PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol.

Section 1: Administrative Information.

**Title.**

*Item 1a. Identification. Identify the report as a protocol of a systematic review.*

The (comparative) effect of group threat, group contact, socialization, media, and the demographic variables age, sex, and education, on majority to minority inter-ethnic attitudes in English, peer-reviewed research articles published in the 2010-2022 period: protocol for a systematic review.

**Registration.**

*Item 2. If registered, provide the name of the registry (such as PROSPERO) and registration number.*

This systematic review protocol was registered on the Open Science Framework (OSF) (registration number NUMBER) and the GitHub page of the corresponding author (<https://github.com/jwgsim/Paper-Meta-Analysis-Interethnic-Attitudes>) on 22-07-2022 and has since not been updated.

**Authors.**

*Item 3a. Contact information. Provide name, institutional affiliation, and email address of all protocol authors; provide physical mailing address of corresponding author.*

Authors: Jan-Willem Simons (JWS)1, Eva Jaspers (EJ)1, Frank van Tubergen (FT)1.

Corresponding Author: Jan-Willem Simons.

Author Affiliations  
1 - Department of Sociology, Utrecht University, Sjoerd Groenmangebouw, Padualaan 14, 3584 CH Utrecht.

Email: Jan-Willem Simons - [j.g.simons@uu.nl](mailto:j.g.simons@uu.nl), Eva Jaspers - [e.jaspers@uu.nl](mailto:e.jaspers@uu.nl), Frank van Tubergen - [f.vantubergen@uu.nl](mailto:f.vantubergen@uu.nl).

*Item 3b. Contributions. Describe contributions of protocol authors and identify the guarantor of the review.*

JWS is the guarantor. JWS will draft the manuscript. All authors will contribute to the development of the selection criteria, the risk of bias assessment strategy, and data extraction criteria. JWS developed the search strategy. EJ provided expertise on contact theory and socialization effects. FT provided expertise on group threat theory and media effects. All authors will read, provide feedback, and approve the final manuscript.

**Amendments.**

*Item 4. If the report represents an amendment of a previously completed or published protocol, identify as such and indicate what changes were made; otherwise state plan for documenting important protocol amendments.*

Version control is used to track and document protocol changes in OSF and on the GitHub page of the corresponding author (<https://github.com/jwgsim/Paper-Meta-Analysis-Interethnic-Attitudes>). In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale for doing so.

**Support.**

*Item 5a. Sources. Indicate sources of financial or other support for the review.*

This systematic review is funded by the Volkswagen Foundation.

*Item 5b. Sponsor. Provide name of the review funder and/or sponsor.*

This systematic review is sponsored by Utrecht University.

*Item 5c. Role of sponsor and/or funder. Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol.*

The Volkswagen Foundation is funding the systematic review. Utrecht University is the sponsor of the systematic review. Neither the Volkswagen Foundation nor Utrecht University are involved in any other aspect of the systematic review. The funder and sponsor will have no input on the interpretation or publication of the study results.

Section 2: Introduction.

**Rationale.**

*Item 6. Describe the rationale for the review in the context of what is already known.*

[Review title: A systematic review and meta-analysis of the comparative effect of group threat, group contact, socialization, media, and the demographic variables age, sex, and education, on majority to minority inter-ethnic attitudes in English, peer-reviewed research articles published in the 2010-2022 period.]

Negative interactions between ethnic groups are associated with a host of adverse outcomes. Negative ethnic attitudes are the antecedent to negative behaviours towards members of other ethnic groups (Wagner, Christ, & Pettigrew, 2008). As such, improving interethnic attitudes can function as a pathway for improving interethnic interactions and relations.

Over the previous decades, much research has been conducted on the determinants of attitudes between ethnic groups. Amongst these, group threat and group contact have generally received the most attention. In recent years, other determinants have increasingly been considered, such as socialization, media, and demographic effects. The cumulative body of work across these various determinants has grown at a speed and size that makes it hard for researchers to keep up. This research is furthermore being conducted across multiple disciplines, such as (social) psychology, sociology, political science, and communication and media studies. As such, research is splintered across multiple fields of research, each using their own definitions and measures, making it difficult to identify coherent messages.

