An integrative framework for planning and conducting NonIntervention, Reproducible, and Open Systematic Reviews (NIRO-SR).

Abstract

Most of the commonly used and endorsed guidelines for systematic review protocols and reporting standards have been developed for intervention research. These excellent guidelines have been adopted as the gold-standard for systematic reviews as an evidence synthesis method. In the current paper, we highlight some issues that may arise from adopting these guidelines beyond intervention designs, including in basic behavioural, cognitive, experimental, and exploratory research. We have adapted and built upon the existing guidelines to establish a complementary, comprehensive, and accessible tool for designing, conducting, and reporting Non-Intervention, Reproducible, and Open Systematic Reviews (NIRO-SR). NIRO-SR is a checklist composed of two parts that provide itemised guidance on the preparation of a systematic review protocol for pre-registration (Part A) and reporting the review (Part B) in a reproducible and transparent manner. This paper, the tool, and an open repository (https://osf.io/f3brw/) provide a comprehensive resource for those who aim to conduct a high quality, reproducible, and transparent systematic review of non-intervention studies.

1. Introduction

Systematic literature reviews are a widely used method for rigorously synthesising existing evidence to answer research questions and to inform best practice and policymaking. The quality of systematic reviews is contingent upon comprehensive, systematic, and transparent identification of all the relevant literature, followed by a balanced critical evaluation and synthesis of the data extracted from that literature. Rigorous implementation can minimise biases and questionable reporting practices that can lead to misleading or inconsistent conclusions (Ioannidis, 2016; Moher et al., 2009; Siddaway et al., 2019). However, the most popular guidelines for designing, reporting, conducting, and critically appraising systematic reviews to date have been designed for the synthesis of healthcare, medical, and intervention-based research. These include the PROSPERO protocol preregistration system and template (Booth et al., 2012); the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al. 2021); Cochrane Handbook for Systematic Reviews of Interventions (Higgins et. al., 2021); and the Assessing the Methodological Quality of Systematic Reviews tool (AMSTAR; Shea et al., 2017). The popularity of these tools is evident through endorsement- from a number of journals (see PRISMA endorsers for an example), university libraries, and collaborative groups specialised in conducting systematic reviews (see the list of recommended systematic review tools by the EQUATOR network). Therefore, these tools are widely chosen by authors of systematic reviews through recommendations, journal requirements, good findability, and/or greater accessibility. Moreover, they are a likely choice for authors who conduct systematic reviews based on studies other than interventions who reach for these tools through similar routes. For instance, some more general and multidisciplinary journals that publish various types of studies encourage or require that all submitted systematic reviews must follow the PRISMA

guidelines intended for intervention studies (e.g. *Systematic Reviews*, *PeerJ* or *PLOS One*), which may not always be well suited.

Intervention studies focus on assessing the efficacy or effectiveness of, for example, healthcare interventions and clinical trials that *a priori* assign participants to different intervention groups (Committee of Medical Journal Editors, 2021). As conceptualised by Glass (1972), the essential aim of intervention studies is to *evaluate* the proposed intervention and its effects. Many other types of research, such as basic, experimental, and exploratory research in the social, cognitive, and behavioural sciences, do not share the same aims as intervention research, and instead aim to explore and explain mechanisms, and thus evidence synthesis of such papers must be approached from a different perspective.

Box 1. Definitions of terminology:

Intervention Research: A study which aims to evaluate the effects of an intervention, often against another intervention, on primary or secondary outcomes of interest.

Non-intervention Research: A study which aims to provide an explanatory framework of an empirical phenomenon, or to provide supporting evidence for a theoretical paradigm (Glass, 1972; DeLucia & Pitts, 2012).

Intervention Systematic Review: A systematic review which synthesises results from intervention research.

Non-Intervention Systematic Review: A systematic review which synthesises results from non-intervention research.

Systematic Review Protocol: A protocol (ideally pre-registered, see "Systematic Review Pre-registration" below) which outlines specific plans for conducting the systematic review. It may be understood as a 'recipe' for the systematic review.

Systematic Review Pre-registration: A systematic review can be pre-registered by submitting the finished protocol to a pre-registration platform, such as the Open Science Framework.

Systematic map: A report of the ongoing research activity on a particular topic, informed by a systematic search and screening strategy, which can be used to identify gaps in research.

Methodological systematic review: Informed by a systematic search, this review summarises methodological practices or questions in a given area.

Research that does not fit the scope of intervention, such as explanatory, experimental, and basic research, should also adopt rigorous and transparent practises of conducting evidence synthesis, particularly in the context of the ongoing paradigm shift that places

emphasis on replicable and reproducible research (Munafò et al., 2017). However, researchers conducting systematic reviews of non-intervention research who wish to follow established guidelines must often resort to adapting the criteria of less applicable guidelines to make it appropriate to assess these types of studies, leading to ad hoc solutions such as filtering, combining, or customising practices from several sources (Macpherson & Jones, 2010). For instance, one popular tool is the "PICOS" framework (Population, Intervention, Comparison, Outcome, Study design) which aids the development of a research question and eligibility criteria for evidence synthesis. PICOS is an important component of the PROSPERO template, Cochrane guidelines, and the AMSTAR tool, and it was only recently removed from PRISMA following the 2020 update. This framework cannot always be directly applied to diverse research designs (Bramer, 2015) and many alternatives have been developed (Booth et al. 2019); for example the SPIDER framework (Cooke et al., 2012) for systematic reviews of research using qualitative methods.

More general guidelines which are not limited to intervention designs also exist. In the field of psychology a comprehensive tool, the Meta-Analysis Reporting Standards (MARS; Appelbaum et al. 2018), was recently proposed by the American Psychological Association (APA) Working Group on Quantitative Research Reporting Standards. The tool advances standards, but there are barriers to its implementation. Not all systematic reviews include meta-analyses, thus for many authors who decide not to include a meta-analysis component, MARS may initially be considered unsuitable. In addition, accessibility of MARS as a tool is limited because it is not an open-access resource. In fact, the uptake of MARS for evidence synthesis has been very limited and described as "non-existent" in a recent review (Hohn et al. 2020). Lastly, MARS is a reporting guideline, which in practice means that researchers may follow it retrospectively for reporting purposes only and are less likely to use it to inform the design of their study. In summary, although valuable resources exist for guiding the design and reporting of systematic reviews, researchers have a limited

choice when it comes to selecting an appropriate and accessible tool for systematic reviews beyond interventional research.

The utility of existing guidelines for high quality systematic reviews is limited by whether they are correctly applied. General problems with adherence to guidelines have been highlighted in the 2009 version of PRISMA (Page & Moher 2017) but also systematic reviews in psychology as a field (Hohn et al. 2020). Page et al. (2021) and Hohn et al. (2020) suggested that low adherence could be related to possible lack of guideline adherence checks during peer review, relaxed demands for adherence by journals, or variation in how checklist items are interpreted by the systematic reviewers. However, adherence rates might be significantly impacted by the guideline's appropriateness to specific fields. Problems with the use of guidelines differ across disciplines and may be driven by discipline-specific interpretation of items which can be further exacerbated by ambiguity and lack of clarity regarding item wording (Rethlefsen et al., 2021). This is specifically problematic for fields where non-intervention research is common (Gates & March, 2016). Therefore, the development of systematic review guidelines that cater beyond interventional designs and are appropriate for explanatory, experimental, and basic research could help to improve a guideline's adherence rate in fields such as psychology, neuroscience, and economics.

The lack of sufficient instructions accompanying guidelines may also contribute to the low adherence problem especially with regards to items designed to facilitate transparency and robustness of systematic reviews. For example, protocol pre-registration is one of the PRISMA items with a very low adherence rate. Considering that pre-registration is widely understood to be an important measure to constrain reporting bias (Nosek et al., 2018), it is of particular concern that this item is only adhered to by 21% of systematic reviews published using PRISMA between 2010 and 2017 (Page & Moher, 2017). This low adherence may be partly due to the uncertainty that surrounds the writing of systematic reviews protocols, their pre-registration, and how to transparently report and justify deviations from protocol when

necessary. For example, it is often considered unclear how immutable a pre-registered protocol is, and when and how systematic reviewers can appropriately deviate from protocol and subsequently report this transparently (DeHaven, 2017). In addition, systematic reviews tend not to report specific search results (48%), or screening and extraction procedures (abstract screening: 18%; full text screening: 20%). Furthermore, specifically in metaanalyses, systematic reviews reported the effect-size in 62% and moderator information in 58%. Finally, only 11% of systematic reviews contained the statistical code required for reproducibility of the analysis (Polanin et al., 2020). This reporting is necessary not only to give context to any additional decisions made during the analysis, but also to give others the information to evaluate key decisions made within the planned review, and improve the reproducibility of evidence synthesis (which is known to be low: Maassen et al., 2020).

