RESTORATION ECOLOGY
The Journal of the Society for Ecological Restoration

SER SOCIETY FOR ECOLOGICAL RESTORATION

UN DECADE ON ECOSYSTEM RESTORATION

UNITED NATIONS DECADE ON ECOSYSTEM RESTORATION 2021-2030

REVIEW ARTICLE

Repeatability of the searching process in reviews of restoration outcomes

João Paulo Romanelli[1,2], Luiz G. M. Silva[3], Maria Carolina P. Gonçalves[4], Rafaela P. Naves[5], Danilo Roberti Alves de Almeida[6], Angélica Faria de Resende[6], Ricardo Ribeiro Rodrigues[1]

In the recently declared United Nations Decade on Ecosystem Restoration (2021–2030), evidence-based research has an essential role in summarizing restoration outcomes toward broad generalizations to advance restoration practice and science globally. However, to present trustworthy, high-quality recommendations, evidence reviews should be based on rigorous methods that minimize bias and enhance systematicity, transparency, objectivity, comprehensiveness, and repeatability. To assess the current value of restoration reviews in terms of methodological repeatability at the searching stage, a fundamental review stage, we evaluated a sample of meta-analyses and narrative syntheses ($n = 79$) and critically appraised how searches were conducted based on the information reported within each study. By assessing whether review methods were reported in sufficient detail to be repeated by an independent party, we found that repeatability varied substantially by the standard we used. Overall, reviews performed relatively poorly, with a median score of 3 out of 9 points available, and a mean score of 3.6. Most reviews ($n = 76/79$) failed in effectively reporting all necessary information to allow repeatability at this stage. We found no statistically significant differences considering review types, suggesting that the type of synthesis alone does not solely reflect the repeatability of a review. Toward increasingly repeatable, reliable reviews in restoration ecology, we recommend that authors incorporate and apply principles of systematic reviews and maps in their review process, and suggest that environmental journals should broadly emphasize ways of performing rigorous reviews, as well as increase efforts for the publication of open access review protocols.

Key words: evidence synthesis, evidence-based restoration, reliability, restoration reviews, review methodology

Implications for Practice

- Restoration reviews are lacking methodological repeatability at the searching stage. Without rigorous reports in this stage, an independent researcher cannot be sure whether the methods fit for the analysis and cannot verify or judge the authors' findings.
- The searching stage is the first step of the review process and lays the foundation for all successive steps. The lack of repeatability in this stage may lead to controversy or skewed conclusions over review findings as soon as the omitted information is added.
- A high level of objectivity and transparency underlies repeatability in reviews.

Introduction

The compilation and synthesis of data is a powerful approach to identify trends and commonalities within broader restoration science (Ladouceur & Shackelford 2020; Romanelli et al. 2020a), given that it improves our ability to forecast future outcomes to support policy, decision-making, and research. In the recently declared UN Decade of Ecosystem Restoration (2021–2030; https://www.decadeonrestoration.org/), evidence-based research represents an important tool to support broad generalizations across a range of shared and diverse conditions to advance ecological restoration practice and science globally (Brudvig 2017; Ladouceur & Shackelford 2020; Romanelli et al. 2020b).

Besides, without a credible and consistent method for evidence reviews (defined in Table 1 and hereafter also referred to as reviews), not all syntheses can be considered equally

Author contributions: JPR, LGMS, RRR conceived the ideas and designed methodology; JPR, RRR collected the data; JPR, RN, MCPG, DRAA analyzed the data; JPR, MCPG, AFR, DRAA, LGMS led the writing of the manuscript; all authors contributed critically to the drafts and gave final approval for publication.

[1]Laboratory of Ecology and Forest Restoration (LERF), Department of Forest Sciences, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Avenida Pádua Dias, 11, Piracicaba, SP 13418-900, Brazil
[2]Address correspondence to J. P. Romanelli, email joaopromanelli@hotmail.com
[3]Department of Civil, Environmental and Geomatic Engineering (D-BAUG), Stocker Lab, Institute for Environmental Engineering (IfU), ETH-Zurich, Zurich, 8046, Switzerland
[4]Enzyme Technologies Laboratory (LabEnz), Chemical Engineering Department, Federal University of São Carlos (UFSCar), São Carlos, SP 13565-905, Brazil
[5]Department of Forest Sciences, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Avenida Pádua Dias, 11, Piracicaba, SP 13418-900, Brazil
[6]Laboratory of Tropical Forestry (LASTROP), Department of Forest Sciences, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Avenida Pádua Dias, 11, Piracicaba, SP 13418-900, Brazil

doi: 10.1111/rec.13496
Supporting information at:
http://onlinelibrary.wiley.com/doi/10.1111/rec.13496/suppinfo

Table 1. Evidence synthesis and review terminology.

Term	*Definition*	*Reference*
Evidence synthesis	An element of the review process where results are generated from the combination of primary studies considering all available evidence. This "occurs once the evidence base has been accumulated and the data of interest extracted"	Pope et al. 2007
Evidence review	A comprehensive term used to describe articles that collate and summarize multiple primary studies related to a specific, policy-relevant question	CEE 2013
Reliability	"Refers to the level of confidence an end-user may place in the review methodology rather than in the accuracy of review"	O'Leary et al. 2016
Systematic review	A review that aims to answer a specific question as precisely as possible in an unbiased way, which uses systematic and explicit methods to collate, critically appraise, and synthesize all available evidence relevant to the question. Statistical methods (e.g. meta-analysis) may or may not be used to analyze and summarize the results of the studies gathered	CEE 2013; Berger-Tal et al. 2019
Systematic map	A review that does not attempt to answer a specific question as do systematic reviews, but instead gathers, characterizes, and catalogs available evidence in a certain topic or question of interest. The studies included in the synthesis can be used to identify evidence for policy-relevant questions, knowledge gaps, and knowledge clusters	James et al. 2016; Berger-Tal et al. 2019
Meta-analysis	A set of statistical methods specially developed to integrate results of two or more independent studies, on the same research question, combining the magnitude of the outcomes (effect sizes) across different data sets	Koricheva et al. 2013
Narrative synthesis	"A process which uses prose to summarize and draw conclusions from primary research and which may be supplemented by the reviewers' own experience. Some narrative syntheses may include limited quantitative analysis"	O'Leary et al. 2016