Against this backdrop, the objective of this systematic review is to describe the relative efficacy of group threat, group contact, socialization, media, and the demographic variables age, sex, and education for explaining attitudes between ethnic groups. In doing so, it seeks to integrate findings across multiple literatures and to present a coherent message about the relative efficacy of these various determinants.

**Objectives.**

*Item 7. Provide an explicit statement of the question(s) the review will address with reference to populations, interventions, comparators, and outcomes (PICO).*

What is the (comparative) effect of group threat, group contact, socialization, media, and the demographic variables age, sex, and education, on majority to minority inter-ethnic attitudes in English, peer-reviewed research articles published during the 2010-2022 period.

Section 3: Methods.

**Eligibility Criteria.**

*Item 8. Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review.*

Studies will be selected according to a set of study eligibility and review eligibility criteria. According to the PRISMA-P guidelines, study eligibility criteria are “the typical PICO elements that form the basis of clinical questions. These include populations, interventions, comparators, outcomes, time frames for follow-up, settings in which the interventions are delivered, and study designs of interest.” The time frames for follow-up criterium is not discussed because it is not applicable in the context of sociological research. The review eligibility criteria identified by PRISMA-P are “likely to include geographical location, languages of publication, publication status (such as inclusion of unpublished material or abstracts), and years of publication.”

Populations.

The populations that will be included in the systematic review are ethnic majority groups that hold an attitude toward an ethnic minority group.

Interventions

The interventions that will be included in the systematic review are group threat, group contact, socialization, media and the demographic variables age, sex, and education. Starting with the demographic variables, we define these as the chronological age, sex, and highest level of education received, as indicated by the respondent. Although these demographic variables have their own complexities, they generally refer to the same concept and are measured in uniform ways.

With respect to the group threat, group contact, socialization, and media interventions, we need to impose an additional layer on both the conceptual and measurement level, because these are not homogeneous terms. On the conceptual level, we a-priori apply a conceptual distinction for each of these determinants on the basis of theory. For group threat we make a distinction between realistic and symbolic threat. For group contact, we make a distinction between direct and extended contact. We do not consider imagined contact or vicarious contact that is not facilitated through media. For socialization, we make a distinction between the effect of parents, peers, and the in-group in which the respondent is embedded. Finally, for media, we make a distinction between parasocial contact - where the respondent observes the ethnic outgroup through media - and vicarious contact - where the respondent observes fellow ethnic ingroup members interacting with ethnic outgroup members through media. Note that this a-priori framework will be used to guide parts of the systematic review. Examples are the keywords that will be used in the literature search and the theoretical categories of a determinant that will be scrutinized and compared in the meta-analysis. From a practical, empirical standpoint, it can furthermore be hard to a-priori specify the outline of a meta-analysis on the basis of theory alone, in the sense that there might not be sufficient information in the form of estimates from articles. As such, we inspect the degree to which the retrieved literature sufficiently represents the a-priori theoretical framework, and introduce empirical, post hoc conceptual distinctions in the event that it is found to be lacking. Put differently, given that the results from a literature search on the basis of the a-priori framework are unable to ensure sufficient statistical power, we identify the set of most common conceptualizations of the determinants within that set, and use these as our post hoc conceptualizations. Finally, with respect to how these various concepts are measured, due to high levels of heterogeneity that we might expect in these various reach fields, we might observe variance such that estimates that measure the same concept become incomparable. One way to account for this is to make a measurement division within a concept, and use that as input for the meta-analysis. Another approach is to use a structural equation model, where the various measures load on a latent factor. For now, our intention is to use the second approach.

