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.

ARTICLE DETAILS

TITLE (PROVISIONAL)	What works for whom and why? Treatment effects and their moderators among forcibly displaced people receiving psychological and psychosocial interventions: study protocol for an individual patient data meta-analysis
AUTHORS	Kurath, Jennifer; Akhtar, Aemal; Karyotaki, Eirini; Sijbrandij, Marit; Cuijpers, Pim; Bryant, Richard; Morina, Naser

VERSION 1 – REVIEW

REVIEWER	Emberti Gialloreti, Leonardo University of Rome Tor Vergata, Department of Biomedicine and Prevention
REVIEW RETURNED	24-Sep-2023

GENERAL COMMENTS	The study protocol is well described. The rationale and research aims are clear. Correctly, this protocol paper does not present results and/or conclusions. Just a few clarifications: Line 91: “We will conduct a systematic literature search…”. This search has been already partially conducted (Line 199). It just needs to be updated with publications published after January 12, 2022. Regarding literature search, did the authors chose a time-range? Looking at the search string files (“OSF_search_strings.pdf” and “search_syntax_.pdf”) it seems that no time-range has been included. I suggest specifying in the manuscript that no time-range has been considered (or, if one was considered, to specify which was the range). It should be taken into account that the older the study the higher the probability of not having access to the requested full individual dataset. It is likely that analysis of IPD will include mainly newer studies, analysis of aggregate data mainly older ones. Line 128: “…multiple imputation will be used…”. Do the authors foresee to use multiple imputation for missing outcome data on the global database (constructed by pooling all the received databases) or on each single database? Line 128: “selection of the reported result is not applicable…”. This is of course right. However, as it is likely that in several cases no full dataset will be available, thus the authors will have to perform analyses of aggregate data, this possible bias should be considered as well. Line 133: “…PTS symptoms assessed at post-intervention (PT) and follow-up (FU)”. How do the authors define PT and FU? PT, when symptoms are assessed immediately after treatment and FU at a later time? Or, it depends upon the definition of each single

	study? In that case, how will the authors combine outcomes assessed at different timepoints? Line 138: "…resettlement as rationale for exclusion". Why excluding these individuals? Because the outcome was not assessed? In that case the exclusion criterium should be no outcome. Maybe there are other reasons for excluding those who have been resettled? Line 146: "…when a large proportion of authors…". Maybe the authors could define what they mean with "large proportion". Page 9: The authors foresee to run many multivariate regression models for primary and secondary outcome variables. Do the authors plan to correct for multiple comparisons? Lastly, I commend the authors for undertaking such a complex endeavor, aiming to better understand how to treat these vulnerable populations. I look forward to the results, which can help us to better respond to one of the main challenges of our times.

REVIEWER	Schäfer, Sarah K Technische Universität Braunschweig
REVIEW RETURNED	01-Oct-2023

GENERAL COMMENTS	The study protocol entitled "What works for whom and why? Treatment effects and their moderators among forcibly displaced people receiving psychological and psychosocial interventions: study protocol of a series of individual patient data meta-analyses" reports on a planned individual participant data analyses for psychological and psychosocial interventions aiming to reduce symptoms of posttraumatic stress disorder (PTSD) and mental distress in forcibly displaced persons. The authors report on their motives that inspired their project, their systematic literature search, and data-analyses. Overall, such projects are highly needed to shed light on mostly participant-level modulators of intervention effects, which may in future allow for better tailoring of interventions. The project is well-planned and will add valuable knowledge to the literature. However, I have some comments that should be addressed in a revised version of the study protocol. Mainly, I think the current state of research needs to be summarized more adequately, review aims should be elaborated more clearly, and analyses require further specification. This would further improve the value of the study protocol and may also help to further improve an overall important project. 1) The following statement is incorrect (p. 4): "While several meta-analyses have shown different psychological interventions to effectively reduce PTS, there is a considerable heterogeneity among studies (e.g., Kip et al., 2020) which may be attributed to differences in characteristics related to beneficiaries, providers, intervention, and study design. However, factors contributing to this heterogeneity have not yet been explored." Many systematic reviews have explored moderators of treatment effects. There is also an IPD analyses on PM+ (Akhtar et al., 2021) and another ongoing much needed IPD analysis on PM+ and SbS (de Graaff et al., 2022, with preliminary results also being available: http://strengths-project.eu/wp-content/uploads/2023/03/STRENGTHS-D7.4-Modelling-moderators-for-health-outcomes-of-PM.pdf). The state of evidence should be reported in greater detail by clearly stating what the current systematic review can (and cannot) add beyond previous