reliable (Pullin & Stewart 2006; Berger-Tal et al. 2019). Notably, much of the recent evidence review guidance published in the realm of environmental sciences and management has been focused on the characteristics that make the evidence review process trustworthy (e.g. Woodcock et al. 2014; Haddaway et al. 2015; O'Leary et al. 2016; CEE 2018). The principles of systematicity, comprehensiveness, transparency, and objectivity have been the center of attention of this literature. Systematicity refers to literature reviews conducted in an organized and orderly manner, following a clear and well-established framework (Paré et al. 2015, 2016). Comprehensiveness is achieved by searching multiple databases and establishing a search strategy sensitive enough to capture as much of the evidence pertinent to the review as possible (Bayliss & Beyer 2015; Abdulla & Krishnamurthy 2016). Transparency is achieved when the elements of the review process are explicitly detailed (Templier & Paré 2018). Objectivity refers to the idea that subjectivity is a source of bias and that it can and must be minimized (Hammersley 2001; Linde & Willich 2003). Although other characteristics of trustworthiness are noted, such as repeatability (Haddaway et al. 2015; 2017), it was not emphasized in assessments of the reliability of environmental reviews as we addressed here.

An outlook on repeatability, which refers to the methods that can be reperformed by an independent party to generate trust in the methodological process used by the authors has prominently figured in recent publications for literature review in the information science (IS) field (e.g. Boell & Cecez-Kecmanovic 2015a; Paré et al. 2016; Cram et al. 2020). In fact, this issue has been claimed to remain less recognized in IS field because of two main factors, which we believe that also apply to the environmental sector (e.g. Grames & Elphick 2020).

The first factor assumes that there are competing perspectives on the precise definition of repeatability, including inconsistencies relative to other disciplines (e.g. Cram et al. 2020). Secondly, the distinctiveness of the repeatability concept may be uncertain and even confounded with others, such as reproducibility and replicability. While the former refers to the ability of an independent party to reperform the methods (e.g. the searching strategy) or duplicate the results of an evidence review using an existing dataset (Cram et al. 2020), the latter refers to an entirely new study that seeks to corroborate or refute a previous study's results based on independent data collection and analysis (Peng 2011; Dennis & Valacich 2014). The differing perspectives presented by authors across different disciplines may introduce uncertainty into the literature review writing process. In other words, it may remain confuse what elements of a literature review process (Fig. 1) could (or should) be repeatable (Fig. 2), as well as the types of literature reviews that the concept of repeatability applies to—this was extensively discussed in Boell and Cecez-Kecmanovic (2015b), Chiasson (2015), and Leidner (2018).

In this study, we seek to clarify the concept and applicability of the principle of repeatability by echoing contemporary perspectives from the IS field (Templier & Paré 2018; Cram et al. 2020). Further, we used a set of criteria based on Grames and Elphick (2020) and Woodcock et al. (2014) for assessing the methodological repeatability of restoration reviews at the searching stage, evaluating whether an independent researcher could repeat review methods, and therefore, to assess the level of trustworthiness researchers and decision-makers (end-users) can place in these syntheses, given this is a fundamental review stage that lays the foundation for all successive steps (Johnson & Hennessy 2019). For clarity, we used evidence review terminology presented in the glossary. We assessed (1) the methodological repeatability of meta-analyses and narrative syntheses based on the information reported within each study; and (2) how

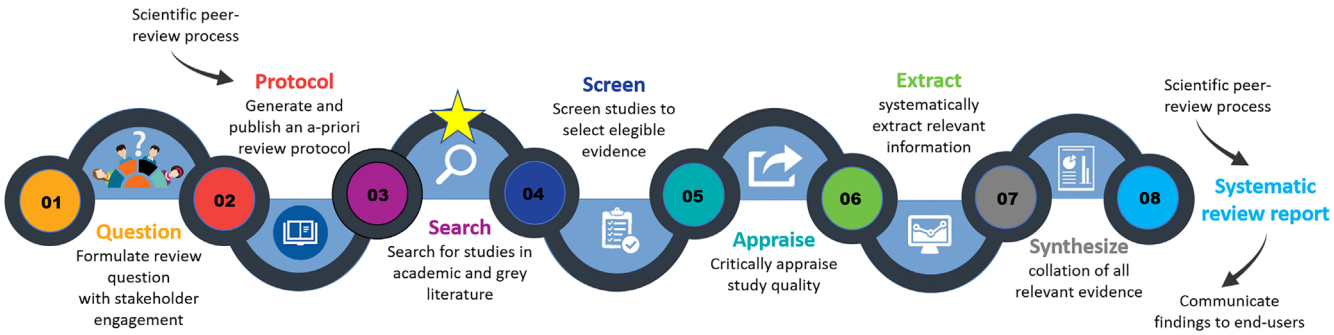

Figure 1. Flow diagram detailing the systematic review stages based on the Collaboration for Environmental Evidence (CEE 2018); highlighting (with a star) the searching stage, which was investigated in this study.

repeatability varies according to the synthesis type. We then discuss how reviews in restoration ecology could improve their methodological repeatability at the searching stage (Fig. 1) by incorporating principles of systematic reviews and maps (CEE 2018), which can further aid the field and related areas in facilitating literature reviews that are increasingly repeatable, thus reliable. We conclude with the implication of our findings for evidence-informed research.

Methods

We followed methods to gather evidence outlined in an a priori protocol (Romanelli et al. 2020b). We deviate from the protocol in that we proposed the assessment of methodological reliability of reviews in terms of objectivity, transparency, and comprehensiveness. Thus, many of the materials related to the dataset and methods used to collate evidence are similar to those presented in Romanelli et al. (2020a). Web searches were undertaken on 5 November 2019, comprising the period 2009–2019, to obtain a sample of recent reviews. Synthesis history is available in Supplement S1.

