STUDY PROTOCOL

Scoping review and evidence mapping of interventions aimed at improving reproducible and replicable science: Protocol [version 1; peer review: 3 approved with reservations]

Leonie A. Dudda[1,2], Magdalena Kozula[3], Tony Ross-Hellauer[4,5], Eva Kormann[4], René Spijker[6,7], Nicholas DeVito[8], Gowri Gopalakrishna[9,10], Veerle Van den Eynden[11], Patrick Onghena[3], Florian Naudet[12,13], Rita Banzi[14], Maddalena Fratelli[14], Monika Varga[15], Yuri Andrei Gelsleichter[16], Inge Stegeman[1,2], Mariska M. Leeflang[9]

[1]Department of Otorhinolaryngology, Head and Neck Surgery, University Medical Center Utrecht, Utrecht, The Netherlands
[2]University Medical Center Utrecht Brain Center, University Medical Center, Utrecht, The Netherlands
[3]Faculty of Psychology and Educational Sciences, Methodology of Educational Sciences Research Group, KU Leuven, Leuven, Flanders, Belgium
[4]Institute of Interactive Systems and Data Science, Technische Universitat Graz, Graz, Styria, Austria
[5]Know-Center GmbH, Graz, Austria
[6]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
[7]Amsterdam Public Health, Medical Library, Amsterdam UMC, Univ of Amsterdam, Amsterdam, The Netherlands
[8]Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, England, UK
[9]Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, The Netherlands
[10]Faculty of Health, Medicine & Life Sciences, Maastricht University, Maastricht, The Netherlands
[11]Research Coordination Office & RDM Competence Centre, KU Leuven, Leuven, Flanders, Belgium
[12]CHU Rennes, Inserm, Irset (Institut de recherche en santé, environnement et travail)-UMRS 1085, CIC 1414 [(Centre d'Investigation Clinique de Rennes)], Universite de Rennes, Rennes, Brittany, France
[13]Institut Universitaire de France, Paris, France
[14]Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy
[15]Institute of Animal Sciences, Hungarian University of Agriculture and Life Sciences, Kaposvar, Hungary
[16]Department of Soil Science, Institute of Environmental Sciences, Hungarian University of Agriculture and Life Sciences, Gödöllő, Hungary

v1 **First published:** 18 Oct 2023, **3**:179
https://doi.org/10.12688/openreseurope.16567.1
Latest published: 18 Oct 2023, **3**:179
https://doi.org/10.12688/openreseurope.16567.1

Abstract

Background

Many interventions, especially those linked to open science, have been proposed to combat the reproducibility crisis. To what extent these propositions are based on scientific evidence from empirical

Open Peer Review

Approval Status ? ? ?

	1	2	3
version 1 18 Oct 2023	? view	? view	? view

1. **Antica Culina**, Rudjer Boskovic Institute, Zagreb, Croatia

evaluations is not clear.

Aims

The primary objective is to identify interventions that have been formally investigated regarding their influence on reproducibility and replicability. A secondary objective is to list any facilitators or barriers reported and to identify gaps in the evidence.

Methods

We will search broadly by using electronic bibliographic databases, broad internet search, and contacting experts in the field of reproducibility, replicability, and open science. Any study investigating interventions for their influence on the reproducibility and replicability of research will be selected, including those studies additionally investigating drivers and barriers to the implementation and effectiveness of interventions. Studies will first be selected by title and abstract (if available), semi-automated, and then by reading the full text by at least two independent reviewers. We will analyze existing scientific evidence using scoping review and evidence gap mapping methodologies.

Results

The results will be presented in interactive evidence maps, summarized in a narrative synthesis, and serve as input for subsequent research.

Review registration

This protocol has been pre-registered on OSF under doi https://doi.org/10.17605/OSF.IO/D65YS

Keywords
Reproducibility, Replicability, Open Science, Transparency, Review

This article is included in the Research Ethics and Integrity collection.

2. **Rubén López-Nicolás**, Universidad de Murcia, Murcia, Spain

3. **Olavo B. Amaral**, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Leonie A. Dudda (l.a.dudda@umcutrecht.nl)

Author roles: Dudda LA: Conceptualization, Data Curation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Kozula M**: Conceptualization, Data Curation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Ross-Hellauer T**: Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Kormann E**: Conceptualization, Data Curation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Spijker R**: Methodology, Writing – Review & Editing; **DeVito N**: Writing – Review & Editing; **Gopalakrishna G**: Writing – Review & Editing; **Van den Eynden V**: Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Onghena P**: Writing – Review & Editing; **Naudet F**: Writing – Review & Editing; **Banzi R**: Writing – Review & Editing; **Fratelli M**: Writing – Review & Editing; **Varga M**: Writing – Review & Editing; **Gelsleichter YA**: Writing – Review & Editing; **Stegeman I**: Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Leeflang MM**: Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The OSIRIS project has received funding from the European Union's Horizon Europe research and innovation programme under the grant agreement No. 101094725. The TIER2 project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101094817.

