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

Marta K Topor*[1,2], Jade S Pickering*[3], Ana Barbosa Mendes[4], Dorothy Bishop[5], Fionn Büttner[6], Mahmoud M Elsherif[7], Thomas Rhys Evans[8], Emma L Henderson[2], Tamara Kalandadze[9], Faye T Nitschke[10], Janneke P C Staaks[11], Olmo R van den Akker[12], Siu Kit Yeung[13,14], Mirela Zaneva[5], Alison Lam[15], Christopher R Madan[16], David Moreau[17], Aoife O'Mahony[18], Adam J Parker[19], Amy Riegelman[20], Meghan Testerman[21], and Samuel Westwood[22,23]

[1]University of Copenhagen, Denmark
[2]University of Surrey, UK
[3]University of Manchester, UK
[4]Erasmus University Rotterdam, Netherlands
[5]University of Oxford, UK
[6]University College Dublin, Ireland
[7]University of Birmingham, UK
[8]University of Greenwich, UK
[9]Østfold University College, Norway
[10]University of Newcastle, Australia
[11]University of Amsterdam, Netherlands
[12]Tilburg University, Netherlands
[13]The University of Hong Kong, HKSAR, China
[14]The Chinese University of Hong Kong, HKSAR, China
[15]University of Liverpool, UK
[16]University of Nottingham, UK
[17]University of Auckland, NZ
[18]Cardiff University, UK
[19]University College London, UK
[20]University of Minnesota, MN, USA
[21]Princeton University, NJ, USA
[22]University of Westminster, UK
[23]King's College London, UK

*Joint first authors

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 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.

Keywords: guidelines, non-intervention research, open research, reproducibility, systematic reviews, transparency

Introduction

Systematic literature reviews are a widely used method for rigorously synthesising existing evidence to answer research questions and to inform best practice and policy-making. 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 pre-registration 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., 2019); 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.

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 and 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 sys-

Table 1

Definitions of terminology

Term	Definition
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).
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.

tematic 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 and Moher, 2017) but also in 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 and 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 meta-analyses, 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 non-intervention 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.

Method

Item bank

Search and 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 (osf.io/8seby/).

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 EQUA-TOR 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.

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.

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

4

Figure 1

Flowchart showing the records identified from searching, and the records included/excluded during screening throughout the development of NIRO-SR.
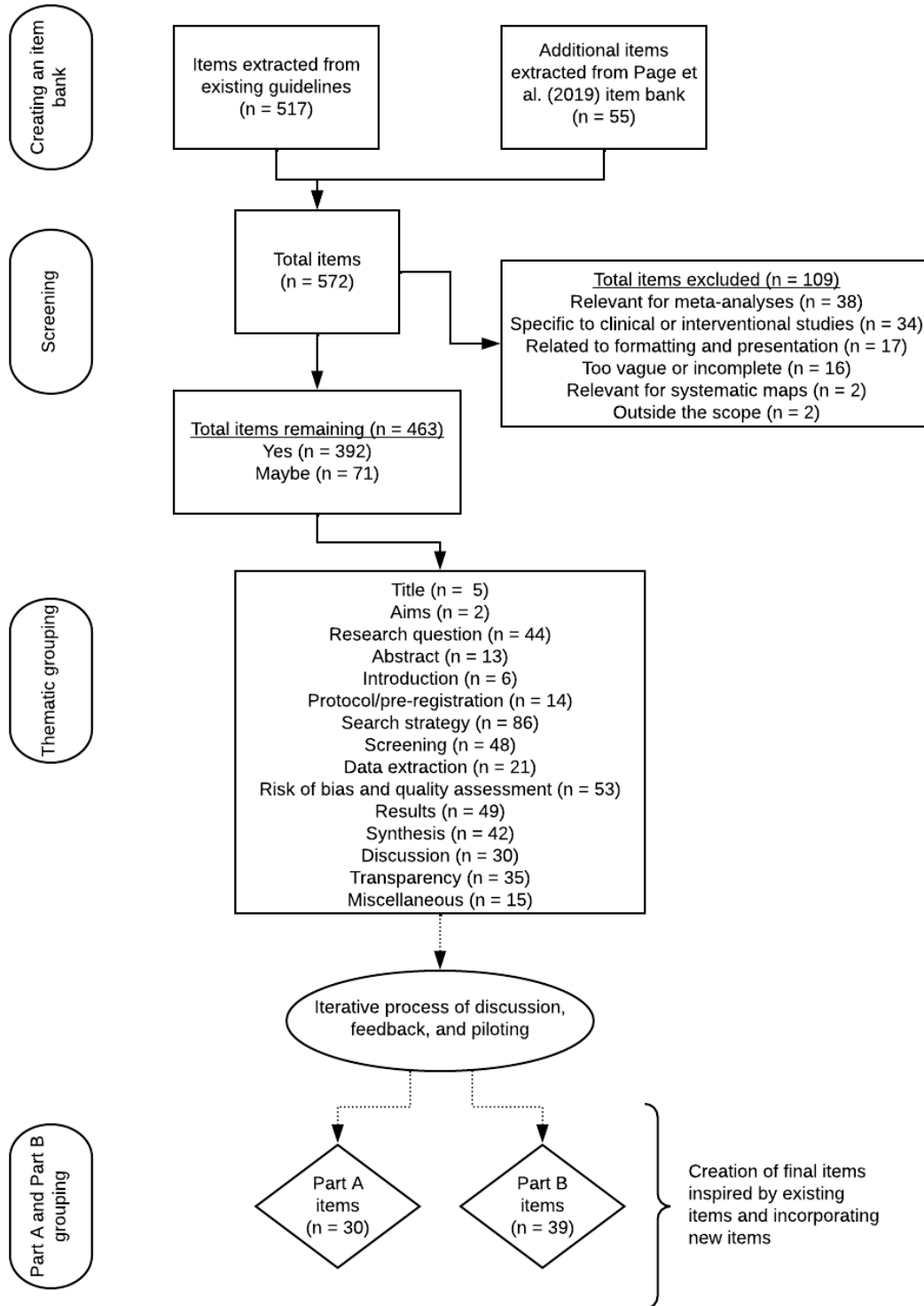
a lack of clarity or incomplete guidance). Disagreements were resolved by consensus, and irreconcilable disagreements were re-evaluated at a later stage of the NIRO-SR tool development following discussions with a larger group of collaborators and experts in systematic review methodology. The final item bank, including decisions about the inclusion and exclusion of items, can be found in the project's OSF repository (osf.io/p2v34).