Systematic review stages	Repeatability	Reproducibility	Replicability
Protocol	✓	✗	✓
Search	✓	✗	✓
Screen	✓	✗	✓
Appraise	✓	✗	✓
Extract	✓	✗	✓
Synthesize: Analyze data	✓	✓	✓
Synthesize: Interprete and formulate conclusions	✗	✓	✓

Figure 2. Definition and applicability of the concepts repeatability, reproducibility, and replicability in each stage of systematic reviews (adapted from Cram et al. 2020). Green symbols distinguish between review stages that could benefit from being repeatable, reproducible, and/or replicable. Red symbols indicate that the concept does not apply (N/A) to the review stage.

Database Building

We assessed a sample of meta-analyses ($n = 61$) and narrative syntheses ($n = 18$) (collectively named hereafter as "reviews") focused on ecological restoration outcomes. All retrieved articles (Supplement S1) were screened according to the following inclusion criteria: (1) reviews should be a synthesis of primary research; cited databases (e.g. data gathered by other studies) were not eligible for inclusion; reviews should be described in the title, abstract or keywords as a systematic review or a meta-analysis; (2) reviews should address the ecological restoration of terrestrial ecosystems; (3) any research outcome was eligible for inclusion. Collected reviews were then classified as either meta-analysis or narrative synthesis (see Table 1) to perform the repeatability assessment based on the type of synthesis conducted in each study.

We aimed to investigate reviews in restoration ecology as a whole, identifying strengths and weaknesses in the population to offer guidelines for improving the overall quality of future evidence reviews, not to "name and shame," as the examples considered here are by no means unique. Consequently, we anonymized the scores assigned in each study. Nonetheless, we provided a list of all included and excluded studies after the full-text assessment (Supplement S1 and S2).

Repeatability Assessment

Based on CEE guidelines (CEE 2018) as well as the scoring system from the Collaboration for Environmental Evidence Synthesis Assessment Tool (CEESAT) (Woodcock et al. 2014), we investigated the searching stage of reviews (Fig. 1). We used a set of criteria (Table 2) for assessing repeatability considering the minimum details needed for an independent researcher to repeat the search, determining if reviews had reported those details. We then assigned categorical scores to each criterion, varying from inadequate standards (0), intermediate (1), or optimal (3), according to how well the criterion was met, with a maximum of 9 points possible.

Scoring decisions between scorers were subsequently analyzed considering the magnitude of disagreement between scorers with a weighted kappa test of agreement. Both JPR and MCPG scored one sample containing 20 random

Table 2. Repeatability assessment criteria for the searching stage. Using this checklist as a tool, we assessed whether restoration reviews reported their methods in sufficient detail to be repeated by an independent researcher. For each criterion, reviews could score 3 points (optimal), 1 point (intermediate), or 0 points (inadequate standard) according to how well the criterion is met. For reviews to be considered repeatable at the searching stage, all items must be satisfied. For a detailed description of the scoring system, see Woodcock et al. (2014) and Grames and Elphick (2020) for additional information on repeatability criteria. To receive 3 points (optimal), all items must be fulfilled; 1 point (intermediate), only one item related to the criterion is not fulfilled; 0 points (inadequate), two or more items related to the criterion are not fulfilled.

1. Searching for studies	An ideal search for literature should possess three fundamental properties: comprehensive (maximizes the number of potentially relevant studies), systematic (avoiding ad hoc search strategies decreases the propensity to bias resulting from, e.g. no well-defined endpoint of search), and transparent (readers should be able to replicate and appraise the search) (Woodcock et al. 2014 [CEESAT]). Searches are also expected to be repeatable and fit for purpose (Livoreil et al. 2017). Here, we evaluate whether the searching stage is described in sufficient detail to be repeated by an independent researcher, allowing a third party to obtain the same full set of articles gathered in the original search
1.1 Bibliographic sources and gray literature searched are listed?	
Optimal (3 points)	All bibliographic sources investigated are listed and described in detail. For instance, when using the Web of Science (WoS) platform, all databases investigated are cited (e.g. Core Collection, SciELO Citation Index, etc.) or this information can be drawn through the methodology
Intermediate (1 point)	Bibliographic sources are partially listed, so a third party cannot repeat the search. For instance, when using the WoS platform, the databases investigated are not cited (e.g. Core Collection, SciELO Citation Index, etc.) and this information could not be drawn through the methodology
Inadequate (0 points)	Bibliographic sources are vaguely listed or not listed. Repeatability is low or not possible
1.2 Does the search provide a full search string used in all bibliographic sources?	
Optimal (3 points)	All search terms, Boolean operators ("AND," "OR" etc.), wildcards, and the field of search (e.g. search by "Topic" in WoS) are clearly stated, so that the exact search is repeatable by a third party
Intermediate (1 point)	The search is partially repeatable by a third party since (1) specific search terms are not stated or (2) Boolean operators were not stated (3) Boolean operators/wildcards/search terms are not stated well (so it is unclear how search terms were combined) or (4) there is no information on the field of search used
Inadequate (0 points)	Search is vaguely defined or undefined. Repeatability is low or not possible
1.3 Do the dates of database access provided?	
Optimal (3 points)	Enough detail is provided to allow the search to be repeated including the name of the database, the interface, and the last update. The date of the search is reported exactly as run, including year, month, and day of access
Intermediate (1 point)	Dates of accesses are partially documented, so a third party cannot repeat the search strategy because either (1) there is no information on the date of access for all bibliographic sources or (2) information on the date of access is missing (e.g. mentioning only the year of the last access, but not the day or at least the month of access)
Inadequate (0 points)	Information on the date of access is not provided. Repeatability is not possible

publications (approximately 25% of the dataset) to account for possible differences in the application of scoring criteria and potential biases introduced by different scorers' expertise. Final kappa values ranged from 0 to 1, and higher kappa values indicated a greater agreement (Cohen 1960; Landis & Koch 1977).

Data Analysis

We used descriptive statistics to enable comparisons between synthesis types. The mean scores between review types (i.e. meta-analysis vs. narrative synthesis) and scores awarded for each criterion were tested using a Mann–Whitney *U* test. We used R to perform statistics and elaborate figures (R Core Team 2019). We created a network map with the VOSviewer software (version 1.16.15) using the text mining functionality to visualize co-occurrences of keywords and then describe our population of reviews. VOSviewer is a software expressly designed to analyze and present bibliometric information (van Eck & Waltman 2010).

