervention directed at children/parents/families at risk. These points are made in Scope and Key Questions, 2nd and 3rd paragraph, Executive Summary. In the main Report, this issue is addressed in both the Population section of the Methods chapter (2nd paragraph) and the Key Findings and Strength of Evidence section of the Discussion in the main Report (2nd paragraph). Please also see response to Reviewer #1 (General Comments).
Peer Reviewer #7	Results	These sections are very detailed-- and in many cases, a little overly detailed and it was hard to remember the primary questions within each of the sections. I found the executive summary and the conclusion section much easier to follow because there was so much detail and repetition in the results sections. I understand the reasoning because of the questions, but the findings were similar across several of the questions so that's where it felt very repetitive.	No action needed.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Discussion	The discussion section was nicely written. The findings and implications were clearly stated. On page 140 (lines 47-50) tries to make a distinction that the at-risk parents and in-risk parents have different clinical needs and I would be more cautious in that statement --because as I stated in my comment on the methods section, there are a few studies that indicate that the two populations are actually more similar than different. This definitely presents a challenge to the systematic review and the need to draw clear boundaries --- so I would rephrase this to acknowledge that there are some similarities.	See earlier response to comment on the Methods section.
Peer Reviewer #7	General	I think the results section could be streamlined a little bit more -- I found it hard to stay focused on the overarching research question and the findings for each program. I am not sure that I have a better suggestion, except to say that some of the tables could be more to an appendix section so that the key findings were much clearer. The executive summary and conclusions section did a nice job with the summary. The recommendations for future research areas were helpful. The conclusions can be used to help inform policy and practice.	We agree that the Results chapter is long (see also response to Reviewer #5 above). We have revised it considerably to streamline as much as possible and reduce repetition.
Peer Reviewer #8	General	Overall, I thought the study was well-constructed and sound methodologically. However, its rigorous methodology may have also resulted in limitations that impact its clinical relevance and usefulness in the field. My overall concern is that the methodology used to rule out certain studies may have severely limited the range of interventions considered. While there may have been few studies conducted that examine the target population of maltreated children, the collective knowledge gained from reviewing other studies that included at-risk children, mixed populations or systems level approaches may have yielded more usable information. Given my knowledge and experience of the range of interventions used to treat children exposed to violence and trauma, there certainly are numerous effective and evidence-based interventions that due to the authors' methodology were not considered for this study. Oftentimes, by definition, children who receive trauma-focused services have had a history of maltreatment (excluding children exposed to community violence or other forms of trauma). Many of the existing studies that have been conducted and reviewed by NREPP could have relevance to understanding the potential applicability to the foster and child welfare population. Although these studies may not have relevance for KQ2, they arguably would be useful in helping to answer the other key questions. Oftentimes, we do not consider the applicability of other research or put children into "silos" based upon their contact with various systems. It has been my experience that children served by behavioral health, child welfare, juvenile justice, education and other major systems often share characteristics (not the least of which is a history of maltreatment) and further move back and forth between these systems. By ruling out evidence drawn from other populations, you may be creating an artificial dichotomy between groups that in practice does not really exist.	As noted above in response to similar comments from other reviewers, we fully recognize the limitations of this CER associated with the stringent exclusion decisions. We agree with the reviewer that our exclusion decisions may be considered a rarified approach by some and have added language to the report clearly acknowledging that our exclusion decisions may have resulted in the exclusion of trials that might bolster evidence for included interventions or support inclusion of other interventions with at least low strength of evidence. We emphasize that our intent was three-fold: 1) to reduce the noise of clinical heterogeneity that currently undermines the extant evidence base, 2) to maintain the rigorous approach for study inclusion that has been employed

Commentator & Affiliation	Section	Comment	Response