Given these issues surrounding uptake, adherence, accessibility, and relevance of existing guidelines, the Non-Intervention, Open, and Reproducible Evidence Synthesis (NIROES) collaboration was set up to create a suite of accessible tools designed to facilitate evidence synthesis of non-intervention research, while also minimising the limitations of existing guidelines. In particular, it is designed to have high utility amongst novice systematic reviewers. This paper presents the Non-Intervention, Open, and Reproducible tool for systematic reviews (NIRO-SR), which is designed with the specific purpose of providing guidelines and a framework for researchers to conduct a systematic review of nonintervention research in line with best practice. We believe this to be particularly applicable to the social, cognitive, and behavioural sciences, as those are the perspectives from which the majority of co-authors have approached the problem, but the guidelines may well prove useful to a wider range of fields given the non-specificity of the items. We acknowledge the importance of conducting meta-analyses as part of systematic reviews, however it is not a strictly necessary part of a systematic review and so this is outside of the scope of the current paper. Our tools provide guidance for creating, planning, and pre-registering a systematic review protocol

(Part A), and conducting and reporting a systematic review (Part B), with the goal of making evidence synthesis as open and reproducible as possible, thereby improving the credibility of the systematic review and reducing the likelihood of biased outcomes and conclusions.

2. Method

2.1. Item bank

2.1.1. Search & Information Sources. The refinement and specification of the aims and the scope of the project (as reflected in the introduction) occurred during conferences and working groups that engaged researchers, librarians, and journal editors predominantly representing experimental and basic behavioural/cognitive fields from January to December 2019 (e.g. Advanced Methods for Reproducible Science Workshop, UK Reproducibility Network 2019; Society for the Improvement of Psychological Science Conference, 2019; NIROES Online Collaborative Hackathon, August 2019). Participants were at different career levels and with varied experience of applying systematic reviews in their work. Discussions during these meetings unveiled personal experiences of barriers for conducting systematic reviews beyond intervention research. In addition, many pre-existing tools to guide systematic reviews across experimental, behavioural and cognitive fields were shared, forming the first step for compiling relevant existing tools that would inform the development of NIRO-SR. Talks and presentations given about the project to date can be found through the project's Open Science Framework page.

The initial list of relevant systematic review guidelines was expanded by two authors (MKT and JSP) who conducted a search of existing guidelines for writing, reporting and quality assessment of systematic reviews, systematic maps, and meta-analyses. This was facilitated through extensive web searches (e.g. "systematic review checklist", "systematic review guidelines", "systematic review reporting"), resources from the EQUATOR network website and further collaborative sessions with the NIROES team until we reached saturation,

i.e. we could not find any more relevant tools using this method. Our search identified 19 guidelines (Appendix A) that provided quality assessment and protocols for systematic reviews, with a total of 517 items.

2.1.2. Item Extraction. All items and explanatory text were extracted verbatim from the 19 sourced guidelines to create an item bank. The PRISMA 2020 update (Page et al., 2020) and accompanying item bank were published after our item bank was compiled and, therefore, was not included in our item bank. We cross-referenced our own with those from PRISMA 2020 and identified 55 items from various additional guidelines that added value to the items we had already included. The final item bank contained 572 items extracted from all sources. The flowchart for this process is presented in Figure 1.

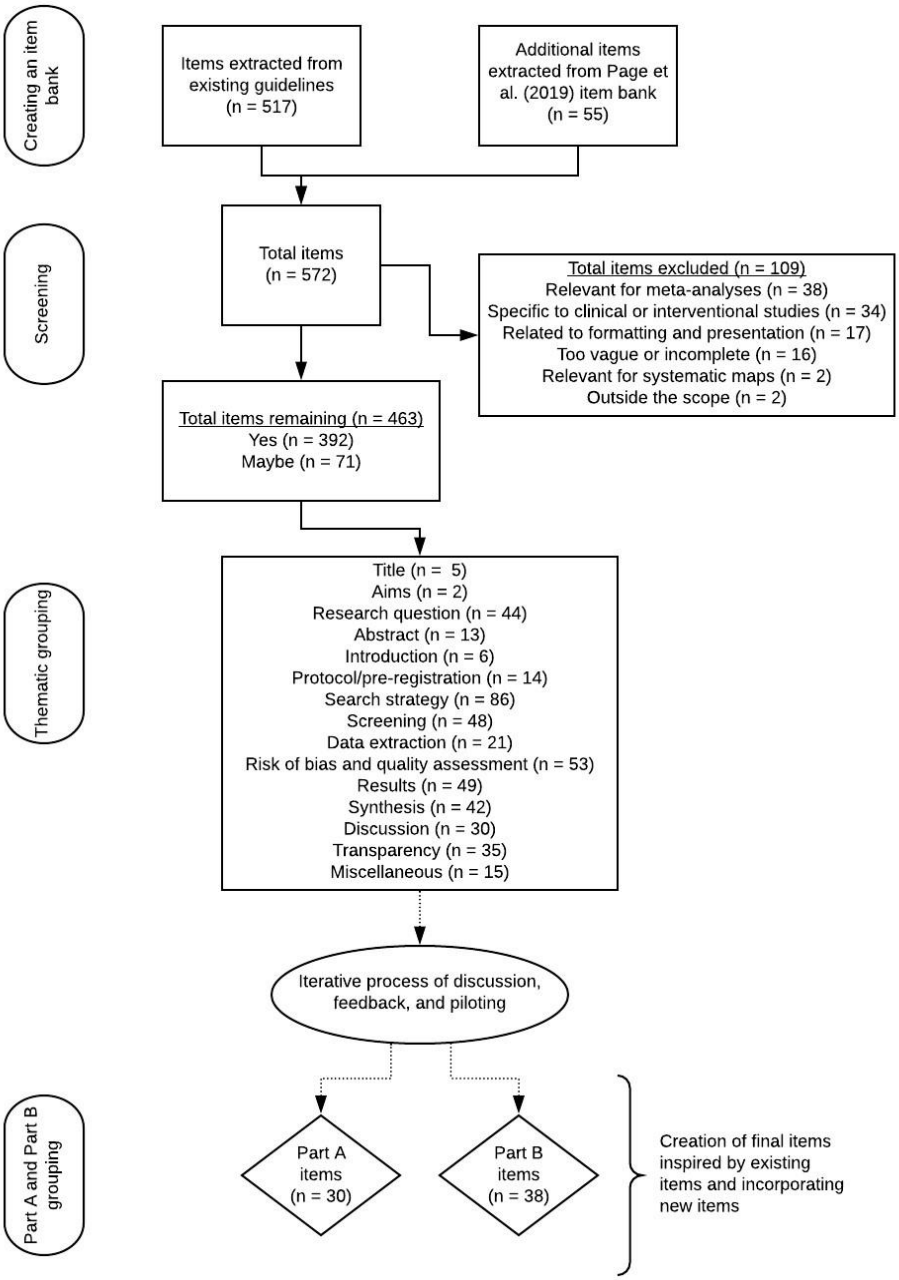

Figure 1.
Flowchart showing the records identified from searching, and the records included/excluded during screening throughout the development of NIRO-SR.

Eligibility was determined by two authors (MKT and JSP) who independently coded each item for potential inclusion as “Yes”, “No” or “Maybe” depending on its broad relevance and application for systematic reviews of non-intervention studies. “Maybe” was defined as having components that were useful but without being directly applicable as a whole item. Exclusion criteria included application (e.g., applicable to meta-analyses and/or

to systematic maps only); relevance (items that were relevant to clinical/intervention research and not adaptable for non-intervention research systematic reviews), formatting and presentation (items which suggested formatting that was not specific to systematic reviews, for instance if they referred only to systematic maps), and ambiguity (e.g., items that had a lack of clarity or incomplete guidance). Disagreements were resolved by consensus, and irreconcilable disagreements were re-evaluated at a later stage of the NIRO-SR tool development following discussions with a larger group of collaborators and experts in systematic review methodology. The final item bank, including decisions about the inclusion and exclusion of items, can be found in the project's OSF repository.

2.2. Item Development

First, eligible items were categorised into the section of a systematic review that was most applicable, which included abstract, title, protocol, introduction, aims, research question, search strategy, screening, data extraction, risk of bias and quality assessment, synthesis, results, transparency, discussion and miscellaneous items (see item bank tab "included items by category"). Second, items were further divided to form two parts of the NIRO-SR tool, the protocol (Part A) and review (Part B). Protocol items were applicable when devising and pre-registering a prospective systematic review protocol of nonintervention studies, and review items were applicable for guiding the process of conducting systematic review and writing a report for publication. Finally, each group of items was either rewritten for clarity or adapted for general use in non-intervention research. This process of rewriting items, splitting complex items, and merging similar items was conducted iteratively and collaboratively over several months and alongside other feedback methods (see section 2.3 and section 2.4). The resulting items resemble the original curated items in theme, depth, scope.

An example of an adapted item is provided in Box 2. Please note that items addressing risk of bias and heterogeneity of reviewed studies were included in the NIRO-SR tool, but to a limited extent. This is because a separate, complementary tool for guiding the assessment of bias and quality in non-intervention research systematic reviews is under development by NIROES.

Box 2.