How to cite this article: Dudda LA, Kozula M, Ross-Hellauer T *et al.* **Scoping review and evidence mapping of interventions aimed at improving reproducible and replicable science: Protocol [version 1; peer review: 3 approved with reservations]** Open Research Europe 2023, **3**:179 https://doi.org/10.12688/openreseurope.16567.1

First published: 18 Oct 2023, **3**:179 https://doi.org/10.12688/openreseurope.16567.1

Introduction

One of the scientific research objectives is to test theories, typically by accumulating knowledge derived from individual studies. This process is contingent on the reliability of previous findings that serve as the foundation for subsequent research. One of the essential prerequisites for trustworthy research findings is reproducibility, which entails obtaining the same results when rerunning (parts of) projects using the same design and data. Unfortunately, research findings frequently cannot be reproduced due to, for example, inadequate information provided to rerun the experiment or analysis (Alsheikh-Ali *et al.*, 2011; Stodden *et al.*, 2018). The EU report on the reproducibility of research results in EU framework programmes (European Commission, Directorate-General for Research and Innovation, 2022) shows that many researchers rarely indicate to share their data, analysis code, or materials. This is problematic not only for the reproducibility of experiments but also for their ability to be replicated in other samples. Ensuring the reproducibility and replicability of scientific results is critical to the rigor and quality of research. When they are ignored, there is a risk that inaccurate, biased, or spurious results can gain undue attention in the literature leading to substantial research waste and flawed decision-making. Strategies to increase reproducibility and replicability have primarily focused on improving research transparency through various open science practices. These aim to ensure that the research process is documented and widely accessible so that it can be checked, critiqued, re-used, and built upon in future research. Ongoing initiatives, such as the ReproducibiliTEA network, aim to facilitate the implementation of these practices in the research process. However, the extent to which these and other practices and interventions have been empirically investigated, and their actual impact on reproducibility and replicability, are unclear. We, therefore, aim to understand the evidence base for interventions for increasing the reproducibility and replicability of research and the documented barriers and facilitators in the process of creating more reproducible and replicable research using scoping review and evidence mapping methodology.

Methods

This protocol was developed prior to the search and will be pre-registered on OSF under DOI 10.17605/OSF.IO/D65YS after including the final search strings. Title and abstract screening started in June 2023 based on a finished protocol. A well-constructed search strategy is the core of the systematic review. Due to the complexity of our search objectives, we revised the search string throughout the T&A screening in order to improve its quality. The final search string will be published with this protocol. The final search string for Medline can be found in the section describing the search strategy. Potential changes to the protocol during the research process will be reported transparently. All additions to the protocol, such as data extraction sheets and data analysis plans, will be added to the OSF folder of this project under DOI 10.17605/OSF.IO/7EF5H whenever they are ready. The draft protocol has been reviewed by all authors. The protocol follows the Preferred Reporting Items for Systematic Reviews and Meta- Analyses: extension for Scoping Reviews (PRISMA-ScR) guidelines (Tricco *et al.*, 2018).

Design

Scoping review methodology will be applied since overviews regarding the evidence of the effect of interventions on reproducibility and replicability still need to be provided. We want to create an overview of the current evidence by systematically mapping available literature using the EPPI Reviewer software (EPPI Centre Software). By doing so, we aim to identify knowledge gaps and gain an overview of the characteristics of the current literature.

Objectives

Our primary objective is to evaluate which interventions, such as open science practices, have been investigated for their effectiveness in improving reproducibility and replicability in science. These interventions can be applied on various levels (researchers, institutes, funders, publishers, editors, etc.). The search strategy will focus on this objective.

The secondary objective is to investigate which drivers and barriers to the implementation and effectiveness of these interventions have been identified in these interventional studies.

Definitions

For the purpose of this review, we embrace the definitions of replicability and reproducibility from Nosek *et al.* (2022) and European Commission's scoping and final report on reproducibility in research (European Commission, Directorate-General for Research and Innovation, 2020 and European Commission, Directorate-General for Research and Innovation, 2022). Following these definitions, we created a list of interventions and outcomes that will be investigated, which can be found in Figure 1 and Figure 2. By adopting this list, we acknowledge that it might not be exhaustive as improving reproducibility in science is a continuous exercise, and there are many maturing ideas tackling the challenge (Munafò *et al.*, 2017).