Results

Population of Reviews

We identified 79 reviews meeting our inclusion criteria (61 meta-analyses and 18 narrative syntheses) published between August 2009 and July 20 ≥1.1, based on the SCImago Journal Rank (SRJ). We found about 340 different author's keywords used to describe the review's subjects, where the most recently addressed research topics among restoration reviews were: "land-use change," "climate change," and "carbon sequestration" (Fig. 3).

Overall Repeatability at the Searching Stage

Scores from individual reviews varied in methodological repeatability ranging from 1 point to 9 (out of the maximum 9 points available; Fig. 4A). Overall, the mean score for all syntheses was 3.6 (median value = 3, and mode value = 1). Meta-analyses obtained a total mean score of 3.5, and narrative syntheses achieved a total mean score of 3.9, with no statistically

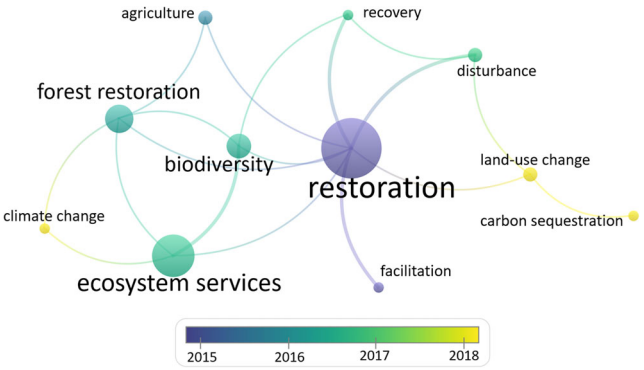

Figure 3. Network analysis of the main research topics addressed by reviews through the author's keywords. Network maps were limited to present items that showed a minimum of three occurrences (the same term or expression). The node size is proportional to the number of occurrences, and the thickness of the edges represents co-occurrences between items. Colors represent which time cluster the item belongs to as well as research subjects over time.

significant difference between synthesis types (Mann–Whitney U statistic = 750.5, $N_1 = 18$, $N_2 = 61$, $p < 0.05$). The mean score for each criterion also varied (Fig. 4B). Nonetheless, we found no statistically significant differences for all criteria considering review types (Mann–Whitney U statistic, $p < 0.05$); mean scores ranging from 0.8 to 1.3 for meta-analysis and from 0.8 to 1.6 for narrative synthesis.

Repeatability Test

Agreement among scores was high for all individual criteria (100% agreement for criterion 1.1, 85% for criterion 1.2, and 95% for criterion 1.3). Weighted kappa test showed 'perfect agreement' between scorers (results ≥0.86), indicating that assigned scores were consistent among scorers.

Discussion

The UN Decade on Ecosystem Restoration (2021–2030) was conceived to move efforts to withstand climate change, protect biodiversity, food security, and water supply (https://www.decadeonrestoration.org/). Evidence-based research in restoration ecology is well-positioned to move toward these goals. However, the reviews we evaluated showed an uneven level of methodological repeatability at the searching stage.

Only three studies were expressly intended to meet the optimal score of nine points (Acosta et al. 2018; Eales et al. 2018; Villemey et al. 2018). This was expected because, although systematic approaches were introduced in the environmental sector nearly 15 years ago (Pullin & Stewart 2006), guidelines expressly intended for the planning and conduct of rigorous environmental reviews are relatively recent (i.e. CEE 2013, 2018). The distribution of total scores shows that most reviews failed in effectively reporting all necessary information to allow an independent party to repeat the searching process and consequently to appreciate their strengths and weaknesses (Livoreil et al. 2017). For both meta-analyses and narrative syntheses, scores ranged from 1 to 9. Results did not vary according to review types, suggesting that the type of synthesis does not solely reflect the repeatability of a review at this stage. However, many reviews achieved the maximum score of 3 points for some criteria isolatedly (criterion 1.1, $n = 15/79$; criterion 1.2, $n = 14/79$; criterion 1.3, $n = 26/79$). Furthermore, some reviews

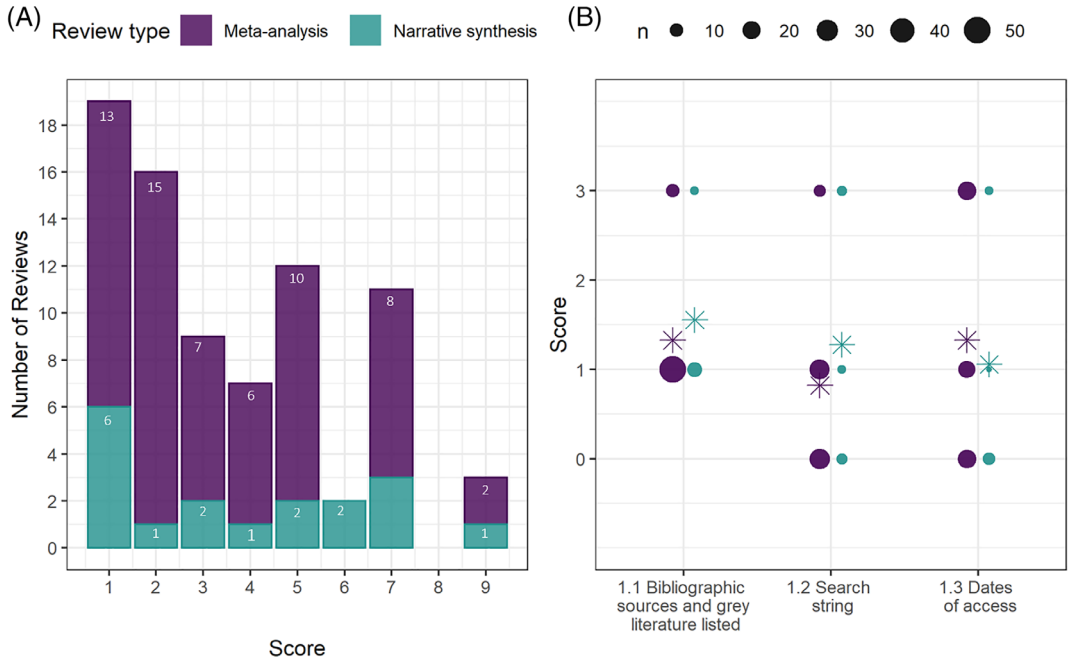

Figure 4. Distribution of reviews by overall scores awarded, according to the synthesis type. The total score is presented in "1" intervals to the maximum available score of 9. Meta-analysis (purple; $n = 61$), and narrative synthesis (blue; $n = 18$) (A); and distribution of individual scores for each repeatability criteria (spheres) and mean scores (asterisk) for meta-analyses and narrative synthesis at the search stage (B).

have even achieved relatively high total scores (e.g. ≥7, $n = 14/79$). Nonetheless, the searching process is only repeatable if all criteria have been fulfilled. Otherwise, reviews are assigned as partial or non-repeatable, depending on the number of items that were attended.