Comparators

Comparators are generally not very prominent in the context of sociological research because most studies use an observational design. As such, the comparator element of PICO is not explicitly incorporated in the research question, i.e., we do not limit our study-level inclusion criteria to such comparators. Comparators will however be relevant for any (quasi-)experimental studies that might be included in the review. However, due to the relative scarcity of these types of studies, baseline comparators are generally not well developed in the field of sociology. As such, inclusion criteria with respect to comparators will not explicitly be formulated, but judged on a case-by-case basis.

Outcomes

Outcomes are inter-ethnic attitudes, specifically majority out-group attitudes such as liking and disliking of, or prejudice towards a minority group. Other forms of attitudes that we consider are xenophobia, immigration attitudes, and immigration policy attitude/preferences. We furthermore limit ourselves to majority to minority attitudes because most studies focus on this attitude direction, and not on alternatives such as minority to majority or minority to minority attitudinal flows.

Setting

There will be no restrictions by type of setting.

Study designs.

Three types of study design will be included: experimental, quasi-experimental, and observational. We expect that most studies will use an observational design, followed by a quasi-experimental and experimental design. All other types of study designs are excluded.

Geographical location

There will be no restrictions by geographical location.

Languages of publication

We will only include articles reported in the English language.

Publication status

We will only include published, peer-reviewed research articles.

Years of publication

We will include only articles published during the 2010-2022 period.

**Information Sources.**

*Item 9. Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage.*

This systematic review intends to use electronic databases as its primary and only information source. Grey literature sources, such as contact with study authors, trial registers, and others will not be used as information sources. This choice was made to limit the scope of the systematic review. The specific electronic database sources that will be used are PsycInfo, Web of Science Core Collection, Scopus, and Sociological Abstracts. These four databases were chosen on the basis of Gusenbauer, & Haddaway (2020), who identify a set of “principal” search systems that conform to a set of usability and replicability standards. These databases and their overlaying search systems are the "most appropriate for evidence synthesis" (Gusenbauer, & Haddaway, 2020) given the topic and objectives of this systematic review. Planned dates of coverage are 01-01-2010 to XX-07-2022, where the exact day in this date depends on the day a specific determinant search was conducted.

**Search Strategy.**

*Item 10. Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated.*

The search strategy consists of two steps. In the first step, EJ and FT select six "gold standard" articles for the group threat, group contact, socialization, and media interventions that they deem to be typical articles to include in a systematic review on these determinants. We do not aim for these articles to be representative of the literature on a determinant, but as the initial input for a search that aims to obtain a substantially representative portion of that literature. More specifically, EJ selected these articles for the group contact and socialization interventions, where FT selected them for the group threat and media interventions. Note that we do not execute a separate search for the demographic interventions, as we anticipate that estimates for these variables will be present in articles on the other determinants. The keywords in these articles, those as indicated by the article authors themselves and those which are obtained by applying the rapid-automatic-keyword-extraction (RAKE) algorithm to the titles and abstracts of these articles, are used to extract a set of candidate keywords. These candidates are filtered manually based on relevance, where relevance is defined as being related in content to either the intervention of interest or the outcome, and whether they appear in the titles and abstracts of the "gold standard" articles for that intervention. In this sense, the literature search is “fit” to the keywords. An initial "naive" search is subsequently carried out. In line with our inclusion criteria, we restrict the time period to the 2010-2022 period and search for English documents peer-reviewed research articles. Note that our date endpoint is determined by the date that we searched the respective databases. This date will be indicated for each search. As noted earlier, planned dates of coverage are 01-01-2010 to XX-07-2022, where the exact day in this date depends on the day a specific determinant search was conducted.