Eligibility criteria

We will include studies evaluating the effectiveness of an intervention in promoting the reproducibility or replicability of scientific methods and findings and those additionally testing drivers and barriers of the process. Studies will be included independent of the specific approach applied to measure the outcome variable(s), such as using proxies or direct indicators of the outcomes.

In addition to studies that explicitly evaluate interventions for reproducibility/replicability, we will also consider studies that assess the effectiveness of interventions related to practices commonly claimed within the literature to support reproducibility/reproducibility to be in the scope of this review. The complete list of such proxies is included in Figure 1 and is based upon two criteria: the project team's synthesis of key, widely cited prescriptive texts that provide a conceptual framework for the selection made (Munafò *et al.*, 2017;

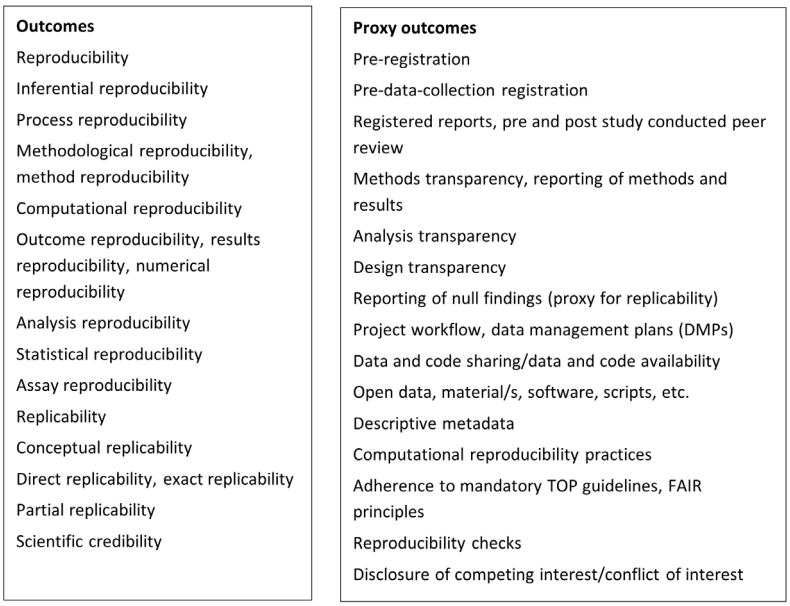

Figure 1. Outcomes.

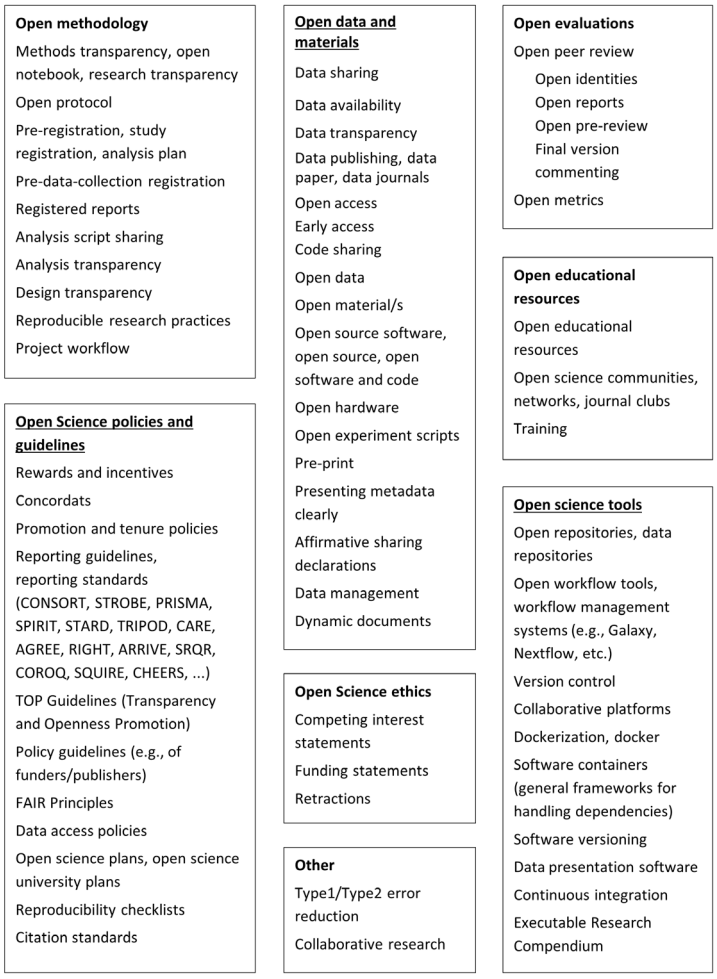

Figure 2. Interventions.