Following the principles of systematic review and maps, the searching stage is required to be based on a detailed reporting of strategies used (Moher et al. 2009), and the same level of detail should be ensured to describe both bibliographic databases and other web-based searching (Haddaway et al. 2017a; CEE 2018). Accordingly, an independent researcher must be able to know where (criterion 1.1. Bibliographic sources and gray literature listed), how (criterion 1.2. Full search string provided), and when (criterion 1.3. Dates of access provided) the search was performed using the information documented in the original study. Without these reports, an independent researcher cannot be sure whether the methods were suitable for the analysis and cannot verify or judge the review outcomes (Haddaway 2017).

Bibliographic Sources and Gray Literature

Knowing where searches were conducted, especially when using electronic databases, is crucial for allowing repeatability. Although all reviews have listed the databases used for searching studies, a common problem involving the Web of Science (WoS) description was evidenced. The WoS was used as a bibliographic source for most reviews in our population ($n = 73/79$). Nevertheless, this is not a single "database" as several authors reported ($n = 64/79$). This led to several reviews achieving the intermediate score for this criterion. The WoS is a platform on which many databases can be interrogated (Haddaway 2017). For example, the WoS Core Collections itself (the main WoS database) consists of eight databases: Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index- Science, Conference Proceedings Citation Index- Social Science & Humanities, Book Citation Index Science, Book Citation Index Social Sciences & Humanities, and Emerging Sources Citation Index. Furthermore, since different institutions subscribe to different databases within the WoS, and to different timeframes within those databases (Calver et al. 2017; Haddaway 2017) only referencing the platform also impaired repeatability. Additionally, reporting any limitations to searches is also an essential part of the searching process to ensure that readers can qualify evidence synthesis results (Haddaway et al. 2017a; Grames & Elphick 2020). In short, nuances of platforms, databases, and subscriptions mean that the central principle of repeatability across reviews can be easily compromised if authors do not report their activities in sufficient detail (Moher et al. 2009; Haddaway 2017). Sound examples of how searches should be reported to allow repeatability considering the standard that we used here can be found in Ceccon et al. (2016), Eales et al. (2018), Pezzati et al. (2018), and Vogel et al. (2018).

Search String

The information provided to depict how a search was conducted needs to include the full search string used (i.e. the combination of search terms, Boolean operators, quotation marks, and wildcards), as well as the field of search used (e.g. search for title, abstracts, and keywords) entered in each electronic bibliographic source. Only providing a poorly structured search strategy is not enough for repeatability because any change in the initial settings can lead to different results. As a consequence, nuances over review conclusions might emerge (e.g. Grames & Elphick 2020). Such problems contradict the optimal searching for literature that should be "systematic" (i.e. minimize risks of errors and biases) (Livoreil et al. 2017; Haddaway et al. 2017b; CEE 2018). Most reviews we assessed ($n = 57/79$) did not include information on the field of search used, which is a serious problem that compromises the assessment of repeatability, and further, the analysis of comprehensiveness (Woodcock et al. 2014; O'Leary et al. 2016; Romanelli et al. 2020a). Comprehensiveness is another important principle of systematic reviews and maps that supports the reliability of evidence synthesis (Haddaway et al. 2015, 2020). The lack of comprehensiveness at the searching stage can lead to so-called 'selection bias' (Bayliss & Beyer 2015), which occurs where the articles included in a review are not representative of the evidence base as a whole (McDonagh et al. 2013). Specifically about search strategies, selection bias affects syntheses through inappropriate arrangements (e.g. the combination and selection of terms) for the subject at hand (Haddaway et al. 2020). In this respect, a considerable number of reviews ($n = 20/79$) presented ambiguous or confusing search strategies, for example *(restoration AND genetic * in the title or with the words' genetic * "and either" restoration ecology "OR" ecological restoration "OR" restoration genetics' OR "revegetation" OR "rehabilitation" AND "min * [to distinguish postmining rehabilitation from medical rehabilitation]")*. In this example, there is a notable lack of clarity regarding how terms were effectively combined. The field of search used is presented along with the search string confounding the reader and compromising the repeatability. Further, there are no truncations limiting compound terms, which make the search strategy non-repeatable at all.

The absence of Boolean operators was also common among reviews ($n = 26/79$), as well as the absence of quotation marks to combine compound terms ($n = 4/79$). Even the total absence of the search terms used ($n = 2/79$) was evidenced, although it occurred less frequently. The following search string is an example of how several important items for achieving repeatability are lacking simultaneously in the same search string: *(limiting similarity, functional groups, functional similarity, invasion, community assembly, removal experiments, and biotic resistance)*. This search string is ambiguous because either the exact combinations of search terms, Boolean operators, truncations, and also the information on the field of the search were not specified, which can lead to different retrievals depending on how terms are combined. Importantly, we did not judge the relevance of the selected search terms used in reviews or if they fit for the

research purpose. For this respect, we recommend that authors engaged with evidence synthesis see Livoreil et al. (2017) for complete guidance on planning, selecting, preparing, conducting, reporting, and updating the search strategy based on systematic principles. For sound examples across restoration reviews on how the search string was effectively reported allowing repeatability (by our standard) see Gongbuzeren et al. (2015), Huang et al. (2019), Land et al. (2016), and Harzé et al. (2018).