In Ovid, we start by clicking on "Advanced Search", followed by "Change" where we select the APA

PsycInfo database. We then click on "Edit Limits", deselect the "Abstracts" box, and select the "Peer

Reviewed Journal" box, leaving the remaining boxes as is. We then click on "Customize Limits, check the "Peer Reviewed Journal" and "English Language" boxes, and set the "Publication Year" tabs to "2010" and "Current", respectively, We then enter the Boolean search in the search box and click "Search". Note that the "Keyword" and "Map Term to Subject Heading" boxes are checked by default and are left as is. The documents are subsequently exported with "Format:" set to "RIS" and "Fields:" set to "Complete Reference". If the total number of documents exceed 1500, this exporting process is executed in batches of size 1500.

In Proquest, we click on "Advanced Search", enter the Boolean search in the first line, set the value of the "in" bar to "Anywhere except full text - NOFT", check the "Peer reviewed" box, select "After this date..." in the "Publication date:" drop-down menu, and set the respective boxes to "January", "1" and "2010", respectively. Under "Source type:" we select "Scholarly Journals" and under "Language" we tick the "English" box. We subsequently click on "Search". On the next page, under "Document type" we subsequently select "Article", where under "Language" we select "English". We set "Items per page" to 100 and select the "EndNote" category under "All save & export options" for

exporting a RIS file for each page.

On the Scopus start page, we set the "Search within" tab to the default "Article title, Abstract, Keywords" tab. We do not set it to the "All fields" tab, because doing so 1) retrieves a large number of irrelevant documents, and 2) makes it difficult to export documents from Scopus. We subsequently enter the Boolean search in the "Search documents" tab. We click on the "Add date range" button and set the "Published from" tab to "2010", leaving the "To" tab to "Present", and the "Added to Scopus" tab to "Anytime". We click "Search" and after having searched, scroll down to "Document type" and "Language" under "Refine results", check the "Article" and "English" boxes, respectively, and click "Limit to". We subsequently select "All" and click on "Export" and select

"RIS Format". Note that I only export "Citation information", "Bibliographical information" and

"Abstracts & keywords" in Scopus because exporting the additional two categories "Funding details" and "Other information" leads to merging issues later on. Note that we stratify the returned document set by years when this number exceeds 2000.

In Web of Science, we click on "Advanced search", enter the Boolean search in the "Query Preview" search box, and click "Search". We once again scroll down to "Document Types" and "Languages" under "Refine results" and check the "Articles" and "English" and "Unspecified" boxes, respectively, if applicable. We subsequently click on "Select all records", "Export" and then "RIS (other reference software):". We subsequently click on "Records from:" and export either the total returned document set if this set is less than 1000, or export them in a per 1000 batch-wise fashion if this number exceeds 1000, with "Record Content:" set to "Full Record".

The naïve search result for an intervention is validated and accepted based on the criterion that all six gold standard articles for that intervention can be retrieved. If not, the search is repeated with different, often more specific keywords until the gold standard articles can be retrieved. Besides validating the naïve search result on the gold standard article set, we also validate it on a second external article set, which we retrieve by taking a snowball sample over the gold standard articles. More specifically, we enter the title of each gold standard article on the "connectedpapers.com" website, and select all papers that are suggested in the resulting network, and all papers under the "Derivative works" tab, that were published between the 2010-2022 period and are relevant to the query of interest as judged by the corresponding author.

Given the “naïve” search results, the second step consists of iterating on this search. This is done by extracting keywords from the corpus of the naive search, those keywords as listed by the authors, and keywords from the titles and abstracts with the RAKE. These keywords are filtered manually on the basis of relevance, and subsequently filtered again on the basis of being the most central in a keyword co-occurrence network (KCN) (Radhakrishnan, Erbis, Isaacs, & Kamarthi, 2017). It is assumed that keywords that occur frequently in articles in the corpus and with certain other words will be more effective in addressing the literature. About the first sixty of these keywords - which are the keywords with the highest network degree - are selected, because most search systems, such as Web of Science, do not allow more than about sixty words in a search string, as such providing a natural cut-off. This threshold is equal to fifty for search the Sociological Abstracts database in ProQuest. This iterative process continues until the coverage increase is less than 5 to 10% relative to the previous iteration. This usually occurs after three iterations. At that point, we assume that most of the literature with the initial naive search as its basis has been addressed. Note that we merge the search results of each iteration with the previous iteration to maximize coverage. Also note that we quantify search precision for each iteration by inspecting the degree to which the gold standard and especially the external article set can be retrieved. This too functions as a stopping criterion, i.e., if the increase in the number of external articles is small or zero between subsequent iterations, this is grounds for terminating the search.