Stodden *et al.*, 2016; European Commission, Directorate-General for Research and Innovation, 2020 and European Commission, Directorate-General for Research and Innovation, 2022) and the team's evaluation which proxies would be measurable.

Various study designs will be considered for inclusion in this research. To evaluate an intervention, comparisons must be made either between or within subjects. Studies that do not have a comparator will be included only if the intervention is explicitly stated and the outcome variable measures reproducibility or replicability (for example, the prevalence of data sharing in journals with open science policies). These studies are included since they provide relevant information when compared with other studies (e.g., the prevalence of data sharing in journals without open science policy). Studies only investigating the prevalence of certain practices, such as data-sharing, are not included. Reviews will be included to apply the snowballing methodology to identify relevant literature, and since they might contain additional relevant information that is retrieved by the comparison and summary of primary studies. Reviews will be summarized separately to avoid duplication of primary studies in the evidence map. Furthermore, we include studies that investigate facilitators and barriers in addition to evaluating the effectiveness of interventions on reproducibility and replicability (e.g., as moderator). When only drivers and barriers are assessed, for example, in the form of a survey or similar describing the opinions of multiple individuals without testing any intervention, these studies will be marked as such in the full-text screening phase. Including these articles in the analysis is beyond this project's scope. However, these papers are essential for understanding researchers' practices and decisions and compiling them might serve as a starting point for further projects. Article types such as position papers, study protocols, and other literature without primary data will not be considered. Additionally, we will search for and exclude retracted articles.

Participants: Researchers, institutes, funders, publishers, editors etc. in any field of research.

Interventions: Any intervention that aims to improve the reproducibility and replicability of science. These can be open science practices, such as the pre-registration of studies, open access publishing, data and code sharing or interventions to promote these open science practices, such as journal editorial or institutional policies, and implementation of guidelines or codes of conduct.

Comparator: Any comparator, including the absence of a comparator (e.g., pre-post comparison).

Outcomes (qualitative and quantitative): Reproducibility, replicability, and their proxies.

Article types: Articles, review articles and early access papers will be included.

Study types: Qualitative and quantitative interventional studies (e.g., pre-post or experimental designs).

Search strategies

Electronic searches. At first, we will identify key papers on the topic across various domains. This will be done by direct consultation with colleagues and reproducibility and open science experts for literature recommendations to fill any remaining gaps. The key papers will be expanded by citation coupling using Connected Papers. The team will screen and discuss the new list, including an information specialist. The resulting set of documents will both inform the building of a search query and validate the proposed search. The articles will be added to a publicly available Zotero folder

Additionally, the team will create a list of interventions and outcomes of interest (see Figure 1 and Figure 2). This list will be informed by the team's knowledge of the field, colleagues' consultation, and the key papers' screening. The search terms will be defined based on the list and key papers. Before searching, the search strategy will be checked by an external information specialist.

We will systematically search Medline, Embase, Web of Science, PsycINFO, Scopus, CAB Direct, Agris, PubAg, AGRICOLA and Eric. The proposed search will initially be developed for Medline and subsequently be translated to all databases. The search string for Medline can be found at the end of this section. No date or language restriction will be used. Grey literature will be searched separately by scanning websites of major policy actors (such as EC, Science Europe, EUA, NSF) for relevant interventional studies. After identifying relevant references, they will be used for reference and citation checking to determine potentially missed studies. The results from the searches will be collected in EndNote 20. Duplicate references will be removed with an in-built deduplication function of EndNote.

Medline Search string. (((data or code or workflow or practices or materials or notebook) adj2 (open or share or shared or sharing or preservation or stewardship)) or "open science" or ((computational or data or open or research or conclusion* or inferential or analytic or conceptual or direct or exact or statistical) adj3 (reproducib* or replicability or replicable)) or (research adj5 (transparen* or credib*))).ti,ab. or (reporting adj3 guideline*).ti.

(Clinical study/ or Case control study/ or Family study/ or Longitudinal study/ or Retrospective study/ or Prospective study/ or Cohort analysis/ or Comparative Study/ or (Cohort adj (study or studies)).mp. or (Case control adj (study or studies)).tw. or (follow up adj (study or studies)).tw. or (observational adj (study or studies)).tw. or (epidemiologic$ adj (study or studies)).tw. or (cross sectional adj (study or studies)).tw. or (comparative adj stud*).mp. or ((("randomized controlled trial" or "controlled clinical trial" or "multicenter study" or "pragmatic clinical trial").pt. or non- randomized controlled trials as topic/ or interrupted time series analysis/ or controlled before-after studies/ or random*.ti,ab. or groups.ab. or (trial or multicenter or "multi center" or multicentre or "multi centre").ti. or (intervention? or effect? or impact? or controlled or control

group? or (before adj12 after) or (pre adj5 post) or ((pretest or "pre test") and (posttest or "post test")) or quasiexperiment* or quasi experiment* or pseudo-experiment* or pseudoexperiment* or evaluat* or "time series" or time-point? or "repeated measur*" or ((experimental or empirical or qualitative) adj5 (study or studies))).ti,ab.)) not ((news or comment or editorial).pt. or comment on.cm.)