	reviews in the field (I think there is still a lot that can be added – also beyond the WHO interventions). 2) The authors also state on p. 5 that interventions are overall effective. I agree that there are many effective interventions, however, there is substantial evidence that some psychosocial interventions are less or even not effective in forcibly displaced persons (e.g., Turrini et al., 2021). I think this should be reported more nuanced as it even makes a greater point for the current project. 3) It is not entirely clear what are the primary and/or secondary outcomes of this review and how were they chosen by the review team. Why is there a focus on PTSD and why is the availability of data on PTSD an inclusion criterion when interventions do not need to focus on PTSD? Why are other symptoms analyzed as comorbid symptoms? In general, the protocol reads like this review focuses rather on any intervention aiming to reduce mental distress in forcibly displaced persons and the focus on PTSD derives from the need of defining a primary outcome. I think this might be critical as some of the included interventions might also be less effective for PTSD symptoms (e.g., Schäfer et al., 2023, for PM+). What is the role of other outcomes like attrition, adverse events, treatment non-responses? Given the great potential of IPD meta-analyses for level-I moderators, I believe that the latter outcomes are of major interest. For example, adverse events are quite seldom in most of the trails, and it might provide very important knowledge on their predictors to have a detailed look at adverse events in an IPD meta-analysis. The same applies to attrition – that is a major issue, and the current project has the potential to shed light on this problem. It might also be good to include a table defining primary and secondary outcomes along with (preliminary) moderators of interest. 4) This goes hand in hand with another rather general remark: It should be clearer from the protocol what are the advantages of IPD meta-analyses above traditional meta-analyses. The great potential of IPD is to shed light on level-I moderators with greater statistical power and based on a larger number of events (e.g., in case of adverse effects). This should be stated more clearly in the Introduction section and should also guide the choice of moderators as results will not be substantially different from standard reviews for study-level moderators (e.g., interventions type, type of intervention providers), but will have the potential to substantially enlarge knowledge on participant-level moderators. 5) Inclusion criteria for the IPD meta-analysis are very broad ranging from psychosocial intervention to standard psychotherapy. From the protocol, it is not entirely clear how these heterogeneous interventions will be handled in the analyses. This should be addressed more explicitly – is it reasonable to assume that moderators are equally important for different types of interventions or will meta-analyses on different intervention types be performed separately? Can one except comparable effects from rather preventive psychosocial interventions and standard (and probably trauma-focused) psychotherapy? Might these broad criteria introduce unwanted heterogeneity in the analyses? I would be very interested to hear the authors' ideas on this issue. 6) Overall, the two-stage approach is sufficiently described, while the multi-level model for the one-stage approach is not described in sufficient detail. What kind of random and fixed effects will be tested? What kind of model diagnostics will be applied? Will the authors check the (potentially non-normal) distribution of intervention outcomes? I would recommend to perform model

	diagnostics using DHARMa (https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html). Minor remarks: 1) The title 'study protocol of a series of individual patient data meta-analyses' might be irritating for readers as it may suggest that there is a series of (independent) IPD analyses. However, the protocol describes the analyses of one large dataset rather than a series of analyses/projects. 2) Information on the combination of a one- and two-stage approach and secondary outcomes is missing in the Abstract. The Abstract in general should draw a more complete picture of what is planned in the review. 3) There is a one- and a two-stage approach IPD, with the two-stage approach using aggregated data (but calculated from IPD). In the manuscript, this is sometimes mixed, with referring to meta-analyses on aggregated data (derived from IPD) as traditional meta-analysis. However, the latter would rather be an analysis based on data reported in a paper. This should be differentiated throughout the manuscript. For example (p. 3): "IPD, compared to aggregate data meta-analysis, will allow more complex analyses to identify moderators of treatment effects while standardisation of variables is facilitated, and missing values can be accounted for." 4) On p. 4: Step-by-step is guided by an eHelper (Carswell et al., 2018), I would not refer to the program as unguided. 5) The authors may also think of automated quality checks for IPD – this could be helpful when obtaining data from a large number of studies. It might also be good to add an estimate how many primary studies will be expected (based on knowledge from pervious systematic reviews in the field). 6) What kind of statistical program will be used for the one-stage and the two-stage meta-analysis? Multiple imputations will be done in Stata, will other analyses be performed in R? 7) It might be useful to describe the multivariate outcome approach for the two-stage approach in greater detail. What kind of multilevel model will be used? How will they examine effects on single outcomes? Will they model primary and secondary outcomes in one model? Will they obtain information on between-outcome correlations? Moreover, it is not entirely clear how standardized mean differences (SMDs) will be calculated? Do the authors compare both groups at post-intervention and follow-up assessment? This should be stated more explicitly. 8) It might be better to make us of contour-enhanced funnel plots instead of simply using funnel plots. Another minor remark: How can Egger's regression test be applied when a multivariate meta-analysis is performed (normally making use of a multilevel approach)? In those cases, one may rather approximate regression tests by including sampling error to meta-regression models. 9) I am also not sure with the following statement (p. 9): "In order to identify moderators of treatment effects, we will run several multiple linear regression models, including intervention condition and all potential moderators as independent variables and PTS symptom scores at PT and FU assessments as dependent variables for each trial separately." Does this refer to meta-regression analyses? In case the authors refer to standard regression models, I would recommend using meta-regressions instead. 10) In same places (e.g., p. 3) the authors use the term 'treatment trajectories'. What is meant by that? It would be good to define this more explicitly. Will they also use other models than linear mixed