All required files for replicating the literature search for each intervention, i.e., the “gold standard” article sets and R-files, will be made available in the GitHub repository of the corresponding author (<https://github.com/jwgsim/Paper-Meta-Analysis-Interethnic-Attitudes>) after the searches have been conducted. Please refer to the annotation of these R-files for additional, more specific information on the search procedure.

**Study Records.**

*Item 11a: Data management. Describe the mechanism(s) that will be used to manage records and data throughout the review.*

Literature search results will be managed in EndNote, R, and Zotero. The literature on an intervention often has to be retrieved from a database in batches when the number of retrieved documents is high, because most search systems do not allow for single, bulk imports. These batches are imported and merged in R. The merged result for a database is subsequently deduplicated in R. We remove duplicates within and between database search results with the suite of functions in the "synthesisr" package. We start with exact matching, and follow up with optimal string alignment to identify potential duplicates. More specifically, we use approximate string matching, which is prone to errors in that either duplicates are identified which are not duplicates - false positives - or duplicates are not identified which are - false negatives. We address false positives by manually inspecting duplicate candidates and only removing them if they are identified to be duplicates. We additionally import the merged, de-duplicated output into EndNote and Zotero to cross reference whether all duplicates have been removed. We use and combine these two reference managers because they have different sensitivity and specificity rates. As such, through manual assessment their combined use will enable us to identify duplicates more thoroughly. After the main de-duplication procedure in R, I start by screening duplicates in EndNote and removing these in Zotero, whether these are flagged by Zotero or not. In general, those documents in the duplicate set are retained which list the most information on the title, abstract, and keywords. In EndNote I use the "Author", "Volume", and "Pages" categories to compare references and identify duplicates. I subsequently screen and flag any remaining duplicates in Zotero that were not identified in EndNote. Zotero uses the title, DOI, and ISBN fields to identify duplicates. If these fields match (or are absent), Zotero also compares the years of publication (if they are within a year of each other) and author/creator lists (if at least one author last name plus first initial matches). After this screening procedure I remove retracted items as flagged by Zotero and EndNote, export the result from Zotero, and import it into R. The duplication choices are not explicitly noted in the R-script, but the retraction choices are. Note that I will store the corpus file that is de-duplicated in the reference managers along with the final de-deduplicated file for an iteration, so that the reader can cross-validate the de-duplication process. Note that it is possible for some duplicates to remain unidentified with this procedure, but not for non-duplicates to be accidentally removed, which is the most important to my judgment. Remaining duplicates will be identified and removed during the review process in ASReview and full-test screening.

*Item 11b: Selection process. State the process that will be used for selecting studies (such as  
two independent reviewers) through each phase of the review (screening, eligibility, and inclusion  
in meta-analysis.*

EJ and FT will independently screen the titles and abstracts yielded by the search against the inclusion criteria in ASReview until 100 articles that match the inclusion criteria are retrieved. EJ will screen the literature retrieved on contact theory and socialization, where FT will screen the literature on group threat and media. JWS will be present during screening to understand the decision process of EJ and FT. On that basis, JWS and three student assistants will independently screen the remaining set of titles and abstracts in ASReview, with a stopping criterion of 20% of total articles reviewed and one-hundred subsequent article exclusions. We will use Cohen’s kappa coefficient (κ) to quantify inter- and intra-rater reliability. We will resolve disagreement through discussion. We will record the reasons for excluding articles. During the screening process in ASReview, the review authors will be blind to the journal titles, the study authors, and the institutions. We will subsequently obtain full reports for all titles that appear to meet the inclusion criteria. EJ and FT will screen a sample of the full text reports and decide whether these meet the inclusion criteria. JWS will screen the remaining full text reports and decide whether these meet the inclusion criteria. We will use Cohen’s kappa coefficient (κ) to quantify inter- and intra-rater reliability. We will resolve disagreement through discussion. We will record the reasons for excluding articles. A simple random sample (SRS) of approximately 75 articles per intervention will subsequently be drawn from the final inclusion article set for final inclusion in the systematic review. We draw 75 articles for two reasons, (1) power calculations show that at least 40 studies are required for each intervention, under the assumption of a small effect size, a moderate sample size, and high study heterogeneity, and (2) we estimate that this is the approximate workload that we will be able to process given the time and resources available to us.