Selection process

The title and abstract of all records identified through the respective search strategies will be screened for potential eligibility in duplicate independently. When there are disagreements, a third reviewer will be involved to resolve discrepancies in the selection of articles. The full-text manuscripts belonging to the titles and abstracts considered potentially eligible will then be read and selected for inclusion in duplicate independently. In this phase as well, disagreements will be solved by a third author.

Data-extraction

The data extraction sheet, including extraction instructions and terminology definitions, will be developed in cooperation with the whole team. After extracting the first ten papers, the extraction sheet will be piloted and potentially adapted. Data extraction will be conducted in duplicate by two or more reviewers independently. Disagreements will be solved through discussion and additional consultation of a third reviewer. Data will be extracted from study documents and recorded in EPPI Reviewer across the items listed below. In case of insufficient data, we will attempt to contact the document's authors for additional information. The following data will be extracted (if applicable):

a. Description of the publication
   - Title
   - Author(s)
   - Year of publication
   - Source (peer-reviewed, preprint, etc.)
   - Journal
   - Academic field

b. Type of publication
   - Systematic review, original research, research letter, etc.

c. Study design (in case the publication was a literature review, this item lists the study designs included in the review)
   - Research question
   - Comparative yes/no
   - Experimental yes/no
   - Qualitative design yes/no
   - Specific name of design

d. Description of the sample
   - Sample
   - Location/context

e. Interventions
   - Classification according to list of interventions in appendix A
   - Stringency (mandatory, optional)
   - Implemented by
   - Target population
   - Stage of target process (before, during, after study conduct, after publication)

f. Outcome variables
   - Outcome variable (reproducibility, replicability, proxy)
   - Metric
   - Instrument

g. Drivers and barriers
   - E.g., time constraints that hinder the implementation of OS practices or programming knowledge reducing mistakes in shared code

h. Results
   - Answer to research question
   - Primary statistical results (effect size, p-value, CI, etc.)
   - Descriptives (means, percentages, etc.)
   - Meta-analytic results
   - Qualitative results
   - Additional results (e.g., drivers/barriers)

Critical appraisal of individual sources of evidence

As this is a scoping review, we will not assess the risk of bias of each individual study included. We will, however, check for each article what the design was and whether a comparison was included (see data-extraction items).

Data synthesis and analysis

All included studies will be summarized in a narrative synthesis.

The evidence will be mapped according to the intervention (rows) and reproducibility/replicability (columns) using the EPPI Reviewer software (EPPI Centre Software). Additionally, an adapted version of the categorization, as described by Davidson *et al.* (2022), will be applied to cluster different kinds of interventions.

The outcomes of reviews will not be included as a row in the evidence map but described separately. Evidence mapping will produce an interactive, searchable database of relevant studies, a list of knowledge gaps, and visualizations such as

evidence atlases, heat maps, or descriptive plots. If possible, this will be done semi-automatically as well (using EPPI-reviewer, for example). These will aid in identifying gaps in the evidence and ways to address them. The exact layout of these figures will be determined after we have piloted the data extraction sheet.

Study status
We are currently at the stage of title and abstract screening.

Anticipated end of the study
The anticipated end date of the study is September 2023.

Ethics and consent
Ethical approval and consent were not required.

Data availability
No data are associated with this article.

References

Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, *et al.*: **Public availability of published research data in high-impact journals.** *PLoS One.* 2011; **6**(9): e24357.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Davidson AR, Barbour G, Nakagawa S, *et al.*: **Taxonomy of interventions at academic institutions to improve research quality.** *BioRxiv.* 2022.
Publisher Full Text

European Commission, Directorate-General for Research and Innovation, Baker L, *et al.*: **Reproducibility of scientific results in the EU: scoping report.** Lusoli, W. (editor): Publications Office of the European Union, 2020.
Publisher Full Text

European Commission, Directorate-General for Research and Innovation: **Assessing the reproducibility of research results in EU Framework Programmes for Research: final report.** Publications Office of the European Union, 2022.
Publisher Full Text