Dates of Access

For the search strategy to be repeatable, it is equally important to report all relevant access dates on which searches were performed. Regarding this criterion, both review types (meta-analyses and narrative syntheses) reached relatively low scores because this information was often incomplete. Several reviews ($n = 31/79$) reported only the years comprising the search, ignoring the information about which month ($n = 18/79$) or day ($n = 40/79$) the search was performed, which led to most reviews ($n = 53/79$) being assessed as unrepeatable for this criterion. Knowing exactly when searches were performed is an essential concern since the number of available primary studies is frequently updated, and from the moment a search is completed, new results may be available, and these could change the review findings and conclusions if they were included (Livoreil et al. 2017). Moreover, the web-search eventually needs to be changed. This may occur when the review extends over a long period (e.g. more than 2 years) or when the publication rate on a relevant topic increases quickly, as it does in restoration ecology (Romanelli et al. 2018). In these situations, the conclusion of the synthesis may be out of date even before the study is published, therefore, it must be updated. In such circumstances, to repeat the web-search, all relevant dates must be objectively documented and well-described (Livoreil et al. 2017). Furthermore, reviews typically become out-of-date within several years of publication (i.e. approximately 1–6 years according to Shojania et al. 2007). Although we will likely be asking better questions in the future as our knowledge evolves, it may be important updates of reviews as new data and technologies are available, mainly if the publication rate of relevant documents on a certain topic is high (Livoreil et al. 2017). Transparently documenting all search activities can substantially support the efforts of updates, given that the study need not be repeated for the years covered by the original review (Haddaway 2018). For good examples of how the dates of access can be effectively reported to allow repeatability considering the standards we used, see Xu et al. (2019), Huang et al. (2019), Mijangos et al. (2015), and Timpane-Padgham et al. (2017).

Implications for Evidence-Informed Research

Reviews are of vital importance for effective environmental policymaking and practice, but their use requires recognizing the reliability of methods by which they were produced, which include the principle of repeatability (Haddaway et al. 2015; Plesser 2018). Many of the problems highlighted here could be readily solved by better reporting review activities, given that if a review explicitly and transparently details how it has been carried out, repeatability would be achieved.

Non-repeatable reviews may hide several undesirable problems. For instance, decisions may be informed by less rigorous (and/or biased) evidence; conflicting interpretations of evidence among reviews addressing similar topics may emerge (Grames & Elphick 2020); and no new (or updated) reviews may be produced on a specific subject, since researchers and decision-makers may be unaware that published reviews can lack rigorous standards to inform precisely the synthesis of research outcomes (cryptic evidence gaps) (Woodcock et al. 2017; Plesser 2018). Thus, assessment of review's repeatability (as we performed here) and other aspects that influence the trustworthiness of the review process are likely to be valuable for several end-users. An independent researcher, for example, can objectively identify questions lacking current high-quality reviews and direct new research, and editors or peer-reviewers can use this information in a preliminary evaluation of manuscripts to select high-quality studies (Seavy & Howell 2010; O'Leary et al. 2016).

Importantly, the criteria used in this study are a relatively crude measure, with no intention to include exhaustively every critical aspect influencing the repeatability of reviews at the searching stage. Nevertheless, we consider trends presented here as an important start toward improving the overall reliability of reviews in restoration ecology. It is also important to realize that there might be situations where less rigorous reviews (i.e. reaching relatively average-good repeatability scores) could still inform decisions if they are treated with appropriate caution (Romanelli et al. 2020a).

The characteristics of ecological studies provide several challenges that should be considered when planning and undertaking searches (Bayliss & Beyer 2015). A range of biases can affect the publication of ecological data in scientific journals (Leimu & Koricheva 2004; Budden et al. 2008), and each stage of the publication process could also be subject to different biases relating to the overall design of the study, or where is published (Lortie et al. 2007). It is expected that the movement toward evidence-based research in the environmental sector will face a structural and cultural shift similar to that seen in medicine, and may lead to improved search infrastructures for ecological studies (Pullin & Salafsky 2010). In the meantime, methods for data searching in ecology should be considered in the context of the sources and data available to attempt to reduce biases inherent to the data retrieval process and the information retrieved (Bayliss & Beyer 2015). Several forms of bias that may be encountered in the data retrieval process in ecology and steps that can be taken to counter them have been detailed extensively in Bayliss and Beyer (2015).

Ways by which authors could increase the overall reliability of their reviews include aspects such as following the evidence review guidance from the Collaboration for Environmental Evidence (CEE)—established in 2008 to act as the coordinating body supporting review efforts in the field of conservation and

environmental management (Haddaway et al. 2020). CEE guidelines (CEE 2018) promote, among other principles, a high level of objectivity and transparency, which underlies repeatability. As a part of the CEE systematic review process, the development of the protocol is also a way to ensure that intelligibility is high, pitfalls have been considered, and the transparency and clarity of the coming review have been taken into account. However, although the protocol tends to favor achieving repeatability, this is not a guarantee, as this depends on the rigor and precision of reporting the final review.

Despite the many advantages of following CEE guidelines to conduct high-quality systematic reviews, this approach is time- and resource-intensive, and such demands can make following the full process prohibitive for some authors operating on limited resources (Haddaway et al. 2015; Berger-Tal et al. 2019). Nonetheless, it is still worthwhile to adopt many of the principles of systematic reviews, such as transparency, to allow for repeatability, minimizing biases to which more traditional reviews are susceptible (O'Leary et al. 2016).

The CEE journal *Environmental Evidence* and *Ecological Solutions and Evidence* were the few avenues that we found for publishing evidence review protocols in the environmental sector. Thus, we reinforce the importance of environmental journals to highlight the parallels of performing rigorous evidence reviews, as well as to increase the financial support for the publication of protocols, considering they may incur a publication fee. Although alternative options for publication exist, they may lack a rigorous peer-review. Accordingly, securing affordable and fair (or free) open synthesis seems to be a key challenge for evidence review journals in the future (Haddaway 2018).

Additionally, a key frontier for future evidence-based research is to understand how individual researchers and their research teams can contribute to the generation of scientific knowledge. In this regard, the concept and framework of "research weaving" have emerged in ecology (Nakagawa et al. 2019). Research weaving seeks to synthesize both the research evidence and influence, combining systematic mapping and bibliometric approaches, where the repeatability of methods relies on as an important principle. For an extensive discussion on research weaving, we recommend Nakagawa et al. (2019).

