PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Parent Skills Training for Parents of Children or Adults with Neurodevelopmental Disorders: Systematic Review and Meta-Analysis Protocol
AUTHORS	Reichow, Brian; Kogan, Cary; Barbui, Corrado; Smith, Isaac; Yasamy, M. Taghi; Servili, Chiara

VERSION 1 - REVIEW

REVIEWER	Ole Jakob Storebø Psychiatric Research Unit
REVIEW RETURNED	16-Jul-2014

GENERAL COMMENTS	This is a systematic review on a very interesting and relevant topic. In the objective only standard care is mentioned, but later on it is described that the control group could consist of also no intervention and waitlist (and TAU). Please clarify this. Regarding the participants: all ages is included and also all subtypes of autism spectrum disorders. This gives a very high clinical heterogeneity and is it likely that parents with adult children will participating in parent skills training? The reason for publishing protocols for systematic reviews is because this reduces the impact of authors' biases, promotes transparency of methods and processes, and reduces the potential for duplication (Higgins 2011). Furthermore to avoid post hoc decisions made when the impact on the results of the research is highly susceptible to risk of bias. That is why it is important to make beforehand decision on primary outcomes. It important also to avoid too many outcomes because of the risk for multiciplicity. In this protocol for a systematic reviews there are far too many primary outcomes. The probability of finding false positives increases considerably. By reducing the amount of outcomes the risk for multiple significance testing and thereby finding false positives will decrease considerably. A risk of bias assessment is planned, but there is no description of how this assessment is done. When is an included study considered high risk of bias? Is this done in the same way as in Cochrane Systematic Reviews?

	The use of SMD is planned, but no mention of what is considered as a clinical relevant high effect? There is no description of how to deal with missing data! Regarding the Hetereogenity**;** with few studies a visual inspection of the funnel plot may be misleading. The chi-square test is conservative and even a p-value of 0.1 may not work. The I^2 index has also large variability (possibly consider values of I^2 >50%). It is possible that the chi-square and the index may give contradictory results.

REVIEWER	Ryan Williams American Institutes for Research, USA
REVIEW RETURNED	06-Aug-2014

GENERAL COMMENTS	This is a proposed meta-analysis of parent skills training programs for children and adults with neurodevelopmental disorders. My comments will focus on addressing some methodological concerns: 1. The authors' selection criteria limits their search to randomized experiments but they do not justify why quasi-experiments would be excluded. I strongly recommend that the authors include non-randomized experiments to the extent to which the meet the other criteria and evaluate them separately or use design as a moderator in a random effects meta-regression. 2. The authors did not discuss how they will handle effect size dependencies, something that will almost certainly come up. I suggest the authors use robust variance estimation (e.g. Hedges, Tipton, and Johnson, 2010; Tipton, 2014) for their standardized mean differences. 3. The authors discuss subgroup analyses as a series of disaggregated analyses. To the extent that the data permit it, I recommend random effects meta-regression using one or more potential study moderators in each analysis. This approach may be more efficient than analyzing different groups of studies separately (i.e. you may be able to use all effects in estimating the moderator effects). I don't know how important it is at this stage but a codebook would also be useful to review at this point.

VERSION 1 – AUTHOR RESPONSE

Reviewer Name Ole Jakob Storebø
Institution and Country Psychiatric Research Unit
Please state any competing interests or state ‘None declared’: None declared

This is a systematic review on a very interesting and relevant topic. In the objective only standard care is mentioned, but later on it is described that the control group could consist of also no intervention and waitlist (and TAU). Please clarify this. Regarding the participants: all ages is included and also all subtypes of autism spectrum disorders. This gives a very high clinical

heterogeneity and is it likely that parents with adult children will participating in parent skills training?

THANK YOU FOR CATCHING THIS INCONSISTENCY. THE ABSTRACT AND METHOD WERE CORRECT, AND THE OBJECTIVE SECTION NOW INDICATES WE WILL EXAMINE NO TREATMENT OR STANDARD CARE. THE PRIMARY OBJECTIVE NOW READS:

The primary objective of our review is to systematically review and meta-analyze evidence to determine if parent skills training programs for parents who have a child with a developmental disorder produce greater benefits than no treatment or standard care on child functioning and on parental or family functioning, as measured across multiple domains, and to use meta-analytic techniques to determine which program components are most reliably associated with the most successful outcomes of parent skills training programs.

WITH RESPECT TO YOUR COMMENT RE PARENTS OF ADULT CHILDREN PARTICIPATING IN SKILL TRAINING GROUPS, WE DO NOT ANTICIPATE FINDING MANY STUDIES, BUT FEEL IT IS IMPORTANT TO NOTE THAT PARENTS WHO CARE FOR CHILDREN WITH DEVELOPMENTAL DISABILITIES FACE NEW CHALLENGES ONCE THEIR CHILD REACHES ADULTHOOD AND THAT SOME OF THESE ISSUES (E.G., EMPLOYMENT, END OF LIFE CARE) COULD BE ADDRESSED USING THIS TYPE OF TREATMENT. HENCE WE HAVE DECIDED TO LEAVE THE INCLUSION CRITERIA OF AGE AS IS.