Munafò MR, Nosek BA, Bishop DV, *et al.*: **A manifesto for reproducible science.** *Nat Hum Behav.* 2017; **1**(1): 0021
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Nosek BA, Hardwicke TE, Moshontz H, *et al.*: **Replicability, robustness, and reproducibility in psychological science.** *Annu Rev Psychol.* 2022; **73**: 719–748.
PubMed Abstract | **Publisher Full Text**

Stodden V, McNutt M, Bailey DH, *et al.*: **Enhancing reproducibility for computational methods.** *Science.* 2016; **354**(6317): 1240–1241.
PubMed Abstract | **Publisher Full Text**

Stodden V, Seiler J, Ma Z: **An empirical analysis of journal policy effectiveness for computational reproducibility.** *Proc Natl Acad Sci U S A.* 2018; **115**(11): 2584–2589.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Tricco AC, Lillie E, Zarin W, *et al.*: **PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation.** *Ann Intern Med.* 2018; **169**(7): 467–473.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ? ? ?

Version 1

Reviewer Report 20 February 2024

https://doi.org/10.21956/openreseurope.17882.r36821

?

Olavo B. Amaral
Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

The paper describes the protocol of a scoping review of interventions to improve research reproducibility. While the goal of the review is extremely worthy, I found some of the details confusing (particularly with respect to the current state of the project and the use of the categories provided). I also believe the protocol would benefit from a clearer assessment of feasibility.

Major points:

1. Perhaps the most obvious issue is that, from the paper, it is not really clear at what point the project is. On the contrary, contradictory verb tenses across the manuscript make this quite confusing.
The Methods section starts out with “this protocol will be pre-registered on OSF after including the final search strings”, and then immediately gives a DOI for the preregistration. It then proceeds with “title and abstract screening started in June 2023”, even though (a) the section describing screening in the Methods is written in the future tense and (b) the text has just stated that the final search strings were yet to be developed.
These contradictions come back in later section, where the Methods state that “the team will create a list of interventions and outcome of interest” and then cites Figure 1 and Figure 2 (in which this list seems to already have been developed). And then again when they state that “the search terms will be defined based on the list and key papers” and that “the proposed search will initially be developed for Medline” and then follows up with the complete MedLine string.
I have no problem with part of the work having been performed, or with preliminary searches being carried out – such pilots tend to be useful for refining the protocol. But the authors have to be clear about what has already been performed and what is yet to be done. This is not the case in the current state of the manuscript and should be thoroughly revised.

2. Another confusing issue is the meaning of the lists in Figure 1 and Figure 2. Are the concepts listed meant to be used as inclusion criteria in the systematic review (e.g. are these the outcomes and types of interventions that qualify for inclusion)? As concepts reflected in the search strings

(although this does not seem to be the case in the provided string)? As categories for grouping interventions and outcomes in the review? I really could not understand, and the lack of legends in the figures (as well as their disconnect from the text that follows their appearance) doesn't help. My further comments on the lists are naturally limited by not understanding what the terms are intended to be used for. That said, I would note that they are quite dense in technical and/or vague terms that are not precisely defined. Some examples:
- What is meant by "process reproducibility", and how does it differ from "method reproducibility"?
- "Descriptive metadata" refers to that? Studies? Datasets?
- What fits under very vague terms such as "methods transparency", "reproducible research practices" or "training"?
If these lists are indeed important for the review process, these and other terms would benefit from a more precise definition and/or references (even if this requires an appendix or supplementary material). In fact, specific guidelines cited (e.g. TOP, FAIR principles) likely require references that have not been included anyway.
Also, if the categories are meant to be used as inclusion criteria (once again, I don't understand if this is the case), I'd question some of the items in the list. "Scientific credibility" as an outcome (unless it is defined in a very specific way) would lead to the inclusion of lots of studies on public perception of science that are not about reproducibility *per se*. And although disclosure of competing interests is important, I'm not sure I'd consider it as a proxy outcome for reproducibility – this has more to do with general research integrity and will probably lead to the inclusion of a literature that does not deal directly with reproducibility.
Finally, I note that some concepts are included both as proxy outcomes and interventions (e.g. "Methods transparency"). Although I can see why this could be the case (e.g. one could test a specific intervention to see whether it improves methods transparency, but also test whether something that clearly improved methods transparency improves reproducibility), it leads to a strange impression of circularity. It's worth thinking whether there might be a better taxonomy to use in order to avoid this. If there isn't, the authors should at least note this apparent contradiction explicitly and explain why some things can be both interventions and outcomes.