The table below presents an item which guides authors on how to prepare and report systematic review research questions. On the left, the PICOS framework sourced from the PRISMA statement. On the right, the same framework is adapted for non-intervention research in NIRO-SR. In the adapted version, the language clearly guides the researchers to state their dependent and independent variables. "Interventions" are excluded from the item and there is an added optional position on the consideration of covariates.

PICOS, PRISMA statement; Moher et al., (2009)	**Item 3, NIRO-SR (Part A)**
Provide an explicit statement of questions being addressed with reference to: • participants, • interventions, • comparisons, • outcomes, • study design	What is the primary review question? The review question must be clearly defined and include the following: • The primary outcome measure(s) of interest (the dependent variables(s); DV) • The primary independent variables (IVs) of interest • The population/participants of interest (e.g., undergraduate students, participants with a specific diagnosis, school-age children etc.) • (optional) Study design(s) of interest, for example: i. observational - measured variables at one time-point ii. cross-sectional - measured variables with different individuals at different ages/timepoints iii. longitudinal - same individuals followed over time; could be prospective or retrospective iv. experimental - examining effect of specific manipulation • (optional) Any covariates of interest or variables you want to control for (e.g. participant(s) age) *NB. If you find that your research question does not fit the above, for instance in exploratory or methodological systematic reviews, you should state this in the protocol for transparency. If you cannot operationalise the DV and IV make sure to clearly define the focus (e.g. methodological variation) and the context (e.g. in working memory research) of your investigation.*

2.3. Initial Feedback; accessibility and understandability

One aim of NIRO-SR was to make it accessible to researchers who had never conducted a systematic review before. In December 2019, feedback on the initial version of the tool was sought from a convenience sample of students and staff (N = 9) in the School of Psychology, University of Surrey (all materials and feedback available on the project's OSF page). None of the participants had published a systematic review at the time of response, and they had little previous experience with conducting systematic reviews, relatively low confidence in this method, and their research areas were non-interventional. Participants were asked to provide general ratings of NIRO-SR using a three-point scale ("1 - Not good enough", "2 - Could be improved" and "3 - Good") across five separate categories: clarity (mean rating = 2.56, SD = 0.53), structure (mean rating = 2.89, SD = 0.33), practicality (mean rating = 2.61, SD = 0.49), relevance (mean rating = 2.86, SD = 0.38), and simplicity (mean rating = 2.44, 0.52). Comments were overall positive about the tool's utility, with suggested revisions limited to improvements in clarity and further guidance in a minority of items. All participants reported that they would want to use this tool when conducting relevant systematic reviews in the future. The feedback guided some initial changes to improve the tool's clarity for non-expert users, which included adding an explanation of the purpose and procedures of pre-registration at the beginning of the tool, and explaining items in further detail. The study procedures involving human participants have been reviewed against the guidelines set out by the Ethics Committee of Faculty of Health and Medical Sciences, University of Surrey and carried out in accordance with the University of Surrey's Code of Conduct on Good Research Practice and the Declaration of Helsinki.

2.4. Final Edits; collaborator feedback

In March 2020, a virtual hackathon was hosted to invite final feedback on the tool from a multidisciplinary team of both existing and new collaborators comprising expert

researchers and librarians experienced in systematic reviews, systematic maps, and metaanalyses as well as more novice researchers with little experience of evidence synthesis. Expert researchers revised the tool to ensure that it covered the breadth of knowledge needed to conduct a systematic review, including adding details that were missing based on their own experiences of preparing pre-registration protocols and writing non-intervention systematic reviews. Novice contributors refined the tool with the aim of making it as accessible and understandable as possible to users of all levels of expertise in reporting and conducting systematic reviews. In the cases where new items were applicable to only certain types of non-intervention studies, they were marked as optional.

Finally, it was identified that certain items could benefit from additional illustrative examples, templates, or detailed guidance. These included:

- A full example of a search strategy
- A decision log template to track the decisions made during the screening and data extraction stages
- An example of a screening manual
- A template for data extraction forms
- A risk-of-bias assessment tool to help with the assessment of credibility of included non-intervention studies.

These are outside of the scope of the current paper, but represent the need for further information and guidance.

Following this feedback process, NIRO-SR Version 0.1 (and version 0.1.1 for subsequent minor fixes) was uploaded to the OSF for any researcher who wanted to use it to guide their systematic review projects. The NIRO-SR tool has already been used by several projects to inform pre-registration and the guidelines have been implemented in some curriculums, including the University of Coventry and the University of the Philippines Diliman. Feedback from users has been very positive, and they provided further suggestions

to improve the tool and increase clarity. These changes were implemented, and the current paper presents the finalised NIRO-SR Version 1.1 (see Appendix B).

3. Results

3.1 NIRO-SR Version 1.0

NIRO-SR comprises two parts, A and B. Part A is a guide for pre-registering a systematic review protocol composed of 30 items, of which 26 items are required and 4 items are recommended for best practice. The items are divided into eight sections: Title, Description and Aims, Research Question, Search Strategy, Screening, Data Extraction, Critical Appraisal, Synthesis, and Transparency. Part B is a 38-item guide for high standards of reporting for non-intervention systematic reviews with the following sections: Title, Abstract, Introduction, Method (Deviations from protocol, Search Strategy, Screening Methods, Data Extraction Method, Critical Appraisal Method, Synthesis Method), Results (Extracted Records Results, Critical Appraisal Results, Synthesis Results), Discussion, and Transparency. If Part A cannot be completed, researchers must give a justification why this is the case and are advised to include as much relevant content from Part A as possible in the final systematic review publication.

4. Discussion

NIRO-SR aims to firstly provide guidelines for conducting systematic reviews of research that do not clearly fit the definition of intervention research, such as explanatory, experimental, and basic research. The guidelines are intended to be particularly applicable to the behavioural sciences and related fields, but may also be used in other fields outside the expertise of the authors of this paper. Secondly, NIRO-SR aims to place emphasis on reproducibility, openness and transparency of systematic reviews. Part A provides guidance for developing and pre-registering a comprehensive review protocol, and Part B guides

authors in writing and reporting systematic reviews. Both parts of the guidelines are designed to be usable on their own, but can also complement existing tools such as PRISMA 2020.

NIRO-SR may particularly benefit psychologists and experimental and behavioural scientists who focus on non-intervention research in their systematic reviews, by providing specific advice on how to develop a review protocol, and to conduct and report a rigorous systematic review. It provides guidance to authors on defining primary review questions (item A3), secondary research questions (item A4), hypotheses (item A5), inclusion and exclusion criteria (item A13), and data extraction processes (items A15 to A17). These are the areas where existing systematic review guidelines are often inapplicable for nonintervention research. NIRO-SR provides a framework that places particular emphasis on the operationalisation of variables of interest (e.g. IVs and DVs) and covariates, whilst still maintaining focus on relevant study designs and participant groups (see Box 2 and item A3).

It is hoped that by providing specific advice for conducting comprehensive systematic reviews of basic research in the behavioural sciences, NIRO-SR will help to begin to standardise and improve the contents of non-intervention systematic reviews protocols.

NIRO-SR may help prevent author bias (which is usually unintentional) through its emphasis on the development and pre-registration of a protocol before conducting a systematic review. NIRO-SR does not make the distinction as to whether the protocol should be publicly available from the outset or upon publication of the review (for example, by preregistering with an embargo period on the Open Science Framework), but it places importance on the availability and transparency of the public record. NIRO-SR advises that the protocol should be available together with the final review and include a statement of transparency which specifies the date of pre-registration and point in the review process at which the protocol was pre-registered (e.g. before the final search was completed, or before data extraction began; see item A26). The protocol benefits the authors as it sets out a detailed and transparent plan for the systematic review, and benefits the reader who can more

confidently reflect on how different decisions made throughout the process of conducting the systematic review may have influenced its outcomes. NIRO-SR also emphasises the importance of reporting all deviations from the original protocol. We acknowledge that such deviations are often necessary, so we recommend that they are justified and transparently declared in the eventual report of the systematic review (item B5).

NIRO-SR recommends a multiple-author approach when conducting systematic reviews, in line with best practice recommendations (Watts & Li, 2019; Page et al., 2021). For example, multiple team members should *independently* screen the titles and abstracts and full texts, and have a clear procedure for solving potential disagreements between systematic reviewers, as well as report a quantitative measure of inter-reviewer reliability (items B13, B14, B18, B20 and B21). This helps facilitate *reproducibility* by increasing the likelihood that a separate team of researchers could follow the exact steps of the original review and reach the same conclusions (i.e. same data, same method, same results; Barba, 2018). Researchers should be able to use the same method (i.e. search strategy, screening process and inclusion/exclusion criteria), on the same data (i.e. the databases and search results) and arrive at the same results (i.e. the final set of papers and the extracted data).

However, subjective decisions must still be made through this process and so, where full reproducibility is not possible, NIRO-SR emphasises the importance of *transparency.* We recommend that a decision log is made available that catalogues important decisions (items B14, B15 and B36). The decision log allows anyone trying to reproduce the results to identify and evaluate the subjective decisions behind any discrepancies.