The reason for publishing protocols for systematic reviews is because this reduces the impact of authors' biases, promotes transparency of methods and processes, and reduces the potential for duplication (Higgins 2011). Furthermore to avoid post hoc decisions made when the impact on the results of the research is highly susceptible to risk of bias. That is why it is important to make beforehand decision on primary outcomes. It important also to avoid too many outcomes because of the risk for multiciplicity. In this protocol for a systematic reviews there are far too many primary outcomes. The probability of finding false positives increases considerably. By reducing the amount of outcomes the risk for multiple significance testing and thereby finding false positives will decrease considerably. A risk of bias assessment is planned, but there is no description of how this assessment is done. When is an included study considered high risk of bias? Is this done in the same way as in Cochrane Systematic Reviews?

WE ORIGINALLY SPECIFIED THAT ADAPTIVE BEHAVIOR WOULD BE PRIMARY OUTCOME FOR CHILDREN AND QUALITY OF LIFE WILL BE PRIMARY OUTCOME FOR PARENTS. WE HAVE NARROWED THE LIST OF SECONDARY OUTCOMES TO TWO CHILD OUTCOMES, THREE PARENTAL OUTCOMES, TREATMENT SATISFACTION, AND ATTRITION. PARAGRAPH NOW READS:

We will examine the effects of parent skills training programs on both child outcomes and parent outcomes. The primary outcome measure for children will be adaptive behavior (e.g., functional skills, daily skills). We will also include two secondary child outcomes: child development and problem behavior. The primary outcome measure for parents will be quality of life. We will also three parental secondary outcomes: psychological health, parent skills, and family quality. Finally, we will examine consumer (parental) satisfaction and attrition.

WITH RESPECT TO THE RISK OF BIAS ASSESSMENT, WE WILL USE METHODS CONSISTENT WITH THOSE DESCRIBED IN COCHRANE HANDBOOK, WHICH WAS REFERENCED IN THE PROTOCOL. SINCE WE WILL NOT BE INCLUDING NON-RANDOMIZED STUDIES, REFERENCE TO CONCERNS FOR NON-RANDOMIZED STUDIES OF REEVES AND COLLEAGUES HAS BEEN REMOVED. SECTION NOW READS:

Two authors will independently assess the study level risk of bias using the methodology and risk of bias tool detailed by the Cochrane Collaboration [18]. We will resolve discrepancies through consensus, and if consensus cannot be reached, a third researcher will make a final judgment on the risk of bias. We will use the tool to assess the following domains: sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other potential sources of bias.

The use of SMD is planned, but no mention of what is considered as a clinical relevant high effect? There is no description of how to deal with missing data! Regarding the Hetereogenity; with few studies a visual inspection of the funnel plot may be misleading. The chi-square test is conservative and even a p-value of 0.1 may not work. The I^2 index has also large variability (possibly consider values of I^2 >50%). It is possible that the chi-square and the index may give contradictory results.

INTERPRETATION OF CLINICALLY RELEVANT EFFECTS VARY ACROSS MEASURES, SO SETTING A SPECIFIC VALUE FOR A HIGH EFFECT IS DIFFICULT. WE PROPOSE TO CONVERT THE SMD EFFECT SIZES WE FIND INTO COMMON LANGUAGE EFFECT SIZES TO CONVEY CLINICAL SIGNIFICANCE AND UTILITY (RELEVANCE). EXAMPLES IN ADDITION TO THE ONE PROVIDED IN THE TEXT, WITH DIRECT RELEVANCE TO DATA WE WILL BE USING INCLUDE EXTRAPOLATING DEVELOPMENTAL OUTCOMES TO SHOW HOW MANY ADDITIONAL SKILLS A CHILD WOULD OBTAIN IN THE TREATMENT GROUP COMPARED TO THE CONTROL GROUP, SHOW THE NUMBER OF FEWER BEHAVIOR PROBLEMS PER WEEK AFTER TREATMENT COMPARED TO CONTROL, OR THE LIKELIHOOD OF A PARENT MOVING FROM A CLINICALLY HIGH LEVEL OF DEPRESSION TO NOT BEING AT A LEVEL OF CLINICAL DEPRESSION.

In our analysis of the standardized mean difference effect sizes, we will consider an effect to be clinically relevant, irrespective of statistical significance, by transforming the standardized mean difference effect sizes into common language effect sizes[24,25]. An example of transforming standardized mean difference effect size into common language effect size is shown in a meta-analysis of social skills group interventions for children with autism spectrum disorders by Reichow et al.,[26] who showed the weighted mean effect size of d=.47 equated to a gain of 24 additional social skills for the treatment group compared to control. Possible transformations that might be possible from data anticipated to be analyzed in this review would also include extrapolation of the number of fewer behavior problems per week for children whose parents receive a parent skills training program or the likelihood of a parent moving below a clinical threshold for depression.