	models (e.g., growth mixture modeling) to examine distinct trajectories? 11) Will moderator analyses only be performed for posttraumatic stress symptoms (p. 5)? I think it might be beneficial to have a broader focus. Also as being diagnosed with PTSD is not an inclusion criterion. 12) The definition used for “non-active control” is very broad. Why is psychoeducation a non-active control? It might be better to give a precise definition than a number of examples. 13) Will the target population be involved in any step of the review process? It might be useful to involve patients to identify (additionally) relevant moderators and to discuss review results with respect to their relevance for patients. 14) What is meant by ‘synchronizing variables’ on p. 7? What type of harmonization will be applied? Might it be useful to extract the better-quality measure for a specific symptom domain (e.g., favoring clinician-administered ratings over self-reports) compared to simply averaging across effect estimates for the same outcome? 15) Will the quality assessment also be performed by two independent raters? 16) Are there any ideas on studying between-moderator interactions? This might be interesting when different level-I moderators are available from a sufficient number of studies.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

1) Line 91: “We will conduct a systematic literature search…”. This search has been already partially conducted (Line 199). It just needs to be updated with publications published after January 12, 2022.

Answer: We thank the reviewer for this keen observation. We have transformed this section into the past tense (see line 109) and now refer to the section “Current status” at the beginning of the “Methods and analysis” (see lines 97-98).

“Disclaimer: the systematic literature search and identification and selection of studies have been partially conducted. For more details see the section “Current status”.”

Moreover, we specify that the search for relevant reviews and meta-analyses as a data source will be updated before conducting analyses (see lines 116-118).

“This search for relevant records provided by newly published reviews and meta-analytic work will be repeated before conducting the analyses.”

2) Regarding literature search, did the authors choose a time-range? Looking at the search string files (“OSF_search_strings.pdf” and “search_syntax_.pdf”) it seems that no time-range has been included. I suggest specifying in the manuscript that no time-range has been considered (or, if one was considered, to specify which was the range).

Answer: We thank the reviewer for this remark. Indeed, we did not choose a time range and specified this now in the method section (see lines 113-114)

“The time range was not specified.”

and in the abstract (see line 10).

“A systematic literature search will be conducted from database inception in the databases PsycINFO, Cochrane, Embase, PTSDpubs, and Web of Science.”

3) It should be taken into account that the older the study the higher the probability of not having access to the requested full individual dataset. It is likely that analysis of IPD will include mainly newer studies, analysis of aggregate data mainly older ones.

Answer: We agree with the reviewer on this point and will discuss this limitation in the main paper. In this study protocol the issue of availability of IPD is mentioned as a bullet point in the strengths and limitations of this study (see line 35).

“IPD meta-analysis is limited by the availability of IPD and their quality.”

4) Line 128: “…multiple imputation will be used…”. Do the authors foresee to use multiple imputation for missing outcome data on the global database (constructed by pooling all the received databases) or on each single database?

Answer: We thank the reviewer for raising this question. We will use multiple imputation per trial and specified this in the method section (see lines 176-177).

“Multiple imputation per trial will be conducted using 100 imputations through the mvn method in STATA software, StataCorp, as recommended by [33].”

5) Line 128: “selection of the reported result is not applicable…”. This is of course right. However, as it is likely that in several cases no full dataset will be available, thus the authors will have to perform analyses of aggregate data, this possible bias should be considered as well.

Answer: We thank the reviewer for mentioning this limitation. We realise that this IPD meta-analysis is limited by the availability of IPD provided by the authors. We added this as a bullet point in the section “Strengths and limitations of this study” (see line 35).