NIRO-SR was developed to both alleviate the barriers preventing researchers from conducting systematic reviews and to encourage novice researchers to conduct systematic review in fields where specific guidelines are currently lacking. We strived to ensure that NIRO-SR is comprehensive, clear, and openly accessible to enable researchers to improve their literature review methodology with a systematic and transparent approach.

4.1 Methodological Limitations

NIRO-SR was developed without a pre-registered protocol or previously published methodological guidance for the development of such tools, which could introduce biases at the item selection stage of the tool development. Unfortunately, the lack of pre-registration was due to the fact that—as far as we are aware—there was no pre-registration template that could serve as an adequate template for developing NIRO-SR. Our web searches to identify appropriate guidelines and tools were therefore not systematic. We minimised biases with the breadth of the collaboration team and, although the sample of nine researchers providing initial feedback was small, we additionally sought input from multiple, independent contributors comprising a cross-discipline mix of academics and librarians with extensive experience of conducting and teaching intervention or non-intervention research systematic reviews. Further, we chose to develop NIRO-SR based on existing, peer-reviewed, consensus-based guidelines of robust methods for rigorous and transparent reporting (see Appendix A).

As with all guidelines, some limitations may only be fully known when NIRO-SR has been widely adopted. Furthermore, whilst the NIROES collaboration represents multiple disciplines and research fields, the dominant field of the authors is the experimental and behavioural sciences, which may reduce its applicability to some fields. Whilst we believe the tool to be particularly applicable to explanatory experimental and basic behavioural/cognitive research, we cannot confidently assess its use for other fields. This paper accompanies the release of NIRO-SR Version 1.1, and we anticipate that further updates will be necessary and may affect the structure, content, and wording of the items. To retain standardisation, these are anticipated to be infrequent. To facilitate future updates, users of NIRO-SR are encouraged to provide feedback to the corresponding authors.

4.4 Implications and Future Directions

NIRO-SR addresses an important gap in the available guidelines to help reviewers produce high quality systematic reviews for research in experimental and behavioural sciences. The project was conceptualised through a collaborative effort during multiple method and metascience oriented meetings including the Society for the Improvement of Psychological Science 2019 conference, Advanced Methods for Reproducible Science 2019 and 2020 workshops and ReproducibiliTea meetings. The growing demand for the tool is also reflected through many presentations about NIRO-SR delivered at psychology-focused or interdisciplinary meetings and conferences including The Organisation for Human Brain Mapping 2020 conference, Metascience 2021 conference and UK Reproducibility Network's meeting for Open Science Working Groups in 2020. A number of pre-registered protocols have already been completed using NIRO-SR, some of which can be found on the NIRO-SR website. Therefore, we expect a further increase in use of the NIRO-SR tool, which we hope will have a significant impact on the quality of systematic reviews in non-intervention research, reducing the need for bespoke customisations of existing guidelines in order to answer specific research questions. A few years after release, we plan to assess the implementation of the NIRO-SR guidelines to further understand the challenges of conducting systematic reviews in our field, as well as to inform future updates. Specifically, we would like to provide an evidence-base for whether there is a demand for the tool as we have anecdotally observed already, and whether reviews using NIRO-SR are of comparable or greater quality to the high quality systematic reviews that have used other pre-existing tools.

The further standardisation of systematic reviews outside of intervention research will also allow for better meta-scientific approaches and comparison of outcomes across multiple systematic reviews in the future. Further, NIRO-SR provides a solid basis for conducting systematic review with a meta-analysis component. Whilst NIRO-SR does not advise on the

methodology specific to meta-analyses, it will help to raise the standard of the systematic approach such as the establishment of the research question, pre-registration, search strategy, inclusion/exclusion criteria, and logging decision making.

Finally, NIRO-SR is tailored for systematic reviews of experimental, cognitive and behavioural research, but future additions to the project could include "plug-ins" for the tool that enhance its existing features (released as needed on the OSF page). For example, additional optional items could assist with reviews of other study designs such as qualitative studies or longitudinal studies, or specific items could be created for other approaches to evidence synthesis such as meta-analyses or systematic maps. Additionally, extensions of the NIRO-SR are currently under development, including further guidance for risk of bias and quality assessment (related, but not necessarily synonymous, endeavours). There are elements of a study that may not directly introduce bias but which are nevertheless important indicators of quality, for example incompleteness in the reporting of the methodology which can lead to problems with replicability.

5. Conclusions

NIRO-SR is a new tool that will allow researchers to follow standardised guidelines for systematic reviews of basic cognitive and behavioural research. It fills an important gap in methodological standards and we hope it will contribute to the improvement of the quality of systematic reviews of research that does not form an intervention.

6. Authorship Contributions

The authorship for this project was determined using the CRediT Taxonomy [38] and the authorship agreement for the NIROES collaboration (available on the OSF). For the current project, authors were divided into four relevant tiers as specified in the authorship agreement. The first tier, "Project Management" specifies the joint lead authors and project co-leads,

M.K. Topor and J.S. Pickering. Within each subsequent tier, authors were listed in an alphabetical order as follows: Tier 2 “Major Contributions” (data curation, formal analysis, investigation, methodology, visualisation, writing - original draft, miscellaneous input into creating the tool): A. Barbosa Mendes, D.V.M. Bishop, F.C. Büttner, M.M. Elsherif, T.R. Evans, E.L. Henderson, T. Kalandadze, F.T. Nitschke, J.P.C. Staaks, O. van den Akker, S.K. Yeung, M. Zaneva; Tier 3 “Feedback and Review” (conceptualisation, writing - review & editing, feedback on the tool): A. Lam, C.R. Madan, D. Moreau, A. O’Mahony, A.J. Parker, A. Riegelman, M. Testerman; and Tier 5 “Senior Supervision” (in addition to Tier 2 “Major Contributions”): S.J. Westwood.

7. Acknowledgements

Special thanks to the attendees of the Society for the Improvement of Psychological Science meeting in Rotterdam, Netherlands, July 2019, as well as attendees of the Advanced Methods for Reproducible Science workshop in Windsor, UK, January 2020. We are also very grateful to all researchers at the University of Surrey for providing their feedback. Additional thanks to: Katie Corker, Margriet Groen, Matt Jaquiery, Sayaka Kidby, Emily Kothe, Marta Majewska, Marissa McBride, and James Montilla Doble.

8. Supplemental Material

All data and supplementary materials have been deposited in an open repository on the Open Science Framework. Relevant links have been provided throughout the paper for access to specific materials. All of these materials are hosted on the open NIRO-SR OSF page (Topor et al., 2021).

9. Conflicts of Interest

No funding has been received for the realisation of this project. Jade Pickering was on the advisory board at Meta-Psychology at the point of submission. No other authors declare any conflicts of interest.

10. References

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3. http://doi.org/10.1037/amp0000191 https://doi.org/10.1186/s13104-022-05944-1

Barba, L. A. (2018). Terminologies for Reproducible Research. *ArXiv*. https://arxiv.org/abs/1802.03311

Booth, A., Clarke, M., Dooley, G., Ghersi, D., Moher, D., Petticrew, M., & Stewart, L. (2012). The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Systematic Reviews*, *1*, 2. https://doi.org/10.1186/2046-4053-1-2 Booth, A., Noyes, J., Flemming, K., Moore, G., Tunçalp, Ö., & Shakibazadeh, E. (2019). Formulating questions to explore complex interventions within qualitative evidence synthesis. *BMJ global health*, *4*(Suppl 1), e001107. http://dx.doi.org/10.1136/bmjgh-2018-001107

Bramer, W. M. (2015). Patient, intervention, comparison, outcome (PICO): an overrated tool [expert searching]. *MLA News*. 55(2).

Campbell Collaboration. (2021). *Campbell Collaboration.* Accessed March 3, 2020. https://www.campbellcollaboration.org/

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ (Clinical Research Ed.)*, *368*, l6890. https://doi.org/10.1136/bmj.l6890

Coeytaux, R. R., McDuffie, J., Goode, A., Cassel, S., Porter, W. D., Sharma, P., Meleth, S., Minnella, H., Nagi, A., & John W Williams, J. (2014). *Criteria Used in Quality Assessment of Systematic Reviews*. Department of Veterans Affairs (US). Accessed December 3, 2020. https://www.ncbi.nlm.nih.gov/books/NBK242394/

Collins, A.M., Vercammen, A.; McBride, M.F., Carling, C., Burgman, M.A. (Under Review) *Reproducibility of systematic reviews in environmental and conservation science.* PLoS One.