Limitations

We acknowledge that we could have underestimated the total number of reviews in restoration ecology relevant to this analysis, considering that documents described with other terms regarding the context (e.g. ecosystem restoration) and the population (e.g. only review* itself) could also be relevant, as well as other articles out of the timespan comprised. Our search strategy was also dependent partly on how the authors described their study. Thus, the external validity of our conclusions must be evaluated in light of these imperfections. However, as we set our search strategy using terms that refer to reviews that imply to be systematic, we expected to avoid so-called traditional reviews (see Haddaway et al. 2015), which are broadly recognized to be susceptible to several biases (e.g. publication bias, selection bias, etc.) (Bayliss & Beyer 2015; O'Leary et al. 2016). Moreover, we inaugurated a platform to approach this subject in restoration ecology, hoping that trends showed here highlight the additional details that future authors should take at the searching stage when conducting a review.

Acknowledgments

This work was supported by the Research Foundation of São Paulo (Fapesp) (grant numbers 2013/50718-5, 2018/18416-2, 2018/21338-3, 2019/23908-4, 2019/24049-5, and 2019/08533-4).

LITERATURE CITED

Abdulla A, Krishnamurthy M (2016) Comparing retrieval of systematic review searches in health sciences areas using two major databases. Reference Reviews 30:1–5

Acosta AL, d'Albertas F, Leite MS, Saraiva AM, Metzger JPW (2018) Gaps and limitations in the use of restoration scenarios: a review. Restoration Ecology 26:1108–1119

Bayliss HR, Beyer FR (2015) Information retrieval for ecological syntheses. Research Synthesis Methods 6:136–148

Berger-Tal O, Greggor AL, Macura B, Adams CA, Blumenthal A, Bouskila A, et al. (2019) Systematic reviews and maps as tools for applying behavioral ecology to management and policy. Behavioral Ecology 30:1–8

Boell SK, Cecez-Kecmanovic D (2015a) Debating systematic literature reviews (SLR) and their ramifications for IS: a rejoinder to Mike Chiasson, Briony Oates, Ulrike Schultze, and Richard Watson. Journal of Information Technology 30:188–193

Boell SK, Cecez-Kecmanovic D (2015b) On being 'systematic' in literature reviews in IS. Journal of Information Technology 30:161–173

Brudvig LA (2017) Toward prediction in the restoration of biodiversity. Journal of Applied Ecology 54:1013–1017

Budden AE, Tregenza T, Aarssen LW, Koricheva J, Leimu R, Lortie CJ (2008) Double-blind review favours increased representation of female authors. Trends in Ecology & Evolution 23:4–6

Calver MC, Goldman B, Hutchings PA, Kingsford RT (2017) Why discrepancies in searching the conservation biology literature matter. Biological Conservation 213:19–26

Ceccon E, González EJ, Martorell C (2016) Is direct seeding a biologically viable strategy for restoring forest ecosystems? Evidences from a meta-analysis. Land Degradation and Development 27:511–520

CEE (2013) Guidelines for authors | Environmental evidence guidelines and standards for evidence synthesis in environmental management. https://www.environmentalevidence.org/wp-content/uploads/2018/02/Review-guidelines-version-4.2-final-update-1.pdf (accessed 4 April 2021)

CEE (2018) Guidelines for authors | Environmental evidence guidelines and standards for evidence synthesis in environmental management. https://environmentalevidence.org/information-for-authors/ (accessed 4 Apr 2021)

Chiasson MW (2015) Avoiding methodological overdose: a declaration for independent ends. Journal of Information Technology 30:174–176

Cohen J (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:37–46

Cram WA, Templier M, Paré G (2020) (re)considering the concept of literature review reproducibility. Journal of the Association for Information Systems 21:1103–1114

Dennis A, Valacich J (2014) A replication manifesto. AIS Transactions on Replication Research 1:1–4

Eales J, Haddaway NR, Bernes C, Cooke SJ, Jonsson BG, Kouki J, et al. (2018) What is the effect of prescribed burning in temperate and boreal forest on biodiversity, beyond pyrophilous and saproxylic species? A systematic review. Environmental Evidence 7:1–33

van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84:523–538

Gongbuzeren, Li Y, Li W (2015) China's rangeland management policy debates: what have we learned? Rangeland Ecology and Management 68:305–314

Grames EM, Elphick CS (2020) Use of study design principles would increase the reproducibility of reviews in conservation biology. Biological Conservation 241:108385

Haddaway NR (2017) Response to "Collating science-based evidence to inform public opinion on the environmental effects of marine drilling platforms in the Mediterranean Sea". Journal of Environmental Management 203: 612–614

Haddaway NR (2018) Open synthesis: on the need for evidence synthesis to embrace open science. Environmental Evidence 7:26. https://doi.org/10.1186/s13750-018-0140-4.

Haddaway NR, Woodcock P, Macura B, Collins A (2015) Making literature reviews more reliable through application of lessons from systematic reviews. Conservation Biology 29:1596–1605

Haddaway NR, Collins AM, Coughlin D, Kirk S (2017a) A rapid method to increase transparency and efficiency in web-based searches. Environmental Evidence 6:1–14

Haddaway NR, Land M, Macura B (2017b) "A little learning is a dangerous thing": a call for better understanding of the term 'systematic review'. Environment International 99:356–360

Haddaway NR, Bethel A, Dicks LV, Koricheva J, Macura B, Petrokofsky G, et al. (2020) Eight problems with literature reviews and how to fix them. Nature Ecology and Evolution 4:1582–1589

Hammersley M (2001) On 'systematic' reviews of research literatures: a 'narrative' response to Evans & Benefield. British Educational Research Journal 27:543–554

Harzé M, Monty A, Boisson S, Pitz C, Hermann J, Kollmann J, et al. (2018) Towards a population approach for evaluating grassland restoration—a systematic review. Restoration Ecology 26:227–234

Huang C, Zhou Z, Peng C, Teng M, Wang P (2019) How is biodiversity changing in response to ecological restoration in terrestrial ecosystems? A meta-analysis in China. Science of the Total Environment 650:1–9

James KL, Randall NP, Haddaway NR (2016) A methodology for systematic mapping in environmental sciences. Environmental Evidence 5:1–13