“IPD meta-analysis is limited by the availability of IPD and their quality.”

As the reviewer rightfully pointed out, in order to compare effects of both type of trials, i.e., those that provided IPD and those that did not, we will have to conduct an aggregate data meta-analysis (see lines 190-194).

“Additionally, to compare effects of both type of trials, i.e., those that provided IPD and those that did not, an aggregate data meta-analysis using a two-stage approach including all IPD (transformed) and available meta-data from study reports will be conducted. This is particularly advisable when a large proportion of authors did not share their datasets [33-34]. Results from both the one- and two-stage approach will be compared and discrepancies will be discussed [35].”

6) Line 133: “…PTS symptoms assessed at post-intervention (PT) and follow-up (FU)”. How do the authors define PT and FU? PT, when symptoms are assessed immediately after treatment and FU at a later time? Or, it depends upon the definition of each single study? In that case, how will the authors combine outcomes assessed at different timepoints?

Answer: We thank the reviewer for this question. We define post-intervention as outcomes assessed immediately after treatment and follow-up as outcomes assessed at any later time. We added this

specification in the method section (see lines 159-161). Additionally, we specified that the time of assessment will be added as a moderator in the models (see lines 166-169).

7) Line 138: "…resettlement as rationale for exclusion". Why excluding these individuals? Because the outcome was not assessed? In that case the exclusion criterium should be no outcome. Maybe there are other reasons for excluding those who have been resettled?

Answer: We thank the reviewer for this question. Originally, we intended to exclude these individuals from analyses as the rationale for exclusion is most likely due to external reasons. However, as the intention-to-treat principle conventionally ignores any kind of rationales for exclusion and does not differ between them (Gupta, 2011), we decided against this exception (see lines 174-175).

8) Line 146: "…when a large proportion of authors…". Maybe the authors could define what they mean with "large proportion".

Answer: We see the value of the reviewer's proposal. Although we would like to further define "large proportion" in the context of unavailability of individual patient data (IPD), the authors of the mentioned references (i.e., Riley, Simmonds, Look, 2007; Stewart & Tierney, 2002) did not specify this in their guidelines. As the proportion of available IPD will not influence the type of statistical analyses we will use, we prefer not defining this matter any further and thank the reviewer in advance for their understanding.

9) Page 9: The authors foresee to run many multivariate regression models for primary and secondary outcome variables. Do the authors plan to correct for multiple comparisons?

Answer: We thank the reviewer for this important question. As we will run several analyses, we will correct for multiple testing for analyses including secondary outcome variables (see lines 185-187).

"As we will run several analyses with different outcome variables, we will correct for multiple testing (i.e., Bonferroni adjusted p-values) for analyses including secondary outcome variables."

Reviewer 2

Major remarks:
1) The following statement is incorrect (p. 4): "While several meta-analyses have shown different psychological interventions to effectively reduce PTS, there is a considerable heterogeneity among studies (e.g., Kip et al., 2020) which may be attributed to differences in characteristics related to beneficiaries, providers, intervention, and study design. However, factors contributing to this heterogeneity have not yet been explored."
Many systematic reviews have explored moderators of treatment effects. There is also an IPD analyses on PM+ (Akhtar et al., 2021) and another ongoing much needed IPD analysis on PM+ and SbS (de Graaff et al., 2022, with preliminary results also being available: http://strengths-project.eu/wp-content/uploads/2023/03/STRENGTHS-D7.4-Modelling-moderators-for-health-outcomes-of-PM.pdf). The state of evidence should be reported in greater detail by clearly stating what the current systematic review can (and cannot) add beyond previous reviews in the field (I think there is still a lot that can be added – also beyond the WHO interventions).

Answer: We thank the reviewer for this important observation. We corrected the statement, extended the state of evidence, and described novel insights to which this IPD meta-analysis can contribute (see lines 54-64).

“While several meta-analyses have shown different psychological interventions to effectively reduce PTS, there is a considerable heterogeneity among studies [14-15], some of which have been investigated and attributed to differences in study characteristics. For example, [16] found that treatment effects of narrative exposure therapy increase if the providers themselves have a displacement background. While randomised-controlled trials (RCT) and trials with an active control group seem to be associated with smaller treatment effects [17-18], findings with regard to treatment dose (i.e., number of sessions) tend to be mixed, with evidence for more sessions boosting the treatment effect [17-18] or having no impact [14]. However, many tested moderators did not seem to influence treatment effects across studies including medication rate, time since displacement [14], residence status [19], use of interpreter [17], type of PTS assessment [17-18], study quality, country where trial was conducted, or ethnicity [18].”