Committee of Medical Journal Editors. (2021) *Clinical Trials.* Accessed March 3, 2020. http://www.icmje.org/recommendations/browse/publishing-and-editorialissues/clinical-trial-registration.html

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*, *22*(10), 1435–1443. https://doi.org/10.1177/1049732312452938

Critical Appraisal Skills Program. (n.d.). *CASP Systematic Review Checklist*. Accessed December 2, 2020. https://casp-uk.net/casp-tools-checklists/

DeHaven, Al. (2017, May 23). *Preregistration: A Plan, Not a Prison.* Accessed March 3, 2020. https://www.cos.io/blog/preregistration-plan-not-prison

DeLucia, C. & Pitts, S.C. (2012). Intervention. In N.J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 630-632). SAGE.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, *75*(3), 301–315. https://doi.org/10.1037/amp0000494

Gates, N. J., & March, E. G. (2016). A Neuropsychologist's Guide To Undertaking a Systematic Review for Publication: Making the most of PRISMA Guidelines. *Neuropsychology Review*, *26*(2), 109–120. https://doi.org/10.1007/s11065-016-9318-0

Glass, G. V. (1972). The wisdom of scientific inquiry on education. *Journal of Research in Science Teaching*, *9*(1), 1–18. https://doi.org/10.1002/tea.3660090103

Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES RepOrting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*, *7*(1), 7. https://doi.org/10.1186/s13750-018-0121-7

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. Wiley. https://doi.org/10.1002/9781119536604

Hohn, R. E., Slaney, K. L., & Tafreshi, D. (2020). An Empirical Review of Research and Reporting Practices in Psychological Meta-Analyses. *Review of General Psychology*, 108926802091884. https://doi.org/10.1177/1089268020918844

Hou, Y., Benner, A. D., Kim, S. Y., Chen, S., Spitz, S., Shi, Y., & Beretvas, T. (2020). Discordance in parents' and adolescents' reports of parenting: A meta-analysis and qualitative review. *American Psychologist*, *75*(3), 329–348. https://doi.org/10.1037/amp0000463

Ioannidis, J. P. A. (2016). The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, *94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210

Joanna Briggs Institute. (n.d.). *Critical Appraisal Tools.* Accessed December 2, 2020. https://joannabriggs.org/critical-appraisal-tools

Macpherson, A., & Jones, O. (2010). Editorial: Strategies for the Development of International Journal of Management Reviews. *Wiley*. https://doi.org/10.1111/j.1468-2370.2010.00282.x

Maassen, E., van Assen, M. A., Nuijten, M. B., Olsson-Collentine, A., & Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PloS one*, *15*(5), e0233107. https://doi.org/10.1371/journal.pone.0233107

Miller, J. (2002). The scottish intercollegiate guidelines network (SIGN). *The British Journal of Diabetes & Vascular Disease*, *2*(1), 47–49. https://doi.org/10.1177/14746514020020010401

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1. https://doi.org/10.1186/2046-4053-4-1

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

National Heart, Lung, and Blood Institute. (n.d.). *Study Quality Assessment Tools.* Accessed December 2, 2020. https://www.nhlbi.nih.gov/health-topics/study-quality-assessmenttools

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Oxman, A. D., & Guyatt, G. H. (1991). Validation of an index of the quality of review articles. *Journal of Clinical Epidemiology*, *44*(11), 1271–1278. https://doi.org/10.1016/0895-4356(91)90160-b

Page, M. J. (2020). *Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement.* Accessed December 2, 2020. https://doi.org/10.17605/OSF.IO/P93GE

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic reviews*, *10*(1), 1-11. https://doi.org/10.1186/s13643-021-01626-4

Page, M. J., & Moher, D. (2017). Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. *Systematic Reviews*, *6*(1), 263. https://doi.org/10.1186/s13643-017-0663-8

Panic, N., Leoncini, E., de Belvis, G., Ricciardi, W., & Boccia, S. (2013). Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and metaanalyses. *Plos One*, *8*(12), e83138. https://doi.org/10.1371/journal.pone.0083138

Parsons, S., Azevedo, F., Elsherif, M.M., Guay, S., Shahimet, O.N., Govaart, G…Aczel, B. (2022). A community-sourced glossary of open scholarship terms. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01269-4

Polanin, J. R., Hennessy, E. A., & Tsuji, S. (2020). Transparency and reproducibility of metaanalyses in psychology: A meta-review. *Perspectives on Psychological Science*, *15*(4), 1026-1041. https://doi.org/10.1177%2F1745691620906416

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J. B., & PRISMA-S Group. (2021). PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews*, *10*(1), 39. https://doi.org/10.1186/s13643-020-01542-z

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ (Clinical Research Ed.)*, *358*, j4008. https://doi.org/10.1136/bmj.j4008

Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, MetaAnalyses, and Meta-Syntheses. *Annual Review of Psychology*, *70*, 747–770.

https://doi.org/10.1146/annurev-psych-010418-102803
Stevens, A., Shamseer, L., Weinstein, E., Yazdi, F., Turner, L., Thielman, J., Altman, D. G., Hirst, A., Hoey, J., Palepu, A., Schulz, K. F., & Moher, D. (2014). Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. *BMJ (Clinical Research Ed.), 348*, g3804. https://doi.org/10.1136/bmj.g3804
Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T. A., & Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *The Journal of the American Medical Association*, *283*(15), 2008–2012. https://doi.org/10.1001/jama.283.15.2008
Thiese, M. S. (2014). Observational and interventional study design types; an overview. *Biochemia Medica*, *24*(2), 199–210. https://doi.org/10.11613/BM.2014.022 The
Cochrane Collaboration. (2021) *Cochrane.* Accessed March 3, 2020. https://www.cochrane.org/
The Methods Group of the Campbell Collaboration. (2019a). *Methodological Expectations of Campbell Collaboration Intervention Reviews: Conduct Standards*. Campbell Collaboration. https://doi.org/10.4073/cpg.2016.3
The Methods Group of the Campbell Collaboration. (2019b). *Methodological Expectations of Campbell Collaboration Intervention Reviews: Reporting Standards*. Campbell Collaboration. doi:10.4073/cpg.2016.4
Topor, M., Pickering, J. S., Barbosa Mendes, A., Bishop, D. V. M., Büttner, F. C., Elsherif, M. M., Evans, T. R., Henderson, E. L., Kalandadze, T., Nitschke, F., Staaks, J., van den Akker, O., Yeung, S. K., Zaneva, M., Lam, A., Madan, C., Moreau, D., O'Mahony, A., Parker, A. J., … Westwood, S. J. (2021). *NIRO-SR: Non-Intervention, Reproducible and Open Systematic Reviews.* https://osf.io/f3brw/ https://doi.org/10.17605/OSF.IO/F3BRW
UK EQUATOR Centre. (n.d). *Enhancing the QUAlity and Transparency Of health Research.* Accessed March 3, 2020. https://www.equator-network.org/
University of York. (n.d.). *Centre for Reviews and Dissemination.* Accessed March 3, 2020. https://www.york.ac.uk/crd/
Voorberg, W. H., Bekkers, V. J. J. M., & Tummers, L. G. (2015). A Systematic Review of Co-Creation and Co-Production: Embarking on the social innovation journey. *Public Management Review*, *17*(9), 1333–1357. https://doi.org/10.1080/14719037.2014.930505
Watts, R. D., & Li, I. W. (2019). Use of checklists in reviews of health economic evaluations, 2010 to 2018. *Value in Health*, *22*(3), 377–382. https://doi.org/10.1016/j.jval.2018.10.006
Whiting, P., Savović, J., Higgins, J. P. T., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J., Churchill, R., & ROBIS group. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, *69*, 225–234. https://doi.org/10.1016/j.jclinepi.2015.06.005

Appendix A

The list of guidelines used to extract items for curation and preparation of NIRO-SR.

517 items have been extracted verbatim from the guidelines below:

- AMSTAR systematic review quality checklist (Shea et al., 2017)
- CASP Checklist for Systematic Reviews (Critical Appraisal Skills Program, n.d.)
- Criteria Used in Quality Assessment of Systematic Reviews (Coeytaux et al., 2014)
- Joanna Briggs Institute Checklist for Systematic Reviews (Joanna Briggs Institute, n.d.)
- MECCIR: Conduct standards (The Methods Group of the Campbell Collaboration, 2019a)
- MECCIR: Reporting standards (The Methods Group of the Campbell Collaboration, 2019b)
- MOOSE: Reporting guidelines for Meta-analysis of Observational Studies in Epidemiology (Stroup et al., 2000)
- National Heart Lung and Blood Institute Checklist for Systematic Reviews (National Heart, Lung, and Blood Institute, n.d.)
- Overview Quality Assessment Questionnaire (Oxman & Guyatt, 1991)
- PRISMA Protocols (Moher et al., 2015)
- PRISMA Statement (Moher et al., 2009)
- PRISMA 2020 update item bank (Page, 2020)
- PROSPERO (Booth et al., 2012)
- Reproducibility of systematic reviews in environmental and conservation science (Collins et al., Under Review)
- ROBIS: Tool to assess risk of bias in systematic reviews (1.2) (Whiting et al., 2016)
- ROSES (Haddaway et al., 2018)
- SIGN Tool based on AMSTAR (Miller, 2002)
- SPIDER - alternative to PICO for qual and mixed research (Cooke et al., 2012)

- Synthesis without meta-analysis (SWiM) in systematic reviews (Campbell et al., 2020)

An additional 55 relevant items were extracted from close inspection of the PRISMA 2020 update itembank (https://osf.io/kbj6v/, Page et al., 2020) which included a number of additional tools and guidelines used across different fields.