		I also question if there could have been other analytic approaches considered. Many of the a priori hypotheses had to be eliminated because there were insufficient samples to meet the pre-identified conditions. Given the status of the field and the limited amount of available research, I wonder if there could have been an initial scan of articles that yielded ad hoc categories for analysis. In other words, rather than trying to fit the range of literature reviewed into rigid hypotheses that ultimately were not applicable or sufficiently inclusive of the actual data, it might have been more useful to use the literature review as a means for generating categories for analysis that might have yielded more useful and conclusive results. I also wonder if the rigor applied to the types of studies and sources of data could also have resulted in useful information not being considered for this study. Other approaches could have included doing a scan of current EBPs being used by child welfare agencies across the country and then further examining if they routinely outcome data which may or may not be published in scholarly journals, but may appear in grey literature. Further, the results of this analysis could be further contextualized by lessons learned from the National Child Traumatic Stress Network, that includes a child welfare subgroup. By contextualizing these results within the framework of the actual practices that are ongoing, the reader may have had a better (and more accurate) sense of the current state of the field. By reading these results alone, one may conclude that there are very few evidence-based programs (and those that exist have low evidence) that are available and are currently being implemented to serve maltreated children. And unfortunately, I do not think that is an accurate depiction of the field. Finally, my only other methodological concern I'd like to mention is that there were only two reviewers—(the report indicated one junior and one senior reviewer) and that a third was used for additional consultation or tie-breaking. I'm concerned about the implications that may be drawn from this study that are based only on two reviewers. I also wonder if these reviewers are conversant with current practices in the trauma field. This seems to be a limitation. Depending on how this study is being planned to be used, I think it is very important to ensure that the audience gets a full sense of the current state of the field and available resources. That by no means negates some of the conclusions about the need for more research or limitations of available treatments, but it might cast the results in a slightly different light.	across AHRQ CERs, and 3) to avert yet more heterogeneity due to inconsistent, vague, or absent definitions of samples of children defined as 'at risk' or an admixture of 'at risk' and maltreated. In particular, we have revised the report to explicitly call attention to the fact that at risk and maltreated populations share similar risk/clinical profiles. However, we also note that intervention work with children known to have been maltreated and directed towards parents with known involvement with CPS presents markedly different therapeutic and operational challenges compared with preventive intervention directed at children/parents/families at risk. These points are made in Scope and Key Questions, 2nd and 3rd paragraph, Executive Summary and in Key Findings and Strength of Evidence, Overview, Discussion, 2nd paragraph. We fully recognize that the scoping of the review and rigorous approach has resulted in an uncomfortable dissonance with the breadth of practices widely in use in the field – many of which are not represented in this review (as the reviewer points out). This dissonance reflects the reality that there is not, in fact, a common definition of 'evidence-based' intervention or systematic application of degree of evidence driving the practice arena.

Commentator & Affiliation	Section	Comment	Response
			Numerous interventions that are widely used are described loosely as 'evidence supported' or 'promising.' As such, a key intent of the review is to stimulate the field to step back and engage in critical self-reflection about the state of intervention for this vulnerable population and the urgent need for adequate funds to be made available that support stronger research. For the practice community, we have added language to the Executive Summary tempering recommendations, emphasizing that our findings regarding low strength of evidence is indicative of the serious challenges and early stage of research in the field. Regarding alternate analytical approaches: we scrupulously followed the process for AHRQ reviews which requires defining the PICOTS and Key Questions in advance (e.g., identifying intervention characteristics apriori). We fully agree that the alternate approaches suggested would yield highly salient information for the field. The EPC team comprised several PhD-level (senior) investigators with relevant expertise, including two experienced clinicians specializing in child trauma treatment – including the director of a Category III NCTSN-funded Community Treatment and Services Center. The review

Commentator & Affiliation	Section	Comment	Response
			process, including consensus discussions, were distributed across the numerous senior staff on the team. We will make this clearer in the Methods section in the Executive Summary and main Report.
Peer Reviewer #8	Introduction	I thought the introduction did a good job explaining the rationale, structure and methodology of the study. It was clearly written. However, I would be interested in knowing more about the "why" the study was conducted and what the hopes were for "how" the results might inform the field. Given my comments above, I think it is important to not jump to potentially erroneous conclusions or use the results to inform policy based upon the rigorous methodology that excluded many potential sources of data.	We added language up front in the Executive Summary and main Report explaining the goals and intended impact of this review: to provide stakeholders with a synthesis of the best evidence in the field of child maltreatment and to identify critical areas to address in future intervention research. In the Introduction to the main Report (Background, 2nd paragraph), we also explicitly acknowledge that the rigorous approach specified for the CER limited the range of interventions that could be included. See also the section entitled Limitations of the Comparative Effectiveness Review (in both the Executive Summary and Discussion).