2) The authors also state on p. 5 that interventions are overall effective. I agree that there are many effective interventions, however, there is substantial evidence that some psychosocial interventions are less or even not effective in forcibly displaced persons (e.g., Turrini et al., 2021). I think this should be reported more nuanced as it even makes a greater point for the current project.

Answer: We thank the reviewer for this comment. We made this statement more nuanced by referring to “many existing interventions showing overall efficacy” (see line 65).

3) It is not entirely clear what are the primary and/or secondary outcomes of this review and how were they chosen by the review team. Why is there a focus on PTSD and why is the availability of data on PTSD an inclusion criterion when interventions do not need to focus on PTSD? Why are other symptoms analyzed as comorbid symptoms? (PTSD is the primary disorder in this population, focus on this one) In general, the protocol reads like this review focuses rather on any intervention aiming to reduce mental distress in forcibly displaced persons and the focus on PTSD derives from the need of defining a primary outcome. I think this might be critical as some of the included interventions might also be less effective for PTSD symptoms (e.g., Schäfer et al., 2023, for PM+). What is the role of other outcomes like attrition, adverse events, treatment non-responses? Given the great potential of IPD meta-analyses for level-I moderators, I believe that the latter outcomes are of major interest. For example, adverse events are quite seldom in most of the trails, and it might provide very important knowledge on their predictors to have a detailed look at adverse events in an IPD meta-analysis. The same applies to attrition – that is a major issue, and the current project has the potential to shed light on this problem.
It might also be good to include a table defining primary and secondary outcomes along with (preliminary) moderators of interest.

Answer: We thank the reviewer for their remark. We have stated the primary and secondary outcomes in lines 159-164.

“As PTSD is the most prevalent mental disorder in FDP [5], the primary outcome will be PTS symptoms assessed at post-intervention (PT; i.e., immediately after treatment) and follow-up (FU; at any later time). However, in order to paint a more complete picture, we will run analyses with secondary outcomes including positive mental health outcomes (e.g., well-being), psychopathology (e.g., depression), disability, functioning, and quality of life at PT and FU assessments, as well as adverse outcomes, attendance, attrition, and treatment non-response.”

As the reviewer correctly points out, PTSD is the primary mental disorder and the most studied outcome in this population. We believe that choosing the most commonly investigated and prevalent symptoms as the primary outcome will likely lead to a higher number of included studies and, therefore, also to a higher availability of IPD. Moreover, findings will more likely be relevant for researchers of the field and the target population. However, as other symptoms, e.g., depression, are

also common in FDP (Blackmore et al., 2020), we included mental health outcomes other than PTSD as secondary outcomes in our analysis to paint a more complete picture. Furthermore, we agree with the reviewer on the importance of other outcomes such as attrition. We included this and other variables as secondary outcomes (see lines 161-164) and mention the advantage of IPD-MA to shed light on moderators of rare events such as adverse events in the introduction (see lines 80-84).

“By merging different IPD datasets with each other, a much larger sample size is reached than when looking at a single-study dataset, an advantage which allows for more complex analyses with statistical power and precision large enough to detect significant moderators of treatment effects and examine predictors of rare events such as adverse outcomes [25].”

4) This goes hand in hand with another rather general remark: It should be clearer from the protocol what are the advantages of IPD meta-analyses above traditional meta-analyses. The great potential of IPD is to shed light on level-I moderators with greater statistical power and based on a larger number of events (e.g., in case of adverse effects). This should be stated more clearly in the Introduction section and should also guide the choice of moderators as results will not be substantially different from standard reviews for study-level moderators (e.g., interventions type, type of intervention providers), but will have the potential to substantially enlarge knowledge on participant-level moderators.

Answer: We thank the reviewer for this remark. We edited the introduction section accordingly (see lines 84-86).

“Additionally, the use of an IPD-MA will allow us to shed light also on moderators of treatment effects at client-level, something previous traditional meta-analyses using reported meta-data could not address as they are restricted to moderators at study-level [27].”

5) Inclusion criteria for the IPD meta-analysis are very broad ranging from psychosocial intervention to standard psychotherapy. From the protocol, it is not entirely clear how these heterogeneous interventions will be handled in the analyses. This should be addressed more explicitly – is it reasonable to assume that moderators are equally important for different types of interventions or will meta-analyses on different intervention types be performed separately? Can one expect comparable effects from rather preventive psychosocial interventions and standard (and probably trauma-focused) psychotherapy? Might these broad criteria introduce unwanted heterogeneity in the analyses? I would be very interested to hear the authors’ ideas on this issue.