Appendix B *(Reproduced here under a CC-BY-NC-SA license)*

Non-Intervention, Reproducible, and Open Systematic Reviews (NIRO-SR) [Version 1.1]

This document provides a tool for the preparation of a systematic review protocol intended for preregistration (Part A). Further instructions facilitate the completion of reporting the systematic review (Part B).

This tool is designed specifically for systematic reviews of non-intervention research.

It does not aim to cover the details of performing a meta-analysis as part of the systematic review as this component is comprehensively detailed elsewhere in the literature. This framework can still be used for systematic reviews that do incorporate a meta-analysis.

Table of Contents

GLOSSARY OF TERMS USED IN THE TOOL:

Protocol: A very detailed, step-by-step plan/proposal for your systematic review covering conceptual and methodological aspects. It should guide the process of conducting your systematic review stepby-step.
Records: These are research reports, articles, and any other other materials that you choose to review in your systematic review.
Report: A detailed documentation of the conducted systematic review and its outcomes. May be intended for publication or as dissertation or a thesis chapter.
Reviewers: Researchers who are involved in the reviewing stages of the systematic review e.g. screening, data extraction, and quality assessment.
Other Terminology: Throughout the tool, terminology sourced from open and reproducible research methodology is being used. For definitions and further explanation of terms, please refer to the FORRT Glossary (Parsons et al., 2022; https://forrt.org/glossary/).

Part A
Preparing the Protocol for Pre-Registration

PRE-REGISTRATION

Why should you pre-register your systematic review protocol?
Pre-registration is becoming the new standard practice among many disciplines. When registered, your protocol is time-stamped and it can be kept private until the publication of your finalised manuscript. Pre-registering your protocol constrains biases and questionable research practices (such as selective reporting) that can undermine robust research synthesis. When the protocol is made public, it enhances the discoverability of your work and helps others to evaluate the quality of your review. You may also consider submitting your protocol to a journal in the format of a registered report so that editors may decide whether to publish your review before the outcomes are known.

When should you pre-register your protocol?
You must pre-register your protocol before conducting the final search and extracting the records. Initial scoping searches which may inform your final search strategy can be conducted before preregistration. Pre-registration ensures that your protocol is date-stamped.

Where can you pre-register your protocol?
It is your decision to choose the platform to pre-register your protocol. Some commonly used opensource platforms include:

- Prospective Register of Systematic Reviews (PROSPERO) - a platform designed for preregistering systematic reviews.
- Open Science Framework (OSF) - a platform which hosts research materials including preregistrations, pre-prints, data, and supplementary materials.

You can embargo your pre-registration so that it remains private until you have completed your systematic review.

What if you need to make changes after the protocol has been pre-registered?

Any changes that need to be made after pre-registration (for example, you decided to change the data management software) must be reported and justified in the Transparency section of your review. If you realise that the initial protocol needs significant changes you should consider updating your preregistration. All versions of your pre-registration should be linked and refer back to the original preregistration.

SYSTEMATIC REVIEW PROTOCOL

All items with an asterisk ***** *are required, and the remaining items are recommended for good research practice.*

Title

A1. * Provide a working title for your study.
The title must include key information that is informative to the reader. It does not have to be the same as the title of the final published report.

Description and Aims

A2. * Provide a brief description of your review topic, including background, purpose and rationale of the review, and overarching research questions. Your exact research question(s) should be given in the next section.

In writing your description, consider why the review is needed and how it contributes to knowledge in the wider research area of interest. Reasons can include:

- The literature requires synthesis and no previous or ongoing systematic reviews exist. Specify how this was checked (e.g. scoping search, consultation with researchers in the field, etc.).
- An update to previous systematic reviews is needed (e.g. significant time since the last relevant review, a significant new body of work is available, theoretical reasons to examine new measures related to the topic, etc.)
- Previous review(s) was/were flawed (e.g. no quality assessment, bias in exclusion/inclusion criteria, etc.)

Research Question

A3. * What is the primary review question? The review question must be clearly defined and include the following:
● The primary outcome measure(s) of interest (the dependent variable(s); DV) ● The primary independent variables (IVs) of interest

- The population/participants of interest (e.g., undergraduate students, participants with a specific diagnosis, school-age children etc.)
- (optional) Study design(s) of interest, for example:
    i. observational - measured variables at one time-point
    ii. cross-sectional - measured variables with different individuals at different ages/timepoints iii. longitudinal - same individuals followed over time; could be prospective or retrospective
    iv. experimental - examining effect of specific manipulation
- (optional) Any covariates of interest or variables you want to control for (e.g. participant age)

NB. If you find that your research question does not fit the above, for instance in exploratory or methodological systematic reviews, you should state this in the protocol for transparency. If you cannot operationalise the DV and IV make sure to clearly define the focus (e.g. methodological variation) and the context (e.g. in working memory research) of your investigation.

A4. Clearly define secondary review questions, if any.
Secondary questions can supplement the primary review question by investigating the effect of interest in more detail (e.g., by considering other relevant methodological aspects, outcome measures, or any additional variables. These questions can be more exploratory (i.e. the researcher might not know what effects can be expected), but they must be defined in as much detail as possible.

A5. * Clearly state any hypothesis/hypotheses.
If applicable, specify the expected direction of the effects for one-tailed hypotheses. If you have no hypotheses and you are planning an exploratory review, explain what your expectations are or state that you have no expectations with regards to the findings.

Search Strategy

NB. You can pre-register your protocol before or after running your search. Please respond to each item in a way that is most appropriate for the stages that you have planned or completed.

A6. * Describe the key components of the systematic search strategy. Examples of components that are often used: population, context, outcomes, (in)dependent variables, study type, etc.
- Describe the process by which the relevancy of the different components for the search strategy were assessed (for example, using a reference set of benchmark titles) Supply the names of the databases that will be used for the systematic search strategy

- If text mining tools were used to identify search terms or assess the relevancy of separate components, specify the tool(s) used

A7. * Supply the full search query for the systematic search strategy in a link or as an attachment

- For database searches, specify for each database its name (e.g. PsycINFO, MEDLINE, ERIC) and the interface or platform through which the database was searched (e.g. Ovid, EBSCOhost, ProQuest)
- Use database-appropriate syntax. This typically includes parentheses, Boolean operators (e.g. AND, OR), field codes (e.g. author, title, abstract, subject headings) and wildcard symbols (e.g. *, #, ?)

A8. * Additional search strategies. If additional search methods were used to identify relevant records, please specify the following;

- **Citation tracking:** If cited or citing reference searches (also called backward and forward citation tracking) were conducted, specify the citation database used (e.g. Google Scholar, Web of Science Core Collection or Scopus), and the date the citation searching was done
- **Expert consultation:** If individuals were contacted to identify records, specify the number and types of individuals contacted (e.g. authors of records included in the review or researchers with expertise in the area)
- **Hand searching:** If tables of content from relevant sources were checked, specify the names of the sources (journal names, conference proceeding) and the volumes (or publication period) checked
- **Other methods:** if other additional searches were conducted in resources not suitable for the systematic search (google scholar, preprint servers, search engines, websites, repositories), list the names of the resource(s), date (range) searched, and method used (e.g. scoping, relevancy ranking)

A9. * Assessing the scope of the search. Were the search strategies used to search for records adequate with regards to your research question? Some considerations:

- With regards to your searches, did you aim for completeness (finding *all* records) or retrieving a representative set of records?
- Were extra efforts made to improve the quality of the search strategy? For example; including a librarian, and/or peer review of the search strategy
- If a date limit was used for the systematic search strategy, was this appropriate and/or justified?
- If the results were limited to only English language records, was this appropriate and/or justified?
- If *extra* effort was made with regards to one of the following aspects, provide where relevant, more information on how this was done in the description of your search strategies:
    i. Unpublished research. For example, contacting experts and authors from included studies to find unpublished manuscripts, searching resources with conference proceedings, research reports, research data, protocols or preprints

ii. Multidisciplinarity of your topic. For example, selection of databases that reflects relevant disciplines, include discipline-specific jargon in search terms, searching book series iii. Studies from other countries, especially lower and middle income countries. For example, include databases or institutional university repositories that explicitly cover research from relevant regions

iv. Studies published in languages other than English. For example, using databases in other languages (e.g. LILACS, CNKI) for the systematic search strategy, or including languages other than English for scoping in Google Scholar

v. Studies published by professionals in the field (e.g. teachers, nurses, NGO's). For example, expert consultation, searching relevant websites with research reports or hand searching professional journal websites **A10. * Do you plan to update your search?**

Describe any plans to update the search. It is good practice to update the search before submitting for peer review, but it can be a good idea to plan more updates in earlier phases, especially if the project is expected to take several years.

Screening

A11. * What software/applications are you planning to use to store and manage the data throughout the review process?

For example, exporting .bib/.ris files from search platforms into reference manager software such as Zotero, Endnote, or Mendeley, applications such as Covidence or Rayyan, or a spreadsheet or other manual method. Indicate whether the same platform will be used to screen for duplicates and to deliver your final set of records to your co-reviewers. This may change through the process of conducting the systematic review and you can clarify any deviations from the original plans in the final report.