Peer Reviewer #8	Methods	As previously stated, although methodologically sound from a scientific perspective, the inclusion and exclusion criteria may have significantly limited the scope of potential evidence based practices that have ample evidence and could be used effectively as interventions with maltreated children. 26 studies (many focusing on the same interventions) is a rather limited sample size to extrapolate and major conclusions about the field—especially treatments that may benefit traumatized children.	See above responses. There is not a consistent definition of EBT in the field. It seems that the reviewer may implicitly follow a particular perspective but "ample evidence" is an inferential statement that does not accord with nor is consistent across leading researchers, registries, or practice review catalogues.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Results	Although the arguments presented for the conclusions were clear, there were some examples of conclusions that I did not entirely agree with, or felt that the rigidity of the study failed to capture useful data. An example is on page Executive Summary-16 where TF-CBT and psychodynamic treatment are compared, and although TF-CBT yields better outcomes the authors note that TF-CBT has a specific trauma focus while psychodynamic approaches do not, thus confounding the results. This conclusion makes no sense to me. Yes, TF-CBT is effective because it is a trauma specific intervention. However psychodynamic treatment, by nature, both naturally includes attending to the trauma (by focusing on unresolved conflicts or interruptions in the developmental process) and also avoids directly addressing the trauma because of the nature of the theoretical orientation (thus you would never a find a condition where you could actually compare trauma focused CBT and trauma-focused psychodynamic treatment). This is an example (and there are others like this) where the seemingly lack of clinical understanding of the authors results in a conclusion that did not make sense to me. In this instance, the comparison between TF-CBT and psychodynamic treatment is appropriate and the conclusion that TF-CBT is more effective is sound.	The reviewers' point is well-taken. The CER process explicitly focused on the descriptions provided by study authors however comprehensive or limited. Yet in many instances, the descriptions of intervention and comparator approaches are not detailed nor do they necessarily give the reader a clear sense of what actually occurs through an intervention. This is true of different interventions, different publications from the same trial, active comparators, and services- as-usual comparators. Indeed, it is difficult to "truly" know the intervention from a publication. This is especially true when studies indicated they were using broad perspectives rather than specific treatments. For example, a psychodynamic approach can vary considerably based on population, practitioner, and sub-theory (classical psychoanalytic, ego psychology, object relations, short term psychodynamic, interpersonal psychotherapy, Hans Strop). The reviewer's definition of how psychodynamic treatment addressed and falls short of addressing trauma is reasoned but is by no means a standard of care or universally accepted formulation.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Discussion/conclusion	I do think the authors did a good job describing their findings and how they reached their conclusions. As previously stated, I am concerned that due to the rigor of this study and the strict exclusionary criteria used, couple with an apparent lack of contextualizing the findings in current practices in the field, that the conclusions reached may be over-reaching and have the risk of impacting policies or support for evidence-based models that may or may not have been included in the final analysis. I would like to see the limitations section address some of my previously identified concerns. Clearly, there is a need for future research, but I think that the call may be for funding to support systematic reviews of the use of EBPs for the child welfare population, along with some guidelines for the inclusion criteria so that the data can be analyzed and used to inform policy and practice. Further, federal demonstrations such as the NCTSN. ACF and OJJDP all have great potential for data collection and analysis of the effectiveness of evidence-based practices and programs. Coordinated research efforts across these initiatives where a wide range of evidence-based models have been implemented, would field highly relevant and useful analyses that could greatly inform the field. In short, I would hope that this study alone is not used to draw conclusions about the utility of established evidence-based models for the child welfare population or to drive policy decisions. Although much useful data was collected and analyzed, and the results clearly indicate that we have much work to do in the field, the potential conclusions that can be drawn by readers could be potentially misleading.	See previous responses.
Peer Reviewer #8	General	I do think the report was clearly written. Please see my answer above pertaining to my concerns about how the conclusions might be used to drive policy and practice. I also think that when thinking about child welfare, it is almost impossible to separate out individual practice approaches from the context of the child welfare and foster care system as a whole. Systemic changes and trauma-informed approaches must not only be reflected at the intervention level, but also should be integrated into the child welfare decision making models, reporting systems, screening, assessment and workforce development and training of child welfare staff and foster parents. When thinking about policy implications, it is very important not to consider these evidence-based models in isolation, but rather to consider how systemic level changes can potentially have a significant impact on individual outcomes for maltreated children and their families.	See previous responses.