*Item 11c: Data collection process. Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators.*

Using standardized forms and a detailed instruction manual, JWS and three student assistants will independently extract data from the sample of research articles. Data will be extracted on characteristics of the study, the data, the methodology, the operationalization of the interethnic attitude variable and the intervention, and finally the effect size. Some article overlap will be induced between reviewers to gauge inter-rater reliability. Reviewers will resolve disagreements by discussion, and the corresponding author (JWS) will adjudicate unresolved disagreements.

**Data items.**

*Item 12. List and define all variables for which data will be sought (such as PICO items, funding sources) and any pre-planned data assumptions and simplification.*

We make a distinction between data extraction on the study and effect size level. We make this distinction because multiple effect sizes of interest can be reported within a single study, for each of which data will need to be extracted. On the level study level, we extract data on the characteristics of the study. More specifically, we assign the study an ID, and extract its title, Digital Object Identifier (DOI), the names of its authors, its publication year, and the type of manuscript in which it was made available. With regards to the effect size, we extract information on the data and the methodology with which the effect size is estimated, the operationalization of the interethnic attitude variable and the intervention, and finally information on the effect size itself. To be more specific, with respect to the data characteristics, we start by assigning to the effect size the ID of the study in which it is embedded, and subsequently record the source of the data and the year and geographic location in which it was collected. For the methodological characteristics, we list the research design of the study, the statistical method that was used, the majority in-group that holds a particular attitude towards a minority out-group, and the minority out-group to which that attitude is directed. We subsequently extract information on the operationalization of the interethnic attitude variable, namely the construct operationalization, measurement operationalization, and the minimum, maximum, whether a single or composite measure was used, and measurement scale of the measurement operationalization variable. We subsequently extract information on the operationalization of the intervention, namely which of the five intervention categories it represents, and subsequently the construct operationalization, measurement operationalization, minimum, maximum, whether a single or composite measure was used, and measurement scale of the measurement operationalization variable. We finally extract information on the effect size. In general, we want to obtain bivariate correlations, for which we extract the value of the coefficient and the associated sample size. We alternatively obtain information that enables us to calculate these correlations. We finally also extract semi-partial correlations, or alternatively information that enables us to calculate these correlations.

**Outcomes and prioritisation.**

*Item 13. List and define all outcomes for which data will be sought, including prioritisation of main and additional outcomes, with rationale.*

We will consider positive and negative inter-ethnic attitudes. More specifically, we consider majority out-group attitudes such as liking and disliking of, or prejudice towards, a minority group. Other forms of attitudes that we consider are xenophobia, immigration attitudes, and immigration policy attitude/preferences. We will not consider attitudes like social trust and social distance. We furthermore limit ourselves to majority to minority attitudes because most studies focus on this attitude direction, and not on alternatives such as minority to majority or minority to minority attitudinal flows.

**Risk of bias individual studies.***Item 14. Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis.*

Included non-randomised studies may or may not have a comparison group. To assess the risk of bias within included studies, the methodological quality of potential studies will be assessed by using the Newcastle-Ottawa scale (NOS) for assessing the quality of non-randomised studies in meta-analyses. The NOS for case-control and cohort studies will be adapted to meet the specific needs of this systematic review. A judgement as to the possible risk of bias for each included study will be made from the extracted information. These judgements will be made by JWS based on the criteria for judging the risk of bias. Disagreements will be resolved by discussion with EJ and FT.