A12. * Clearly define the inclusion/exclusion criteria of the records. Use the following for guidance:

- The main effect(s) of interest (dependent variable(s)) described in a standardised form (e.g. percentage of correct responses to reflect task performance, overall score on a self-perception questionnaire etc.) and whether this is within-groups, betweengroups, or both
- The independent variable(s) described in a standardised form (e.g. task difficulty reflected by different trial types, participant anxiety levels reflected by a total sum of questionnaire scores)
- Detailed description of participant groups of interest to match the aims (elderly adults over 80 years old in the general population) including details of any other relevant participant characteristics especially in the case of clinical groups (e.g. diagnostic information)
- Study design (for examples, see item A3)

- Method of data collection (e.g. what experimental paradigm or questionnaires were used, any specific requirements to ensure that the method is comparable across the included records)

Explain how these decisions help to address your research question in an objective way. It is important to consider what factors may impact the outcomes of the review and how/whether you should control for these.

A13. * **Describe the screening manual for any co-reviewers.**

These should include detailed instructions on the screening process to facilitate the standardisation of this process across all co-reviewers. You should also consider piloting the screening manual if possible with a blind reviewer who can give you feedback about their experience with its implementation. At a minimum, the manual should include all relevant inclusion and exclusion criteria for the different stages of the screening process e.g. participant demographics, methods, article types etc. The manual should be attached to the pre-registration document. It is possible that the screening manual may change once the systematic review begins, due to the iterative nature of the process. In this case, transparency in the final screening manual is fundamental, and reviewers should state any changes with clear justifications.

A14. * **How many reviewers will screen the records? How will any disagreement between co-reviewers' screening decisions be resolved?**

You should specify the number of reviewers that will screen the records in at least two stages including screening titles and abstracts as the first stage, and screening the full texts as the second stage. Include the following:

- Whether at stage one the titles and abstracts will be reviewed simultaneously or sequentially.
- Whether and how you will 'blind' reviewers' decisions
- Whether all titles, abstracts, and full texts will be screened by the same reviewers.
- Your instructions for resolving disagreement between the co-reviewers i.e. when two (or more) co-reviewers judge differently whether a record should be included. This may include resolution by discussion or resolution by a third (or further) reviewer.

You should also keep a decision log where reviewer screening decisions are documented and explained. For best practice, you could prepare a template for this and add it to the preregistration as a supplement.

N.B. In order to avoid bias, more than one reviewer should contribute to this process and there should be enough overlap between the items that each reviewer screens to calculate consistency in decision-making.

Data Extraction

A15. * **Specify the data that you plan to extract from the records, which may include:**

- Participant groups/sample size and characteristics (e.g. adult, children, clinical) ● Study design
- Methods and tools used for obtaining the variables of interest (e.g. if you are reviewing anxiety, report the type of questionnaires or physiological readings within the reported study)
- The IV, DVs, and any covariates
- Results, for example means, standard deviations, effect sizes (or relevant data that can be used to calculate effect sizes), statistical tests used, p-values, Bayes factors
- Any additional data relevant for a meta-analysis, if needed ● Any specific limitations or other notes about the study

Provide justification for the proposed data extraction in relation to your research question.

A16. **Create data extraction forms that specify what information should be extracted from the obtained records.**

You can add it as a supplementary file or an appendix within the pre-registration.

A17. **Explain your strategy for dealing with missing data during the extraction process.**

For instance, you may want to extract effect sizes but you notice that they have not been reported in some records. You could try to compute the effect size with other information provided. If this is not possible, you could contact the authors and ask for the missing information to complete your data. You should also specify what you will do if you are unable to obtain complete data (e.g. no response from the authors).

A18. * **How many reviewers will extract data?**

Describe how reliability of the decisions will be assessed if only one reviewer will be involved. You may consider keeping a decision log for this process.

N.B. In order to increase reliability, more than one reviewer should contribute to this process and there should be enough overlap between the data that each reviewer extracts to calculate consistency in decision-making.
Alternatively, if you are planning to use a data mining software, describe in detail how you will extract this information, whether you have piloted this process, and how reliable it is (e.g. sensitivity and specificity ratings for your search criteria).

Critical Appraisal

A19. * State how you will assess the risk of bias and/or methodological quality within the included records.

You should aim to assess the methodological quality of the records included in the systematic review, as well as the reporting quality. It is best practice to use an existing, validated tool, although currently there are few tools appropriate for reviews of nonintervention research. Provide all the details about the tool used and why this tool was selected. If there is no validated tool available that is appropriate to assess the records included in your review, clearly explain any alternative methods of quality assessment (e.g., power analysis) or if you intend to adapt an existing tool. If you intend to adapt an existing tool, clearly state and justify what changes you made to the tool and why these changes were made. You should also state whether any validation of the adapted tool will be undertaken.

A20. * How many reviewers will assess the risk of bias and/or methodological quality of each record?

To avoid bias, more than one reviewer should complete the critical appraisal. Each reviewer should do so independently, and discrepancies should be resolved by consensus or by an additional reviewer. If only one reviewer will conduct the critical appraisal, describe how you intend to avoid bias and ensure reliability. You may want to keep a decision log for this process.

A21. * State how you are planning to assess possible publication bias which may have an effect on the outcomes of the review.

For example, if you are planning a qualitative synthesis you should state an intent to discuss the likely impact of publication bias on your conclusions. If you are planning a meta analysis, you should state which method you will choose to correct the estimated effect sizes for publication bias. If such assessment is not possible, you should explain why.

Synthesis

A22. * How will the results be synthesised?

Will you only use qualitative synthesis? Is there a planned meta-analysis? If you are combining methods (e.g. qualitative synthesis and meta-analysis), describe in detail how each approach will address the aims of the study. If you will synthesise evidence from records based on a given characteristic (e.g. studies in low-income vs high-income settings), provide specific detail on how records will be grouped and whether there is a minimal threshold for the number of records included.

A23. * How will you assess the methodological heterogeneity of the records included in the review?

You should consider the level of variability in the methodologies from the individual

studies included in your systematic review. For instance, different cognitive task software, task settings or mode of on-task responses could fundamentally alter the assumptions of the underlying mechanisms (e.g. a reaction time task with manual hand movements compared to keyboard responses). These differences could impact the interpretation of individual studies and could make it difficult to compare the results between studies.

A24. ***** **How are you planning to weigh the evidence of the included records based on their risk of bias/quality assessment rating?**
Will you only include records with high methodological quality? Will you compare the results from records with high and low methodological quality? You should at least plan to discuss this when presenting your results in the final manuscript.

A25. ***** **How many reviewers will synthesise the results?**
Describe how the reliability of the decisions will be assessed if only one reviewer will be involved.

N.B. In order to avoid bias more than one reviewer should contribute to this process.

Transparency

A26. ***** **Provide a clear statement clarifying the stage of your systematic review at the time of pre-registration (include all that apply):**

- None of the below (provide an explanation)
- Scoping searches completed
- Final database search completed and results exported/download into review management software/spreadsheet
- Grey literature search completed
- Reference lists and/or forward citation search completed
- Screening of titles and abstracts begun
- Any other later stage with justification for why the pre-registration was not completed earlier

A27. ***** **Have you made any changes or have you updated the protocol since the first pre-registration?**
Include details, justifications, and dates of any changes.

A28. ***** **Declare if you or any of the review co-authors are an author of one of the records that will likely be included in the review (based on your search strategy).**
Describe the process for protecting against bias when reviewing articles authored by a member of the review team at all stages of the review process.

A29. List all additional documents that will be attached or submitted together with the pre-registration.

These will include your screening manuals, data extraction forms and decision logs and any other supplementary materials.

A30. * Pre-register the protocol on a designated online platform with a link protected from expiration or deactivation.

The pre-registered protocol can be uploaded under an embargo until you are ready to share it, such as during the peer-review process, or at the point of publication. It is good practice to make your protocol public and accessible at the point of publication (if not before).

Part B

Reporting the Review

Writing your review

In your systematic review report, you can use the same wording as in your pre-registration protocol where appropriate.

If you have not previously followed the accompanying pre-registration protocol guidelines, or did not pre-register your study, include all relevant information as much as possible from the items in Section A into the final review. Where there is no clear existing protocol, emphasise transparency.

Supplementary materials

Supplementary materials should include all resources used to conduct your systematic review such as decision logs and data extraction forms. They should be submitted with your finalised manuscript to a journal if appropriate. You can also place them in an online repository and provide the links in your systematic review report for easy access.

Pre-printing your review

We highly recommend that, upon the completion of the systematic review, it should be made available as a preprint prior to publication. This can be on any of the appropriate arXiV preprint servers, or a public repository such as the OSF. You can use the Sherpa Romeo website to check if your eventual target journal allows submitted manuscripts to be uploaded as preprints.

Making your review available as a preprint allows for early dissemination of your work which is particularly important as reviews can get outdated in the (often lengthy) process of formal peer-review at a journal. It also acts as a fail-safe to ensure that your review does not fall victim to the so-called "file drawer".

This section will guide the write-up of the systematic review following the completion of a preregistered protocol.

In this section, all items are required.

Title

B1. The title must include key information that is clear to the reader.