3. I'm a bit concerned with the feasibility of the review. With the snowballing approach, the very general definitions of interventions and outcomes, the large number of databases and the inclusion of grey literature, this could turn out to be a lot of work. Do the authors have any estimate of this workload, as well as the expected time to perform it, given the size of the team? In order to evaluate the protocol's feasibility, it would be useful if the authors could provide an estimate of the number of articles retrieved by the search (perhaps using their included search string), of the number of included articles (perhaps based on the inclusion rate of a pilot set of article), the number of people involved in each step and their expected deadline for concluding the review. As a reviewer, I would feel much more comfortable accepting the protocol for publication after looking at this data, in order to check whether it seems feasible and likely to be followed and become a systematic review in the future.

4. The interventions to improve reproducibility, outcomes to measure it, and in fact the very notion of "reproducibility" and its relative importance are likely to vary a lot from one scientific field to another (see Leonelli, S. (2018)[Ref-1] and others). This discussion, however, is almostly completely lacking in the protocol. The review seems set to include interventions in any field of research, but will interventions really be commensurable between them? Will any attempt be made to summarize interventions in a discipline-specific manner? This topic likely deserves some attention in the introduction and in the description of the analysis.

Minor points:

Abstract:

- Instead of “combat the reproducibility crisis”, perhaps the authors should use a less charged term such as “improve reproducibility”. There is discussion of whether the “crisis” narrative is appropriate, and it probably does not apply equally to all scientific fields. As it seems unnecessary that there is a crisis for interventions to work, it might be worth removing the term here.

- It is unclear what the term “semi-automated” refers to. Does it refer to the title and abstract screening? If so, perhaps it should be “semi-automatedly”? I also note that there is no mention of automating screening in the methods, so this should be checked.

- Can the authors be more specific on what constitutes an “interactive evidence map”.

Introduction:

- There seems to be a contradiction in the initial definitions of reproducibility. The first definition given for reproducibility is in line with the one by the US National Academy of Sciences, (e.g. obtaining the same results with the same data). The next sentence, however, mentions “rerunning experiments”, which seems to refer to what would be called replicability in this framework – unless it is talking about computational experiments, which would not be intuitive for most readers. The authors then start using “replicability” as well, but don’t define the concept. They later mention using the definition by Nosek et al. (2022)[2] and the European Commission, but only in the Methods section - if that is the case, shouldn’t these references be cited upfront in the introduction?

- Can the authors provide more examples of initiatives to integrate open practices in the research process besides ReproducibiliTea (which I would argue is only very indirectly linked to this purpose and would not be the first example that comes to my mind)?

Methods:

- Please define the abbreviation in “T&A screening” (it probably means title and abstract, but this is not specified”.

- PRISMA-ScR is a reporting guideline for scoping reviews, so I’m not sure it applies to protocols. There is a PRISMA extension for protocols Moher et al.,(2015) [Ref-3]. Shouldn’t this be used instead?

- The “design” section says little about the design. Once again, what do the authors mean by “systematically map available evidence”. It would be useful if they at least mention in which dimensions they plan to be mapping it.

- *“In addition to studies that explicitly evaluate interventions for reproducibility/replicability, we will also consider studies that assess the effectiveness of interventions related to practices commonly claimed within the literature to support reproducibility/reproducibility to be in the scope of this review”*

The distinction is not clear in this sentence: in the way this is written, isn't the second category contained in the first? The rest of the paragraph seems to imply that the distinction has to do with using actual reproducibility/replicability vs. indirect proxies (e.g. transparency) as the outcome. However, this is not what's written in this sentence, which by itself does not make sense to me.

- "*Comparisons must be made either between or within subjects*". What counts as a "subject" here? Is it a researcher? Could it be an article, an institution or a journal? Although this information seems to be included in "participants" later, what is meant by "subject" should be stated before this sentence for it to make sense.

- "*This will be done by direct consultation with colleagues and reproducibility and open science experts for literature recommendations to fill any remaining gaps."* How will these experts be defined and how many of them are expected to be consulted?

- Ten papers sound like a very small sample to develop the extraction sheet. I'd assume that the types of studies may vary a lot in this review; thus, I'd recommend that a larger pilot sample is used in order to have a better idea of what can come up and guide the extraction process.

- "*Classification according to list of interventions in appendix A".*
What is Appendix A? This seems to refer to the figures (which were an Appendix in the OSF protocol), but the paper itself mentions no appendices.

- The "results" field in the data extraction section seem to go beyond what would be analyzed in a scoping review, mentioning extraction of primary statistical results, meta-analytic results, etc. Will this be analyzed in the current review? Will it be used in later steps? Please clarify.