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 the 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 a 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 Table 2. Please note that items addressing the 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.

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 osf.io/f3brw). 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.

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 meta-analyses 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

Table 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 timepoints/variables
	  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.

- 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 osf.io/c9wer 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.

Results

NIRO-SR Version 1.0

NIRO-SR comprises two parts (osf.io/c9wer), 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.

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 non-intervention research. NIRO-SR provides a framework that places particular empha-

sis 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 Table 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 pre-registering 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 (Page et al., 2021; Watts and Li, 2019). 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 throughout 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.

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.

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 OSF (osf.io/f3brw). 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 reviews 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; osf.io/f3brw). 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.

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.

Author Contributions

The authorship for this project was determined using the CRediT Taxonomy 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.

Supplemental Material

The NIRO-SR tool is available at osf.io/c9wer. 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 (osf.io/f3brw; Topor et al., 2023).

Author Contact

Corresponding author: Marta K. Topor, marta.topor@hotmail.co.uk (ORCID: https://orcid.org/0000-0003-3761-392X). MKT and JSP contributed equally and are joint first authors.

Open Science Practices

This article is purely theoretical and as such is not eligible for Open Science badges. The entire editorial process, including the open reviews, is published in the online supplement.

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. *The American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Barba, L. A. (2018). Terminologies for reproducible research. *arXiv*. Retrieved October 30, 2020, from 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. https://doi.org/10.1136/bmjgh-2018-001107

Bramer, W. M. (2015). Patient, intervention, comparison, outcome (PICO): An overrated tool. *MLA News*, *55*(2).

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* (tech. rep.). Department of Veterans Affairs (US). Retrieved March 3, 2021, from https://www.ncbi.nlm.nih.gov/books/%7BNBK242394%7D/

Collins, A., Vercammen, A., McBride, M., Carling, C., & Burgman, M. (n.d.). *Reproducibility of systematic reviews in environmental and conservation science.*

Committee of Medical Journal Editors. (2021). Clinical trials. Retrieved March 3, 2020, from http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/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. Retrieved December 2, 2020, from https://casp-uk.net/casp-tools-checklists/

DeHaven, A. (2017). Preregistration: A plan, not a prison. Retrieved March 3, 2020, from https://www.cos.io/blog/preregistration-plan-not-prison

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., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (Version 6). Cochrane. 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

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. Retrieved December 2, 2020, from https://joannabriggs.org/critical-appraisal-tools

Maassen, E., van Assen, M. A. L. M., 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

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

Methodological expectations of campbell collaboration intervention reviews: Conduct standards (tech. rep.). (2019). The Campbell Collaboration. https://doi.org/10.4073/cpg.2016.3

Methodological expectations of campbell collaboration intervention reviews: Reporting standards (tech.

rep.). (2019). The Campbell Collaboration. https://doi.org/10.4073/cpg.2016.4

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., & Group, P. (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., & Group, P.-P. (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., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021. https://doi.org/10.1038/s41562-016-0021

National Heart, Lung, and Blood Institute. (n.d.). Study quality assessment tools. Retrieved December 2, 2020, from https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

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., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89. 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

Polanin, J. R., Hennessy, E. A., & Tsuji, S. (2020). Transparency and reproducibility of meta-analyses in psychology: A meta-review. *Perspectives on Psychological Science*, *15*(4), 1026–1041. https://doi.org/10.1177/1745691620906416

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J. B., & Group, P.-S. (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, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, *70*, 747–770. https://doi.org/10.1146/annurev-psych-010418-102803

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

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. T., 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. (2023). Non-interventional, reproducible, and open systematic review (niro-sr) guidelines. Retrieved March 27, 2023, from https://osf.io/f3brw/

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., & group, R. (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

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 (*Methodological expectations of Campbell Collaboration intervention reviews: Conduct standards*, 2019)

- MECCIR: Reporting standards (*Methodological expectations of Campbell Collaboration intervention reviews: Reporting standards*, 2019)

- 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 and Guyatt, 1991)

- PRISMA Protocols (Moher et al., 2015)

- PRISMA Statement (Moher et al., 2009)

- PRISMA 2020 update item bank (Page et al., 2021)

- PROSPERO (Booth et al., 2012)

- Reproducibility of systematic reviews in environmental and conservation science (Collins et al., n.d.)

- 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 qualitative 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 item bank (https://osf.io/kbj6v/, Page et al., 2021) which included a number of additional tools and guidelines used across different fields.