The title should be a specific and informative description of the planned systematic review. As a minimum, the title must identify the project as a systematic review (and/or metaanalysis, when applicable) to enhance discoverability. The title should include key information about the review (e.g., effect of interest, independent variables of interest, outcomes, and study design).

Abstract

B2. Prepare a concise abstract for your review. The following details should be added at the least:

- An introduction to the reviewed topic
- The aims and objectives of the review
- The research question
- The method of data synthesis (qualitative synthesis, meta-analysis or both)
- The number of included records
- A summary and an explanation of the main result (in context of critical appraisal, risk of bias, and publication bias findings where appropriate) ● Implications of the review

NB: this section can be prepared last upon completion of your report.

Introduction

B3. Include the following items, which should have been specified in your pre-registration protocol already:

- A description of your review topic, including background *(see item A2)*
- Why the review is needed *(see item A2)*
- Aims, including primary and secondary (if relevant) research questions *(see items A3 and A4)*
- Hypotheses (if applicable) including the direction of expected effects if one-tailed *(see item A5)*

Method

Deviations from protocol

B4. State whether a pre-registered protocol exists for the review.

If a pre-registered protocol exists:

- Advise how to access the protocol. Provide a link or the name of the database and any registration numbers.

If a pre-registered protocol does not exist:

- Explain the reason why and reflect on how the absence of a pre-registered protocol might affect the quality of this review.

B5. Include one of the following transparency statements:

- A "Deviations from the pre-registered protocol" section. Deviations from protocol are allowed, if they are justifiable. Here you should include full details of any deviations and a full justification for any changes, including aspects of the protocol that were not implemented. If lengthy explanation is required, expand this in the supplementary materials.
- A declaration that there were no deviations from the final pre-registered protocol.

Search strategy

B6. Provide an overview of your search strategy.

Include the following details:

- Justification of the search components selected *(see item A6)* ●

Specify all text mining tools that were used

B7. Supply the full search query.

Include all search queries for all databases used alongside database names and interfaces or platforms used. This can be done as supplementary material, appendix or a link to a repository where this information can be found (*see item A7).*

B8. State all additional search strategies used.

These could include citation tracking, expert consultation, hand searching and other methods (*see item A8*).

B9. Describe the scope of the search.

Were the search strategies used to search for records adequate with regards to your research question (*see item A9)*?

B10. State the date(s) that the searches were performed including the dates of search updates if applicable.

Please be as precise as possible, particularly if the search dates differed across databases and sources *(see item A10)***.**

Screening Method

B11. State the software, applications, and/or methods that were used for storing and managing the data throughout the review process *(see item A11)***.**

B12. State the inclusion/exclusion criteria that you used to screen the search results *(see item A13)***.**

- Main effect(s) of interest (dependent variable(s))
- Independent variable(s)
- Participant groups
- Study design
- Method of data collection

B13. State the number of reviewers engaged in the screening process and report the level of agreement (e.g. reliability metrics, consistency) between the reviewers *(see item A14).* You should do this for both**:**

- Titles and abstracts
- Full texts

B14. How did you resolve any discrepancies between reviewers where appropriate? *(see item A14)*

You should describe the general strategies of resolving discrepancies. In addition, within your supplementary materials, you should provide a decision log with all reviewer decisions, notes, and outcomes of discussions for each of the results from the initial search.

B15. Provide a full list or a table of records that were excluded during the screening process together with reasons and explanations for the exclusion (this could be included as supplementary materials).

You should specify how many records were excluded when screening titles and abstracts, how many were excluded when screening the full text and at any other point during the process if applicable. If you've already provided a decision log in the item above, this does not need to be as in-depth.

Data Extraction Method

B16. Specify the data that was extracted from the included records *(see item A15)*.

This may include:

- Participant groups/sample size and characteristics (e.g. adult, children, clinical) ● Study design
- Methods and tools used for obtaining the variables of interest
- The IVs, DVs, and any covariates
- Results, for example means, standard deviations, effect sizes (or relevant data that can be used to calculate effect sizes), statistical tests used, p-values, Bayes factors
- Any additional data relevant for a meta-analysis, if needed
- Any specific limitations or other notes about the study

B17. Provide the details of any missing data and how it was dealt with *(see item A17)***.**
For example, did you attempt to contact the authors of the records to gain more information? Did you exclude the records with incomplete information?

B18. State the number of reviewers engaged in the data extraction process and report the level of agreement between the reviewers *(see item A18)*.
This should be included in a decision log along with the method by which consensus was achieved in the event of disagreement.

Critical Appraisal Method

B19. Provide full details of the methods of the critical appraisal *(see item A21)***.** You should be transparent about the tools you used to guide your critical appraisal and, if it was necessary to adapt an existing tool, describe the adaptations and why they were needed. You should also explain whether your approach for critical appraisal was validated e.g. piloted with blinded reviewers.

B20. State the number of reviewers engaged in critical appraisal and report the level of agreement (e.g. reliability metrics, consistency) between the reviewers *(see item A20)***.**
This should be included in a decision log along with the method by which consensus was achieved in the event of disagreement

Synthesis Method

B21. State the number of reviewers engaged in the process of data synthesis *(see item A25)***.**

B22. Describe in detail how the results were synthesised to address your research question(s) (*see item A22*)**.**

Results

Extracted Records Results

B23. Report the number of records found in the search. You should specify the number of records:

- found using each search method e.g. database searches, grey literature searchers, contacting authors
- removed as duplicates
- removed after titles and abstracts were screened
- removed after the full texts were screened
- removed for other reasons with clear justifications (e.g. retractions, unable to access the full text)
- included in the final synthesis (qualitative and/or quantitative)

N.B. You should present this in the form of a flowchart. We recommend the PRISMA 2020 flowchart template. Alternatively, you can describe this in text.

B24. Present the data extracted from each record in a table.

This should include data for each individual record, rather than summaries of groups of records. The table should include the items that were identified for extraction in the preregistration protocol. It should also include, where applicable, an overall score/rating based on critical appraisal.

Critical Appraisal Results

B25. Provide full details of the risk of bias and quality assessment results for each study included in the review.

This could be represented in a table with the ratings of each item and included as supplementary materials. Summary ratings or overall scores can be included in your overall data extraction table (*see item B24*).

B26. Report the results of your publication bias assessment.

If the assessment was qualitative this can be addressed in B28, but if publication bias was assessed statistically you should report the results in full.

Synthesis Results

B27. Describe the methodological heterogeneity of the records included in the review (e.g. how comparable are the methods used?) *(see item A23)***.**

B28. How have the methodological heterogeneity, publication bias, and risk of bias/quality assessment outcomes affected the synthesis of the results?

Refer to your pre-registration protocol if needed to emphasise whether the results were weighed based on the methodological quality (as determined by critical appraisal).

B29. Provide a descriptive summary of findings for all planned analyses and effects of interest.

These summaries should be aligned with the research questions and any relevant hypotheses. Make sure to clearly present the results in relation to the primary and secondary research questions.

B30. You should present any exploratory analyses separately.

Exploratory analyses are those conducted in addition to the planned analysis. These normally emerge during the process of analysing data to answer your pre-planned research question(s). This may happen if you observed something interesting in your data that you had not taken into consideration before. You must make it clear that these exploratory observations are separate from the pre-planned analyses.

Discussion

B31. Summarise the main conclusions of your review with regards to the main research question(s).

B32. Within this summary consider the overall direction of the effects of interest in the context of the quality of the included records and the strength of the body of evidence.

Present a balanced view of the obtained data. It is important that your conclusions do not overemphasise what is evident in your data.

B33. Present conclusions from exploratory analyses separately to confirmatory analyses and explain how these may impact the field and the future direction of research.

B34. Discuss the problems, gaps and limitations within the reviewed literature. Provide future research recommendations.

B35. Discuss any strengths and limitations regarding the process of conducting this systematic review which may have an effect on the overall outcome (e.g. incomplete detection of relevant record).

Transparency

B36. Keep a record of your data extraction and decision processes throughout the different stages of the review and upload the records as open data and/or supplementary materials for transparency.

For example, for each record in the original search, log whether they were included/excluded at each stage by each reviewer (removed as a duplicate, removed at the screening of titles and abstracts, or screening of full text), how any discrepancies were

resolved, and any additional information that contributed to each decision where appropriate. You should also upload the associated data for each record from both the data extraction and critical appraisal (if appropriate) stages.

B37. State the name and version number of all software and any packages used, if applicable.

For example, `revtools v0.4.0` with `RStudio v3.6`.

Analysis scripts should be made available (e.g. on the Open Science Framework or Github) alongside the data where possible and should include a persistent identifier, such as a Digital Object Identifier (DOI). The scripts should be reproducible, meaning that they should include enough instructions for other researchers to computationally reproduce the outcomes.

B38. Acknowledge the contribution of each author. You could do this using the CRediT taxonomy.

B39. Declare your funding and any non-financial support you have received for this systematic review project as well as conflicts of interest, such as whether you authored any of the records that were included in the review or whether you receive any benefit as a result of performing the review.

Make a clear statement where there are no conflicts of interest.