References

1. Leonelli S: Rethinking Reproducibility as a Criterion for Research Quality. **36**: 129-146 Publisher Full Text
2. Nosek BA, Hardwicke TE, Moshontz H, Allard A, et al.: Replicability, Robustness, and Reproducibility in Psychological Science.*Annu Rev Psychol*. 2022; **73**: 719-748 PubMed Abstract | Publisher Full Text
3. Moher D, Shamseer L, Clarke M, Ghersi D, et al.: Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement.*Syst Rev*. 2015; **4** (1): 1 PubMed Abstract | Publisher Full Text

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Metaresearch

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 20 February 2024

https://doi.org/10.21956/openreseurope.17882.r36826

?

Rubén López-Nicolás
Universidad de Murcia, Murcia, Region of Murcia, Spain

In this manuscript, the authors present a protocol for a scoping review of the existing evidence regarding interventions aimed at enhancing reproducibility and replicability. I find the topic highly compelling and pertinent; shedding light on this issue is undoubtedly necessary.

However, I have a few comments, mainly concerning the clarity and presentation of the issue at hand:
While the study focuses on reproducibility and replicability, exploring various aspects of both (as indicated in Fig. 1, Outcomes), the introduction explicitly defines only reproducibility. Additionally, although both reproducibility and replicability are later referenced in the introduction, a distinction between the two is absent in this section. Improving the introduction by providing clearer definitions for the various concepts that will be central to this work would enhance its overall clarity and understanding. While two references are indeed mentioned later in the 'Definitions' section, an earlier clarification would be very helpful.

On the other hand, I also noticed a lack of a brief discussion of the state of the art of the topic at hand in the introduction. The studies of interest seem to be papers evaluating the effectiveness of interventions to improve reproducibility and replicability. It might be useful to mention papers in this regard in the introduction, commenting on what their approach was, what intervention they evaluated, what they found, etc. It would help to put the focus on the specific topic.

The methodology, arguably the most relevant section of the protocol, is adequately described in my opinion. The iterative approach to the electronic search and the design of the data extraction sheet also seems to me to be adequate. It might be relevant to mention whether any kind of quantitative synthesis is planned if the conditions become suitable and what those conditions would be. But there may not be any intention of this being a scoping review.

Best of luck in the progress of this interesting work.

Is the rationale for, and objectives of, the study clearly described?
Yes

Is the study design appropriate for the research question?
Yes

Are sufficient details of the methods provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Research synthesis, reproducibility, meta-analysis

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 13 November 2023

https://doi.org/10.21956/openreseurope.17882.r35721

?

Antica Culina
Rudjer Boskovic Institute, Zagreb, Croatia

The manuscript is a study protocol identifying the rationale and methodology for conducting evidence mapping of interventions for increasing reproducibility and replicability. I find the study very interesting and relevant given the current need to identify whether and how open science improves credibility and impact of research. Overall, methodology is well described and developed for this stage of the research. I only have several relatively minor comments/suggestions for the MS.

My main comment is about the scope of the mapping which is somewhat not consistently described through the MS. This is relevant for two aspects:

1. Reading the Background, I've got the impression this work will look into the strength of the evidence (the last sentence of the Background). However, it is mapping the literature that contains the evidence, rather than calculating some effect sizes from the evidence. I think it would be useful that the text reflects this. E.g. last sentence of the Background could be something like ' Much of the literature has looked at the evidence that xxx but this literature has not been systematically mapped.' Similar, the sentence of the Introduction: ' However,

the extent to which these and other practices .... Etc' gives the impression that some effect sizes will be calculated from the obtained evidence.

2. The title and the aims are about any practices that can improve reproducibility and replicability. However, the Interventions that are listed in the Fig 2 are only about Open practices.

Some minor comments:

1. In the Background section of the abstract, please consider adding '/replicability' to read '.. combat the reproducibility/replicability crisis.' While I understand that reproducibility crisis is the original term that was used to describe non-replication of study results, maybe it would be easier for the reader to have both terms mentioned in the Background section.

2. I do not understand why are the listed proxy outcomes (Fig 1) in fact proxies for the main outcomes. Could you explain the reasoning behind these?

3. Fig 2: Under ' Other' it is unclear why these exact terms were chosen? Why not p-hacking reduction, or HARKING reduction?

4. The MS could better highlight that mapping the barriers and facilitators is not fully systematic, as it will be done based on the literature obtained via the main search (which is to identify practices and their potential impacts), and not by an interdependent search of the literature on facilitators/barriers.

5. Could you please provide a very brief explanation on why Risk of Bias will not be assessed?

6. Maybe change the date for the 'Anticipated end of the study'?

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Open science, open data, open codes, pre-registration, ecology, evolution

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.