**Data synthesis.***Item 15a. Describe criteria under which study data will be quantitatively synthesised*

If studies are sufficiently homogeneous in terms of design and risk of bias, we will conduct meta-analyses using a meta-analytic structural equation model (MASEM).

*Item 15b. If data are appropriate for synthesis, describe planned summary measures, methods of handling data, and methods of combining data from studies, including any planned exploration of consistency (such as I2, Kendall’s τ).*

Planned summary measures

Our planned summary measure is the Pearson correlation. We will use the equality of standardized regression coefficients in structural equation models using Wald tests to compare effect sizes based on this planned measure. We will additionally execute an exploratory analysis with semi-partial correlations as our summary measure.

Missing data  
When there are missing data, an imputation method will be used. We will use sensitivity analysis to assess the impact on the overall treatment effects of inclusion of articles with missing data.

Assessment of heterogeneity.

Statistical heterogeneity will be tested using the Chi2 test (significance level: 0.05) and I2 statistic (0% to 40%: might not be important; 30% to 60%: may represent moderate heterogeneity; 50% to 90%: may represent substantial heterogeneity; 75% to 100%: considerable heterogeneity). If high levels of heterogeneity among the trials exist (I2 >=50% or P <0.1) the study design and characteristics in the included studies will be analysed. We will try to explain the source of heterogeneity by subgroup analysis and sensitivity analysis.

Data synthesis  
Each outcome will be combined and calculated in R with the help of the “lavaan”, “metaSEM”, and “metafor” packages. The Mantel-Haenszel method will be used for the fixed effect model if tests of heterogeneity are not significant. If statistical heterogeneity is observed (I2 >=50% or P <0.1), the random effects model will be chosen. If heterogeneity is substantial, we will not perform a meta-analysis; a narrative, qualitative summary will be done instead.

*Item 15c. Describe any proposed additional analyses (e.g., sensitivity or subgroup analyses, meta-regression).*

In addition to the main analysis, we will conduct a sensitivity analysis, a subgroup analysis, and a meta-regression in the MASEM context.

*Item 15d. If quantitative synthesis is not appropriate, describe the type of summary planned.*

If quantitative synthesis is not appropriate, a systematic narrative synthesis will be provided with information presented in the text and tables to summarise and explain the characteristics and findings of the included studies. The narrative synthesis will explore the relationship and findings both within and between the included studies.

**Meta-bias(es).**

*Item 16. Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies).*

We will conduct publication bias and sensitivity analyses to investigate meta-bias(es).

**Confidence in cumulative estimate.**

*Item 17. Describe how the strength of the body of evidence will be assessed (such as GRADE).*

The quality of evidence for all outcomes will be judged using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group methodology. The quality of evidence will be assessed across the domains of risk of bias, consistency, directness, precision, and publication bias. Additional domains may be considered where appropriate. Quality will be adjudicated as high (further research is very unlikely to change our confidence in the estimate of effect), moderate (further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate), low (further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate), or very low (very uncertain about the estimate of effect).

References

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Research synthesis methods, 11(2), 181-217.

Nudelman, G., & Otto, K. (2020). The development of a new generic risk-of-bias measure for systematic reviews of surveys. *Methodology*, *16*(4), 278-298.

Pettigrew, T. F., Christ, O., Wagner, U., Meertens, R. W., Van Dick, R., & Zick, A. (2008). Relative deprivation and intergroup prejudice. Journal of Social Issues, 64(2), 385-401.

Radhakrishnan, S., Erbis, S., Isaacs, J. A., & Kamarthi, S. (2017). Novel keyword co-occurrence

network-based methods to foster systematic reviews of scientific literature. PloS one, 12(3), e0172778.