Answer: We thank the reviewer for their thought-provoking comment. Since we are using a data-driven approach, as is often customary (Lyman & Kuderer, 2005), we refrain from making any assumptions such as specific moderators having a differential effect on different types of interventions. Therefore, we will not run subgroup analyses for different types of interventions. However, we will explore the impact of intervention type (i.e., low-threshold interventions vs. specialised therapy) on treatment effects by including this variable as a moderator in our analyses (see lines 169-171).

“In order to examine differences in treatment effects, we will include type of intervention (i.e., low-threshold interventions vs. specialised therapy) as a moderator in the analyses.”

Additionally, we added the advantage of including both type of interventions in our IPD-MA (see lines 87-89).

“Moreover, by including trials using specialised and low-threshold interventions we will be able to examine whether interventions delivered by specialists and non-specialists differ in terms of treatment effects and moderators thereof.”

6) Overall, the two-stage approach is sufficiently described, while the multi-level model for the one-stage approach is not described in sufficient detail. What kind of random and fixed effects will be tested? What kind of model diagnostics will be applied? Will the authors check the (potentially non-normal) distribution of intervention outcomes? I would recommend to perform model diagnostics using DHARMa (https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html).

Answer: We thank the reviewer for this comment and suggestion to use DHARMa for model diagnostics. Before running any main analyses, we will run all necessary assumptions for linear regression models as described in lines 171-173.

“Before running any main analyses (see below), we will first test all assumptions necessary for linear regression models using DHARMa (https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html).”

We have specified the random and fixed effects in our multilevel model for the one-stage approach (see lines 191-194).

“To investigate treatment effects of psychological and psychosocial interventions, we will perform a multilevel mixed-effects linear regression model with a random effect for each trial and fixed effects for intervention condition (treatment vs. control) and severity of PTS symptoms at baseline.”

Minor remarks:
7) The title ‘study protocol of a series of individual patient data meta-analyses‘ might be irritating for readers as it may suggest that there is a series of (independent) IPD analyses. However, the protocol describes the analyses of one large dataset rather than a series of analyses/projects.

Answer: We understand the reviewer’s concern with regard to the title. We agree, we will mainly run analyses with one large dataset and, therefore, changed our title accordingly.

“What works for whom and why? Treatment effects and their moderators among forcibly displaced people receiving psychological and psychosocial interventions: study protocol of an individual patient data meta-analysis”

8) Information on the combination of a one- and two-stage approach and secondary outcomes is missing in the Abstract. The Abstract in general should draw a more complete picture of what is planned in the review.

Answer: We thank the reviewer for this remark. We adjusted the abstract accordingly (see lines 10-22).

“Methods and analysis: A systematic literature search will be conducted from database inception in the databases PsycINFO, Cochrane, Embase, PTSDpubs, and Web of Science. Only studies published in English, German, French, Spanish, Portuguese, and Dutch will be considered. Retrieved records will be screened for eligibility. Randomised controlled trials on adult FDP receiving psychological and psychosocial interventions aimed at alleviating symptoms such as PTS compared to a control condition without intervention will be included in this IPD-MA. Subsequently, authors of eligible studies will be contacted to request individual patient data (IPD). All datasets obtained will be synthesised into one large dataset which will be analysed using a one-stage approach by conducting

mixed-effect linear regression models (i.e., primary analysis). Additionally, an aggregate data meta-analysis using a two-stage approach by conducting a multivariate regression model including all IPD (transformed) and available meta-data from study reports (i.e., secondary analysis). PTS will serve as primary outcome measure, while mental health outcomes other than PTS, attendance, attrition, treatment non-response, and adverse outcomes will be examined as secondary outcomes."

9) There is a one- and a two-stage approach IPD, with the two-stage approach using aggregated data (but calculated from IPD). In the manuscript, this is sometimes mixed, with referring to meta-analyses on aggregated data (derived from IPD) as traditional meta-analysis. However, the latter would rather be an analysis based on data reported in a paper. This should be differentiated throughout the manuscript.
For example (p. 3): "IPD, compared to aggregate data meta-analysis, will allow more complex analyses to identify moderators of treatment effects while standardisation of variables is facilitated, and missing values can be accounted for."

Answer: We thank the reviewer for this pointing out this important difference in terminology. We now refer to "aggregate data meta-analysis" for meta-data derived from IPD, while the term "traditional meta-analysis" refers to meta-data from reports.

10) On p. 4: Step-by-step is guided by an eHelper (Carswell et al., 2018), I would not refer to the program as unguided.