Johnson BT, Hennessy EA (2019) Systematic reviews and meta-analyses in the health sciences: best practice methods for research syntheses. Social Science and Medicine 233:237–251

Koricheva J, Gurevitch J, Mengersen K (2013) Handbook of metaanalysis in ecology and evolution. Princeton University Press, Princeton, New Jersey

Ladouceur E, Shackelford N (2020) The power of data synthesis to shape the future of the restoration community and capacity. Restoration Ecology 29:e13251

Land M, Granéli W, Grimvall A, Hoffmann CC, Mitsch WJ, et al. (2016) How effective are created or restored freshwater wetlands for nitrogen and phosphorus removal? A systematic review. Environmental Evidence 5:1–26

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174

Leidner DE (2018) Review and theory symbiosis: an introspective retrospective. Journal of the Association for Information Systems 19:552–567

Leimu R, Koricheva J (2004) Cumulative meta-analysis: a new tool for detection of temporal trends and publication bias in ecology. Proceedings of the Royal Society B: Biological Sciences 271:1961–1966

Linde K, Willich SN (2003) How objective are systematic reviews? Differences between reviews on complementary medicine. Journal of the Royal Society of Medicine 96:17–22

Livoreil B, Glanville J, Haddaway NR, Bayliss H, Bethel A, Lachapelle FF, et al. (2017) Systematic searching for environmental evidence using multiple tools and sources. Environmental Evidence 6:1–14

Lortie CJ, Aarssen LW, Budden AE, Koricheva J, Leimu R, Tregenza T (2007) Publication bias and merit in ecology. Oikos 116:1247–1253

McDonagh M, Peterson K, Raina P, Chang S, Shekelle P (2013) Avoiding bias in selecting studies. Methods guide for effectiveness and comparative effectiveness reviews. Agency for Healthcare Research and Quality. https://www.ncbi.nlm.nih.gov/books/NBK126701/ (accessed 4 Apr 2021)

Mijangos JL, Pacioni C, Spencer PBS, Craig MD (2015) Contribution of genetics to ecological restoration. Molecular Ecology 24:22–37

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement (reprinted from annals of internal medicine). Physical Therapy 89:873–880

Nakagawa S, Samarasinghe G, Haddaway NR, Westgate MJ, O'Dea RE, Noble DWA, et al. (2019) Research weaving: visualizing the future of research synthesis. Trends in Ecology & Evolution 34:224–238

O'Leary BC, Kvist K, Bayliss H, Derroire G, Healey J, Hughes KM, et al. (2016) The reliability of evidence review methodology in environmental science and conservation. Environmental Science and Policy 64:75–82

Paré G, Trudel M, Jaana M, Kitsiou S (2015) Synthesizing information systems knowledge: a typology of literature reviews. Information and Management 52:183–199

Paré G, Tate M, Johnstone D, Kitsiou S (2016) Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. European Journal of Information Systems 25:493–508

Peng (2011) High-impact astronomical observatories. Science 334:1226–1227

Pezzati L, Verones F, Curran M, Baustert P, Hellweg S (2018) Biodiversity recovery and transformation impacts for wetland biodiversity. Environmental Science and Technology 52:8479–8487

Plesser HE (2018) Reproducibility vs. replicability: a brief history of a confused terminology. Frontiers in Neuroinformatics 11:1–4

Pope C, Mays N, Popay J (2007) Synthesizing qualitative and quantitative health evidence. A guide to methods. Maidenhead: Open University Press.

Pullin AS, Salafsky N (2010) Save the whales? Save the rainforest? Save the data! Conservation Biology 24:915–917

Pullin AS, Stewart GB (2006) Guidelines for systematic review in conservation and environmental management. Conservation Biology 20:1647–1656

R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Romanelli JP, Fujimoto JT, Ferreira MD, Milanez DH (2018) Assessing ecological restoration as a research topic using bibliometric indicators. Ecological Engineering 120:311–320

Romanelli JP, Meli P, Naves RP, Alves MC, Rodrigues RR (2020a) Reliability of evidence review methods in restoration ecology. Conservation Biology 35: 142–154

Romanelli JP, Boschi RS, Almeida DRA, Rodrigues RR (2020b) Is the methodology used in reviews of restoration outcomes reliable? A systematic map protocol. Ecological Solutions and Evidence 1:e12030

Seavy NE, Howell CA (2010) How can we improve information delivery to support conservation and restoration decisions? Biodiversity and Conservation 19:1261–1267

Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D (2007) How quickly do systematic reviews go out of date? A survival analysis. Annals of Internal Medicine 147:224–233

Templier M, Paré G (2018) Transparency in literature reviews: an assessment of reporting practices across review types and genres in top IS journals. European Journal of Information Systems 27:503–550

Timpane-Padgham BL, Beechie T, Klinger T (2017) A systematic review of ecological attributes that confer resilience to climate change in environmental restoration. PLoS One 12:1–23

Villemey A, Jeusset A, Vargac M, Bertheau Y, Coulon A, Touroult J, et al. (2018) Can linear transportation infrastructure verges constitute a habitat and/or a corridor for insects in temperate landscapes? A systematic review. Environmental Evidence 7:1–33

Vogel HF, McCarron VEA, Zocche JJ (2018) Use of artificial perches by birds in ecological restoration areas of the Cerrado and Atlantic forest biomes in Brazil. Neotropical Biology and Conservation 13:24–36

Woodcock P, Pullin AS, Kaiser MJ (2014) Evaluating and improving the reliability of evidence syntheses in conservation and environmental science: a methodology. Biological Conservation 176:54–62

Woodcock P, O'Leary BC, Kaiser MJ, Pullin AS (2017) Your evidence or mine? Systematic evaluation of reviews of marine protected area effectiveness. Fish and Fisheries 18:668–681

Xu S, Liu X, Li X, Tian C (2019) Soil organic carbon changes following wetland restoration: a global meta-analysis. Geoderma 353:89–96

Supporting Information

The following information may be found in the online version of this article:

Supplement S1. Synthesis history.

Supplement S2. Summary of results.

Coordinating Editor: Stephen Murphy

Received: 20 April, 2021; First decision: 10 June, 2021; Revised: 26 June, 2021; Accepted: 26 June, 2021