WITH RESPECT TO MISSING DATA, WE HAVE INDICATED THAT WE WILL CONTACT AUTHORS FOR THE DATA (THIS IS INDICATED IN THE THIRD SENTENCE OF THE DATA EXTRACTION SECTION OF METHODS, “IF CRITICAL INFORMATION IS MISSING FROM A REPORT, WE WILL CONTACT THE AUTHORS TO INQUIRE OF THE DATA”. IF THEY ARE UNABLE OR UNWILLING TO PROVIDE MEANS AND STANDARD DEVIATIONS FOR OUTCOMES WE WILL ESTIMATE EFFECT SIZES FROM STATISTICAL INFERENCE TESTS REPORTED IN THE STUDY REPORT. FOR OTHER DATA THAT WE ARE UNABLE TO EXTRACT FROM A STUDY (E.G., MODERATOR VARIABLES SUCH AS AGE, IQ, NUMBER OF SESSIONS PER WEEK) WE WILL NOT IMPUTE DATA, BUT WILL LEAVE THE STUDY OUT OF THE MODERATOR ANALYSIS FOR EACH MISSING VARIABLE ON A CASE-WISE BASIS.

EXTRACTION OF INDIVIDUAL STUDY ESTIMATES NOW INCLUDES FURTHER CLARIFICATION OF MISSING OUTCOME DATA PROCEDURES:

If means and/or standard deviations are not provided in the report, we will contact authors to request the data. If authors do not provide data, we will estimate effect sizes if possible from p-values, t-values results, or F-values if they are provided using the effect size calculator of the Campbell Collaboration, which uses the formulae provided in the text of Lipsey and Wilson.

WITH RESPECT TO HETEROGENEITY, WE WILL STRESS THE I^2 INDEX, AS PER COCHRANE METHODOLOGY AND RECOMMENDATIONS. A REFERENCE TO THIS HAS BEEN ADDED AS FOLLOWS:

Given the limitations of the Q-statistic, we will emphasize the I2 values in our analyses and reporting of results, as suggested by the Cochrane Collaboration (Deeks et al., 2008).

Reviewer Name Ryan Williams
Institution and Country American Institutes for Research, USA
Please state any competing interests or state ‘None declared’: None declared

This is a proposed meta-analysis of parent skills training programs for children and adults with neurodevelopmental disorders. My comments will focus on addressing some methodological concerns:

1. The authors' selection criteria limits their search to randomized experiments but they do not justify why quasi-experiments would be excluded. I strongly recommend that the authors include non-randomized experiments to the extent to which the meet the other criteria and evaluate them separately or use design as a moderator in a random effects meta-regression.

WE APPRECIATE THAT MANY RESEARCH METHODS CAN PROVIDE HELPFUL EVIDENCE AND WE CONSIDERED INCLUDING NON-RANDOMIZED STUDIES, BUT ULTIMATELY DECIDED AGAINST IT IN ORDER TO REDUCE RISKS OF BIAS ASSOCIATED WITH NON-RANDOMIZED STUDIES. THE FOLLOWING SENTENCE HAS BEEN ADDED TO THE TYPES OF STUDIES SECTION OF THE STUDY SELECTION METHODS:

We will limit our inclusion to studies conducted using randomized control trial designs. We have chosen to limit the inclusion to studies using randomized control trial designs to decrease the likelihood of including studies with high risk of bias, which are more likely in studies conducted using quasi-experimental designs.

ADDITIONALLY, ANY REFERENCE TO METHODS FOR DEALING WITH NON-RANDOMIZED STUDIES THAT WE LOCATED IN THE MANUSCRIPT, WHICH WERE LEFTOVER FROM A PREVIOUS DRAFT BY ERROR HAVE BEEN DELETED.

2. The authors did not discuss how they will handle effect size dependencies, something that will almost certainly come up. I suggest the authors use robust variance estimation (e.g. Hedges, Tipton, and Johnson, 2010; Tipton, 2014) for their standardized mean differences.

AS OUTLINED IN THE PROTOCOL, WE ARE ONLY TAKING ONE STUDY ESTIMATE PER OUTCOME, REGARDLESS OF HOW MANY MEASURES A STUDY MIGHT HAVE REPORTED ON FOR THAT OUTCOME, AS RECOMMENDED BY COCHRANE METHODOLOGY (E.G., HIGGINS ET AL. 2008). TO DETERMINE WHICH ESTIMATE TO USE FROM A STUDY, WE CREATED A PREDEFINED HIERARCHY TO REDUCE INTRODUCTION OF BIAS. THUS, WE DO NOT FEEL THAT EFFECT SIZE DEPENDENCE SHOULD BE AN ISSUE.

3. The authors discuss subgroup analyses as a series of disaggregated analyses. To the extent that the data permit it, I recommend random effects meta-regression using one or more potential study moderators in each analysis. This approach may be more efficient than analyzing different groups of studies separately (i.e. you may be able to use all effects in estimating the moderator effects).

IT WOULD BE DESIRABLE TO HAVE THE POWER TO CONDUCT MODERATOR ANALYSES WITH MULTIPLE VARIABLES, HOWEVER, WE DO NOT ANTICIPATE HAVING THE POWER TO DO SO, THUS WE HAVE CHOSEN TO EXAMINE MODERATORS INDIVIDUALLY.

I don't know how important it is at this stage but a codebook would also be useful to review at this point.