Answer: We thank the reviewer for rightfully pointing out that "Step-by-Step" is not a good example for an unguided program. We actually used it as an example for a guided self-help program and now specified this in the text more clearly (see lines 49 and 103).

11) The authors may also think of automated quality checks for IPD – this could be helpful when obtaining data from a large number of studies. It might also be good to add an estimate how many primary studies will be expected (based on knowledge from previous systematic reviews in the field).

Answer: We thank the reviewer for this suggestion. We agree and, therefore, added automated quality checks for IPD to our protocol (see lines 137-138). We also added the success rate of obtaining IPD from authors, as well as the measures taken to increase this number (see lines 133-136).

"According to [28], the success to obtain IPD from authors is moderator (i.e., 58% success rate). In order to incentivise authors to share their data, we will offer two co-authorships per trial and contact all authors of each article at least three times, as suggested by [29]."

12) What kind of statistical program will be used for the one-stage and the two-stage meta-analysis? Multiple imputations will be done in Stata, will other analyses be performed in R?

Answer: We thank the reviewer for raising this question. We specified the programs used for each analysis in the main text (see lines 187-190).

"Analyses of the one-stage approach will be conducted using the STATA software, while all analyses of the two-stage approach and assumptions tests will be performed using the statistical program R (https://www.r-project.org/)."

13) It might be useful to describe the multivariate outcome approach for the two-stage approach in greater detail. What kind of multilevel model will be used? How will they examine effects on single outcomes? Will they model primary and secondary outcomes in one model? Will they obtain

information on between-outcome correlations? Moreover, it is not entirely clear how standardized mean differences (SMDs) will be calculated? Do the authors compare both groups at post-intervention and follow-up assessment? This should be stated more explicitly.

Answer: We thank the reviewer for this input. We have accordingly described the two-stage approach in greater detail (see lines 199-210).

“Two-stage approach: analysis of aggregate data (secondary analyses). First, we will calculate effect sizes for each trial separately and then compare them across studies by running aggregate data meta-analyses including both, trials providing IPD and studies providing only meta-data, in order to examine potential discrepancies in results. Thus, we will run multivariate meta-analyses with standardised mean differences (i.e., Hedges g; [36]) using a random-effects model estimated by restricted maximum likelihood accounting for differences in trials [37-38]. In order to identify moderators of treatment effects, we will first run several multiple linear regression models, including all potential moderators as independent variables and change in PTS symptom scores from baseline to PT and FU assessments as dependent variables for each trial separately. The obtained standardised regression coefficients will then be used as dependent variables when running several multivariate regression models with a random effect controlling for trial for each moderator separately. This procedure will be repeated for all secondary outcome variables mentioned above.”

14) It might be better to make use of contour-enhanced funnel plots instead of simply using funnel plots. Another minor remark: How can Egger's regression test be applied when a multivariate meta-analysis is performed (normally making use of a multilevel approach)? In those cases, one may rather approximate regression tests by including sampling error to meta-regression models.

Answer: We thank the reviewer for their suggestion and implemented these in the main text with regard to the funnel plots (see lines 217-219).

“Publication bias (two-stage approach). We will assess publication bias by creating “contour-enhanced funnel plots” for a visual evaluation of asymmetry [42] and applying the “trim and fill” method [43].”

The reviewer rightfully pointed out that Egger’s regression test cannot be applied to a multilevel approach. We therefore deleted this part in the main text.

15) I am also not sure with the following statement (p. 9): “In order to identify moderators of treatment effects, we will run several multiple linear regression models, including intervention condition and all potential moderators as independent variables and PTS symptom scores at PT and FU assessments as dependent variables for each trial separately.“ Does this refer to meta-regression analyses? In case the authors refer to standard regression models, I would recommend using meta-regressions instead.

Answer: We thank the reviewer for raising this question. We realised that this part in the text caused confusion. This step describes how we will obtain standardised regression coefficients which will then be analysed in multivariate regression models in the next step. We have clarified this text (see lines 204-207), which now reads:

“In order to identify moderators of treatment effects, we will first run several multiple linear regression models, including all potential moderators as independent variables and change in PTS symptom scores from baseline to PT and FU assessments as dependent variables for each trial separately. The obtained standardised regression coefficients will then be used as dependent variables when

running several multivariate regression models with a random effect controlling for trial for each moderator separately."

16) In same places (e.g., p. 3) the authors use the term 'treatment trajectories'. What is meant by that? It would be good to define this more explicitly. Will they also use other models than linear mixed models (e.g., growth mixture modeling) to examine distinct trajectories?

Answer: We thank the reviewer for pointing this out. We understand that the term "treatment trajectories" may be misleading. We changed it to "treatment outcome" as we do not plan to run any additional analyses that have not already been described in the section "statistical analysis".

17) Will moderator analyses only be performed for posttraumatic stress symptoms (p. 5)? I think it might be beneficial to have a broader focus. Also as being diagnosed with PTSD is not an inclusion criterion.

Answer: We thank the reviewer for this remark. As we described in lines 197-198 and line 210, we will perform moderator analyses not only for PTS symptoms but also for all other secondary outcomes.

18) The definition used for "non-active control" is very broad. Why is psychoeducation a non-active control? It might be better to give a precise definition than a number of examples.

Answer: We thank the reviewer for addressing this point. As the purpose of this IPD-MA is to determine the pooled efficacy of trials that evaluated psychological or psychosocial interventions and moderators thereof, we contrasted these interventions with interventions that did not involve these components. To that end, we defined the control conditions as treatments comprising no treatment, wait-list, or care-as-usual. The reviewer rightfully pointed out that psychoeducation is not a non-active control condition. Therefore, we deleted it as an example and will exclude control conditions using psychoeducation. Furthermore, we changed the term "non-active control group" to "control condition without intervention" (see line 103-104).

19) Will the target population be involved in any step of the review process? It might be useful to involve patients to identify (additionally) relevant moderators and to discuss review results with respect to their relevance for patients.

Answer: We thank the reviewer for this question. Although we generally think that patient involvement can be beneficial when conducting scientific research, we will not consult patients for this particular project. As the moderator variables, which will be included in the final analyses, will depend on availability of IPD, an a priori evaluation of relevant moderators will not affect the decision process. For the discussion of the results and their relevance for patients we will, however, consult clinicians working with forcibly displaced patients.

20) What is meant by 'synchronizing variables' on p. 7? What type of harmonization will be applied? Might it be useful to extract the better-quality measure for a specific symptom domain (e.g., favoring clinician-administered ratings over self-reports) compared to simply averaging across effect estimates for the same outcome?

Answer: We thank the reviewer for this remark. We explain the process of synchronisation in more detail in the main text (see lines 141-145).

"After confirming the accuracy of each dataset, we will first synchronise variables of interest to the same scale or categorical order and then merge the data into one large IPD meta-analytic dataset. If

variables were assessed by several measures, the method with the highest quality standard will be selected (e.g., clinical interviews will be favoured over self-report measures)."

21) Will the quality assessment also be performed by two independent raters?

Answer: We thank the reviewer for this question. We have indicated in line 149 that we will indeed have two independent raters for the quality assessment.

22) Are there any ideas on studying between-moderator interactions? This might be interesting when different level-I moderators are available from a sufficient number of studies.

Answer: We thank the reviewer for this interesting question. As we are following a data-driven approach, we do not make any assumptions with regard to between-moderator interactions and, therefore, did not include these in the section "statistical analysis".

References

• Blackmore, R., Boyle, J. A., Fazel, M., Ranasinha, S., Gray, K. M., Fitzgerald, G., ... & Gibson-Helm, M. (2020). The prevalence of mental illness in refugees and asylum seekers: A systematic review and meta-analysis. PLoS medicine, 17(9), e1003337.
• Gupta, S. K. (2011). Intention-to-treat concept: a review. Perspectives in clinical research, 2(3), 109.
• Lyman, G. H., & Kuderer, N. M. (2005). The strengths and limitations of meta-analyses based on aggregate data. BMC medical research methodology, 5(1), 1-7.
• Riley, R. D., Simmonds, M. C., & Look, M. P. (2007). Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. Journal of clinical epidemiology, 60(5), 431-e1.
• Stewart, L. A., & Tierney, J. F. (2002). To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Evaluation & the health professions, 25(1), 76-97.

VERSION 2 – REVIEW

REVIEWER	Emberti Gialloreti, Leonardo University of Rome Tor Vergata, Department of Biomedicine and Prevention
REVIEW RETURNED	01-Jan-2024

GENERAL COMMENTS	In my opinion, the major inaccuracies of the previous version have now been corrected. I believe that the revised version of this study protocol, amended based on the comments of the editor and reviewers, is ready for publication.

REVIEWER	Schäfer, Sarah K Technische Universität Braunschweig
REVIEW RETURNED	03-Jan-2024

GENERAL COMMENTS	I would like to thank the authors for their very careful and comprehensive revision of the review protocol. All my comments have been adequately addressed and the revised version of the protocol is much clearer and more stringent. I therefore recommend the acceptance of the protocol in this form.